

Stochastic Processes and Modeling /

Discrete Time Dynamical System and Probability Theory

Office Hours: MTuTh 5-6pm

8/31/15

In terms of abstract mathematics, a stochastic process is a random function that maps a parameter domain into a state space.

Random: The value of the mathematical object depends not only on the explicit variables but also on some uncertain influence.

Parameter Domain: The set describing the explicit independent variable of the stochastic process, usually time, but can be space.

State Space: Range of the random variable(s) of interest.

Stochastic Process $X(t)$ or X_t where $t \in T$ (parameter domain) and $X \in S$ (state space)

The most basic classification of stochastic processes is in terms of the nature of their parameter domains and state space.

Discrete: finite or countably infinite, typical examples are finite sets like $(Z)_m = (1, 2, \dots, m)$, $(Z)_m^k = (1, 2, \dots, m) \times (1, 2, \dots, m) \times \dots \times (1, 2, \dots, m)$ or countably infinite sets like $(Z)_{\geq 0} = (0, 1, 2, 3, \dots)$, $(Z)_{\geq 0}^k = (0, 1, 2, \dots) \times (0, 1, 2, \dots) \times \dots \times (0, 1, 2, \dots)$ or modifications like $.25 \times (Z)_{\geq 0} = (0, .25, .5, .75, \dots)$.

Continuous: uncountable; typically continuous subsets of $(R)^d$ or $(C)^d$.

This class will only consider discrete state spaces.

As for the parameter domain, we will focus only on one-dimensional parameter domains with the independent variable typically being time. (could also, in principal, apply to one-dimensional space). Multi-dimensional spatial random fields(turbulent): require different stochastic frameworks, the theory of random fields see Yaglom, Correlation Theory of Stationary Random Functions

We will consider both discrete and continuous one-dimensional parameter domains, starting with discrete parameter domain.

See figure in class notes uploaded by Kramer

9/3/15

Read Karlin and Taylor 1.1-1.3 for Monday, HW 1 posted soon, due Thursday October 1st at 5 pm

Check out Bertsekas & Tsitsiklis, Introduction to Probability and Ross, A First Course in Probability. Probability theory was only properly mathematicized in the 1930s by Kolmogorov using measure theory. This movement separated mathematics from philosophy/intuition. Measure theory is a subject in real analysis which is concerned with mathematical structures that map sets into numbers. The upshot of the measure theoretic development is that a mathematically well-defined probability model consists of three ingredients:

1. A sample space Ω (note that this is not the state space of a r.v.) - This is an abstract space that encodes all possible outcomes of the components of the system of which we are uncertain, and which are relevant to the questions of interest. It's the "universe of possibilities." For example, suppose we want to consider the number of migrants entering the Budapest train station over the next month. Then the sample space might look something like this: Every point $\omega \in \Omega$ (called an elementary event) has the following information: # of migrants of migrants leaving Syria on day $n \in [1, 30]$ and the weather in Balkans for each day n , then for each migrant, their intended destination and whether they successfully pass a milestone on their journey to Europe.

Thus, Ω is the space of all possible values of this data.

2. σ -algebra \mathcal{F} of "measurable sets" or "events" - For the case of discrete sample spaces (which will apply for most of the class), we can just take \mathcal{F} to be the set of all subsets 2^Ω . In more complex settings (particularly with continuous spaces), one must pay more attention to this object for two reasons: (1) technical mathematics; for continuous sample spaces, it is not possible to "measure" every possible subset in a satisfying way and (2) encoding information flow, particularly in finance.

A σ -algebra is a mathematical structure that has the property that it is closed under the operations of finite unions, countable unions, finite intersection, countable intersections, and complements.

3. Probability measure: $P : \mathcal{F} \rightarrow [0, 1]$ with the following axioms:

- $P(A) \in [0, 1]$ for any $A \in \mathcal{F}$
- $P(\Omega) = 1$
- The probability measure is finitely and countably additive, meaning that given any collection $\{A_i\}_{i \in I} \subseteq \mathcal{F}$ where I is a finite or countable index set which has the property of being disjoint or mutually exclusive ($A_i \cap A_j = \emptyset$ for $i \neq j$). Then $P(\cup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$. In words: Probability that A_1 or A_2 or $A_3 \dots$ will happen.

Any choice of these three objects consistent with the above constraints gives a probability space or probability model: (Ω, \mathcal{F}, P) . All the modeling goes into how these mathematical structures are chosen to represent a given system. Once this probability model is specified,

then the calculations regarding the properties of the model can be put on mathematically rigorous and unequivocal foundations.

How does the modeling step (setting up the probability model) work in practice?

- sample space: define the relevant uncertain inputs and parameters
- σ – algebra: almost always implicitly taken to be the richest possible
- Probability measure: implicitly defined by giving a sufficiently detailed model to calculate any desired probability.

