Report - 1
Probability Basics required for SLAM

P Sai Ramana Kiran

July 25, 2017

Notation

Probability Spaces

Probability: They are numbers associated to some phenomena in the real world, in particular to the possible outcomes of some stochastic process.

Simple definitions before we go into in depth details.

Def: Outcome: Possible result of the experiment. Each possible outcome of a particular experiment is unique, different and mutually exclusive. Mutually exclusive, only one outcome can occur on each trail of experiment.

Def: Event: Collection of outcomes which are sets to which probabilities are assigned. Outcomes can also be modeled as singleton random events

Uncertainty in robotics: It comes from 4 levels:

→ Sensors: working limitations,lack of precision, transient effects

→ Actuators: Inaccurate model; Limitations; Transient response;

→ Environment: unless we have a controlled environment, examples static obstacles, environment introduces lot of noise

→ Processing Decisions: Floating point accuracy; psuedo random process.

Probability Space: This space is mathematically modeled as a tuple of 3 elements : <Need to insert the diagram of all the spaces>

- Sample Space (Ω): Set of all different and possible outcomes
- Random Events(E): Set of events or situations that can occur whose elements are sets of outcomes of sample space (Ω)
- Probability function (P): A function which maps an outcome from Random Events space (E) to a real number. Mathematically, $P : E \to [0, 1]$.

It formally defines the set of individual outcomes of the stochastic process of interest (Ω), a useful and well-formed grouping for those outcomes into random events (E) and a probabil- ity measure over those events (P)

Properties of Random Events (E): E is a *σ-algebra* of Ω. It is not an algebra. It is a collection of the subsets of Ω including empty set with few defined operations from set algebra and Calculus. Moreover, E is a subset of power set 2^Ω (all possible collections subsets of Ω). It has following properties:

- Operations: All set operations are valid. Some of the important ones are Union, Intersections, complements, differences and ordered pair.
- Closed: To cope with any kind of event using the above operations E must be closed. In other words, no matter which operation we perform on an event in E the resulting event must be still be in E.
  More formally, E must be closed under countable unions and intersections
  Simple example, if we take difference of same subset we get Φ. For it to satisfy closedness, we include Φ in the σ -algebra
- Measures: This is the probability function which assigns outcome to a non-negative real number. This quantifies the size or volume of the set of the σ-algebra
- Limits of sets: We can perform limit operations on set using this. For this, closure under countable unions and intersections is paramount. Two important limit concepts widely used in sets are supremum and infimum

Simultaneity of random events and outcomes: Several events are considered simultaneous if outcomes from all of them occur *within a given period of time.*
Example, Intersection and union operation of two or more events.

- Intersection of sets $\{A_1, A_2, A_3, \ldots, A_n\}$ is denoted by $\underset{i}{\cap} A_i$. It tells about the *all* the events occurring simultaneously.
- Union of sets $\{A_1, A_2, A_3, \ldots, A_n\}$ is denoted by $\underset{i}{\cup} A_i$. It tells about the *one or more* events occurring simultaneously.

From these examples, one point to be noted is that *time* is included in the probability spaces formalism *indirectly* through simultaneity concept. Also notice here that, till now we talked about different events (set of outcomes). Similarly, simultaneity of outcomes tells us that two different outcomes *cannot* occur simultaneously because of their *mutually exclusive* nature.

Properties of probability function (P): From above formulation of intersection and union, some important properties or constraints or axioms are as follows,

1. *Maximum Probability* tells us that occurrence of any outcome out of all possible outcomes (sample space Ω) is maximum and equal to 1

$$P(\Omega) = 1 \tag{1}$$

2. *Mutual exclusivity* of two events (let's say A_i and A_j) is denoted by $A_i || A_j$ and implies that $A_i \cap A_j = \Phi$. That is, two events won't occur simultaneously. Implication of this in union of two random events is

$$P(\underset{i}{\cup} A_i) = \underset{i}{\Sigma} P(A_i) \tag{2}$$

3. *Mutual Independence* of two events say A_i and A_j is denoted by $A_i \perp A_j$ and is occurred when probability of them occurring simultaneously can be calculated directly from product of their individual probabilities i.e.

$$P(\underset{i}{\cap} A_j) = \underset{i}{\Pi} P(A_i) \tag{3}$$

This implies that these events *occur* independently and their probabilities are not affected on what events of other set do. This property is of great importance in the context of robotics and different from mutually exclusive where they *cannot occur* simultaneously.

4. *Complementary probability* of an event A is denoted by $P(\overline{A})$ and is equal to

$$P(\overline{A}) = 1 - P(A) \tag{4}$$

5. It is important to note that non zero probabilities are assigned only to countable subset of events. For sample space which is continuous zero probability is assigned to every singleton random event unlike in discrete sample space case.
Best example to illustrate this point is from a continuous set of real numbers between $[0, 1]$ probability of selecting 0.5

Some points on Random Variables

1. Random variable(X) is defined as a function that maps outcomes to numerical quantities (labels), typically real numbers or vector of real numbers. In mathematical terms this statement is equivalent to , $X : \Omega \rightarrow \mathbb{R}^n, n \in \mathbb{N}$
2. Value of the random variable changes as the underlying the outcome changes. Hence it is a stochastic mapping.
3. **Support** of the random variable contains the subset of co domain for which value gives non-zero probability and is denoted by $supp(X)$
4. Similar to all functions, inverse function of random variables gives the set of outcomes from the probability space (Ω). Domain and co-domain for this function is defined appropriately.
5. **Cumulative Distribution function (CDF)** of a random variable is a function which yields probability of a particular set of values composed of any value that is less than or equal to a given number. It is denoted as F_X. Mathematically,

$$\begin{aligned} F_X : supp(X) \rightarrow [0,1], \forall a \in co(X) \\ F_X(a) &= P[X^{-1}(b) : b \leq a] \\ &= P[\{w \in \Omega : X(w) \leq a\}] \end{aligned} \tag{5}$$

   Here P is defined over probability space of random variable. P is denoted square brackets to indicate that it not a simple function, but a more complex mathematical operator

Some important properties of CDF

1. CDF yields probabilities of random variable unlike probabilities of random events. Even though implicitly we find probabilities of random events. Hence, its domain is $[0,1]$
2. CDF is a monotonically non-decreasing function. As the name suggests, it is cumulative distribution. Since, range is non-negative function it can either increase or remain constant. Mathematically, $F_X(a) \leq F_X(b), \forall a < b$
3. F_X is a *right continuous* for discrete and *continuous* for continuous probability spaces. To illustrate the difference, if F_X is both left and right continuous then probability of occurring an event is as follows

$$\begin{aligned} P[X = a] &= F_X(a) - \lim_{x \rightarrow a^-} F_X(x) \\ &= F_X(a) - F_X(a) \qquad [\because \text{ it is both left and right continuous}] \\ &= 0 \end{aligned} \tag{6}$$

   This illustrates that for continuous random variable, probability of occurrence of a singleton event, in this case a is zero. This actually matches with our intuition as mentioned in the previous section
4. Simple property, $\lim_{a \rightarrow max(supp(X))} F_X(a) = 1$[1]. Since, when maximum of support is considered, CDF calculates probability of entire sample space Ω
5. Similarly, $\lim_{a \rightarrow min(supp(X))} F_X(a) = 0$[1] Since, when minimum of support is considered, CDF calculates probability of no event in sample space Ω

[1]Here, a is considered to be a real number not a vector. This concept can be easily extended to a vector of real values.

Methods of Capturing Uncertainty!

Modeling Stochastic process:

CDF might be good probability measure, but it doesn't capture the underlying uncertainty of the stochastic process. To capture this, one needs to represent the process in a minimalistic fashion which will then bring out all the features of the uncertainty. CDF fails in this scenario due to its cumulative nature of representation. To this extent, one can define a function which will measure singleton event (in discrete case at least) probability rather than cumulative probability.

- In discrete case, this measure is called **probability mass function (pmf)** denoted by f_x[2]. Mathematically,

$$\forall a \in \mathbb{R}^n, f_X(a) = \begin{cases} P(\{w \in \Omega, X(w) = a\}), & \text{if } a \in co(X) \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

  One important condition, $\sum\limits_{a \in supp(X)} f_X(a) = 1$

- In continuous case, since probability of a singleton event is 0, a equivalent density function is defined over a region of support. This kind of function is called **probability density function (pdf)**. Mathematically,

$$\begin{aligned} &\forall (a, b] \subseteq supp(X), \\ &P[x \in (a, b]] \triangleq \int_{x \in A} \rho(x) dx \\ &\qquad = F_X(b) - F_X(a) \end{aligned} \tag{8}$$

  Few points on pdf:

  → PDF doesn't yield probability of a random variable x. Instead, it yields probability density of x

  → Since they don't define probabilities, their range can be outside $[0, 1]$.

  → Analogous to pmf, their integral over the entire domain of random variable yields 1

Quantifying Uncertainity

Now that stochastic process is modeled as pdf/pmf, time to address the ultimate question: How to quantify uncertainty ? What makes two stochastic process different? both qualitatively and quantitatively.