From the axioms of probability theory, one can deduce some fundamental relations:

- $P(\emptyset) = 0$
- $P(A^c) = 1 - P(A)$
- $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$
- Given $P(A_1), P(A_2)$ there is no general way to determine $P(A_1 \cap A_2)$

To make progress in computing intersections or unions, we need to develop a mechanism that describes the relationship between two events:

Definition 1 *Conditional Probability*

For any events $A, B \in \mathfrak{F}$ we define the conditional probability of event A given event B as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ provided } P(B) > 0$$

A very useful fundamental relationship between conditional probabilities is given by Bayes' rule:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

9/10/15

Law of Total Probability - Let $\{B_j\}_{j=1}^m$ be a partition of the sample space, meaning that:

- $B_j \cap B_{j'} = \emptyset$
- $\cup_{j=1}^m B_j = \Omega$

(intuitively, a chunking of sample space into an exhaustive set of mutually exclusive cases)

$$P(A) = \sum_{j=1}^m P(A|B_j)P(B_j)$$

This is very useful in simplifying calculations of probabilities of complex events A by introducing partial information B_j . The law of total probability also remains valid if the number of elements in the partition is infinite (but countable).

Independence

A collection of events $\{A_i\}_{i \in I}$ is said to be independent provided that for any finite subcollection indexed by a finite subset $J \subseteq I$, we have:

$$P(\cap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$$

In particular, two events A_1 and A_2 are said to be independent provided that:

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$

The more intuitive way to understand why this mathematical definition makes sense is to observe the following relations are equivalent:

$$P(A|B) = P(A) \iff \frac{P(A \cap B)}{P(B)} = P(A) \iff P(A \cap B) = P(A)P(B) \iff P(B|A) = P(B)$$

Random Variables

Intuitively a random variable is an uncertain number or collection of numbers. Mathematically, a random variable is a measurable mapping from sample space into a *state space*. $X : \Omega \rightarrow S_X$.

State space S_X of a random variable X gives a collection of possible values of the random variable X . (It's OK if it's bigger than the actual range of X).

Example 2 X is the number of migrants that arrive on day 3. Then $S_X \in \mathbb{Z}^+$. Notice that state spaces tend to be low-dimensional subspaces of \mathbb{R}^d or \mathbb{Z}^d . For example, in this model, we could also take a random variable $\vec{X} = (X_1, X_2, \dots, X_{30})$ corresponding to the number of migrants that arrive each day $n = 1, \dots, 30$. Random variables are essentially lower-dimensional projections of the randomness in the full probability model, which are more manageable to work and compute with.

How does one compute properties of random variables from the underlying probability model?

First, there is a technical restriction on random variables that will not concern us, namely we always demand that if $B \subseteq S_X$ is a reasonable (Borel) subset, then $X^{-1}(B) \in \mathcal{F}$. The uncertainty about the random variable can be described by a **probability distribution** which is the push-forward of the probability measure on sample space to the state space S_X via the mapping X . For any reasonable (Borel) subset $B \subseteq S_X$ we define the probability distribution of X as the measure:

$$P_X(B) \equiv P(X \in B) = P(X(\omega) \in B) = P(\omega \in X^{-1}(B)) = P(X^{-1}(B))$$

Probability distribution P_X is still a measure, which can be somewhat awkward for computations. It is desirable to reformulate in terms of functions.

- For finite-dimensional state spaces, one can always associate a cumulative distribution function (CDF, $P(X \leq x)$) to a probability measure, but this becomes still somewhat awkward to work with in more than one dimension.
- For practical calculations, it's often easier to use special-purpose frameworks if the random variables satisfy certain nice properties:
 1. Discrete
 2. Absolutely Continuous

First we'll talk about random variables that have discrete (finite or countably infinite) state space S_X . All subsets of S_X can be considered reasonable (Borel), and we can relate the probability distribution of X in terms of probabilities of elementary outcomes:

- $p_x = P_X(x) = P(X = x)$

By countable additivity,

$$P_X(B) = \sum_{x \in B} p_x$$

We must naturally have: $p \geq 0$, $\sum_{x \in S_X} p_x = 1$ Therefore, the probability measure has been related to a simple function p_x on the state space S_X . This function is known as the **probability mass function** (PMF).

Examples of discrete random variables associated to the migration model:

- the number of migrants arriving on a given day ($S_X = \mathbb{Z}^+$)
- characterizing the weather in the Balkans on a given day by a finite set of qualitative descriptors (1=fair, 2=difficult, 3=dangerous)

A probability mass function is completely informative about a discrete random variable. There are various summary descriptors that are simpler partial representation of the uncertainty.

- Mean or expected value or expectation of a random variable

$$X : \mu_X = \mathbb{E}X = \langle X \rangle = \sum_{x \in S_X} x p_x$$

- To characterize the uncertainty about the random variable, the simplest summary descriptor is the variance, or standard deviation (root mean square distance). Standard deviation σ_X , variance: σ_X^2

$$\text{Var}(X) = \sigma_X^2 = \mathbb{E}((X - \mathbb{E}X)^2) = \sum_{x \in S_X} (x - \mu_X)^2 p_x$$

In a practical sense, one can think roughly that a random variable with mean μ_X and standard deviation σ_X is quite likely to take values in the interval $(\mu_X - \sigma_X, \mu_X + \sigma_X)$. But notice that the mean and standard deviation do not completely specify the full probability distribution.