In literature there are number of ways to quantify uncertainty, but two important ways are

1. Moments of Random variable
2. Information and Entropy of Random Variable

Moments

They describe how the density/mass of the probability distribution is spread over the support of the random variable, in this case $c \in supp(X)$. Generally c is chose to be the first order moment μ_1^0. These moments are called central moments. An n-th order moment around a given value $c \in supp(X)$ is defined as follows :

$$\mu_n^c \triangleq \begin{cases} \int_{x \in supp(X)} (x - c)^n f_X(x) dx & \text{if } x \text{ is continuous} \\ \sum\limits_{x \in supp(X)} (x - c)^n f_X(x) & \text{if } x \text{ is discrete} \end{cases} \tag{9}$$

[2]Captial F is for indicating cumulative distribution (CDF). Small F is for individual distribution(pmf/pdf)

In probabilistic robotics, only 1st and 2nd moment are of primary interest, since they contain all the information which quantifies the shape of uncertainty of a stochastic process. They (1st and 2nd moments) are also popularly known as *mean* or *expectation* $E_X[X]$ and *variance* $V_X[X]$ of probability distribution respectively. Popularly denoted as μ_X and σ_X^2.
Mathematically for continuous case, [3]

$$E[X] \triangleq \int_{x \in supp(X)} xf(x)dx$$
$$V[X] \triangleq \int_{x \in supp(X)} (x - E[X])^2 f(x)dx$$

Properties of moments

1. Expectation is the value which one would expect to see if that stochastic process occurs. This value may not be in the support of the random variable.
2. Variance on the other hand denotes the average distance (in a crude sense) of the values that random variable can take with respect to the mean. Mathematically, $V[X] = E[(X - E[X])^2]$.
   Intuitively, a random variable with large variance has a flat shape (more dispersion from the center), while a random variable with small variance (less dispersion from the center) has peak shape.
3. Variance and Expectation are algebraically related as given. $V[X] = E[X^2] - (E[X])^2$. This relation is, again, intuitive since variance is measuring average distribution about the mean.
4. Expectation is a linear operator. That is, $E[aX + bY] = aE[X] + bE[Y]$. Where as, Variance is not.
5. Extension of Expectation to a function of random variable gives, $E[g(x)] = \int g(x)f(x)dx$, for continuous random variable and $E[g(x)] = \Sigma g(x)f(x)$

Self Information and Entropy

Self Information is another form of measuring or capturing uncertainty of the random variable. It forms a part of information theory. This measures the amount of information relayed (or need to be relayed) by a particular random variable. That is, as the confidence in an occurrence of th event increases the amount information relayed by the random variable decreases. In information theory, Self Information is also called surprisal. Mathematically, it is defined as

$$\forall\ a\ \epsilon\ supp(A), I(a) \triangleq log\frac{1}{P[X = a]} = -log f_X(a)$$

Entropy is also another measure for capturing the uncertainty derived from the information theory. But it is different from Self Information. Entropy is the expected information from a random variable. The uniform distribution has more entropy since the amount of uncertainty associated to that distribution is high. For example, consider a particle filter in robot localization setting. In the initial step, it assumes that the robot is everywhere in the given map. That is it's distribution of robot location is distributed uniformly. This implies that this kind of distribution has lot of information. Mathematically, Entropy is defined as

$$H[X] \triangleq E[I(X)] = -\underset{a \in supp(X)}{\Sigma} - P[X = a]logP[X = a] \quad \text{in Discrete case}$$
$$= \int_{f \epsilon support(X)} f_X(a)log f_y(a)dy \quad \text{in Continuous case}$$

[3]for discrete case replace $\int_{x \in supp(X)}$ with $\underset{x \in supp(X)}{\Sigma}$

For continuous case, instead of differential entropy, famous KL divergence is used (Kullback - Leibler Divergence). It is also known as relative entropy. It measures the distance between two probability distributions. That is, how one probability distribution diverges from another. Mathematically, it is defined as below. It is measuring how the probability distribution of P is diverging with Q

$$D_{KL}(P||Q) = \sum_i P(i) log \frac{P(i)}{Q(i)}$$

MultiVariate Probability

In robotics, particularly in Simultaneous Localization and Mapping, a single random variable cannot capture the uncertainty associated with it. Thus arises the need for mutlivariate probability.

Joint Probability and Marginalization

Joint Probability represents the relations between random variables (either defined on the same probability space or not). Joint probability distribution provides information about the probability that all the random variables have values with in the specified limits *simultaneously* All the properties of a regular probability are applicable to the Joint probability distribution.

For example, CDF defined on Joint Probability must satisfy the Maximum probability constraint defined in properties of probability function.

Conditional Probability

**continuing from the page 99. Need to write till then*

Conditional Expectation

It should be noted that a function of random variable in itself is a random variable. Going with this, Conditional Expectation is as it sounds an expectation over a set of conditioned events or in this case random variables. i.e. $E[X|Y]$. Mathematically, this is equivalent to,

$$E[X|Y] = \begin{cases} \sum\limits_{x \epsilon supp(X)} x f_{X|Y}(x) & \text{Discrete case} \\ \int\limits_{x \epsilon supp(X)} x f_{X|Y} dx & \text{Continuous Case} \end{cases}$$

It should be noted that this expectation($E[X|Y]$) itself has some uncertainty associated with it(Why??). This is due to its dependence of expectation on random variable Y. An interesting result with conditional expectation is as follows,

$$E[E[X|Y]] = E[X]$$

Probabilistic Graphical Models

Fusion of Graph theory and Probability is PGM. It is a representation of the conditional dependencies and their assumptions.
Need to port the hand written notes to the digital format