Some specific probability distributions on state spaces appear frequently in models; we will present some of these important distributions.

1. **Uniform Distribution** on a finite state space $\{1, 2, \dots, M\}$

- $p_x = \frac{1}{M}$ for all $x \in \{1, \dots, M\}$

2. **Poisson Distribution** on the state space \mathbb{Z}^+

- $p_x = \frac{e^{-\mu} \mu^x}{x!}$ for $x \in \mathbb{Z}^+$
- described by a single real parameter $\mu \geq 0$, in fact $\mu = \langle X \rangle$
- Poisson distribution arises in many probabilistic and stochastic models because of the **Poisson limit theorem**
 - If one is quantifying the total number of occurrences of some incident over some time interval, and if the incidents themselves can be represented/modeled as a sum of a large number of independent, rarely occurring incidents, then the total number of incidents will be approximately Poisson.
 - common to model the number of incoming agents to a node when the rareness/independence assumptions are good (cars entering an expressway, demand arriving at a server, signal arriving at a neuron)

3. **Geometric Distribution**

$$p_x = p(1 - p)^x \text{ for } x \in S_X = \mathbb{Z}^+$$

with $0 < p < 1$. Geometric distribution is a good model for the number of failures that occur before a success if the probability for success in each trial is p and if each trial is independent. - Memoryless property...

9/14/2015

Mathematical Framework for Stochastic Processes

Reading:

- Karlin and Taylor Sec. 1.1-1.3

Homework 1 due Friday, October 2 at 5PM

Continuous random variables

Probability mass functions don't work well because typically the probability for a random variable with a continuous state space to take any particular value is just $P(x = x) = 0$. So we must come up with an alternative way to represent probability distributions for continuous random variables. For an important class of continuous random variables, namely the "absolutely continuous" random variables, one can define a **probability density function (pdf)** $p_X(x)$ which has the property that for any reasonable (Borel) subset $B \subset S_X$:

$$p(X \in B) = \int_B p_X(x) dx$$

PDFs always have the following properties:

- $p_X(x) \geq 0$
- $\int_{S_X} p_X(x) dx = 1$

But PDFs are not probabilities, so they don't have to sum up to 1, nor must it be the case that $p_X(x) \leq 1$. Any PDF with the above true properties (and not pathological properties) will give rise to a valid absolutely continuous random variable.

What kind of continuous random variables are not absolutely continuous?

- "singular continuous": rarely occur in practice

If a PDF doesn't describe a probability, then what is it? It describes a density of probability, just as in continuum mechanics, mass densities describe densities of mass. That is, if $p_X(x)$ is continuous at point x_0 (which it doesn't need to be), then one can use the MVT to show that:

$$\lim_{\epsilon \rightarrow 0} \frac{P(|X - x_0| < \epsilon)}{2\epsilon} = \lim_{\epsilon \rightarrow 0} \frac{\int_{x_0-\epsilon}^{x_0+\epsilon} p_X(x) dx}{2\epsilon} = p_X(x_0)$$

The formulas for expectation also involve integrals:

$$\mathbb{E}X = \int_{S_X} xp_X(x) dx$$

Important examples

1. Uniform Distribution

$$X \sim U(a, b) \text{ (uniformly distributed on the interval } (a, b) \text{)}$$

We can take $S_X = \mathbb{R}$ or any subset of \mathbb{R} that contains the interval (a, b) . Specify the PDF:

$$p_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x < a, x > b \end{cases}$$

One can calculate that:

$$\mathbb{E}X = \int_{S_X} xp_X(x)dx = \int_a^b \frac{x}{b-a}dx = \frac{a+b}{2}$$

The primary use of the uniform distribution in this class is that random number generators are typically built up from an algorithm that generates a $U(0, 1)$ random variable.

2. Exponential Distribution

$$S_X = \mathbb{R} \text{ (or } \mathbb{R}_+ \text{)}$$

$$p_X(x) = \begin{cases} \frac{1}{\mu} e^{-\frac{x}{\mu}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

One parameter μ which turns out to be equal to the mean of the random variable:

$$\mathbb{E}X = \int_{S_X} xp_X(x)dx = \int_0^\infty \frac{x}{\mu} e^{-\frac{x}{\mu}} dx = \mu$$

The exponential distribution has the important property of being memoryless when interpreted as a model for a time delay. Mathematically, it has the property that for $0 < s < t$:

$$P(T > t | T > s) = P(T > t - s) \text{ if } T \sim \text{Exp}(\mu)$$

Interpretation: Think of T as the time required to wait until some incident occurs. Then if we have waited a time s without incident, then we know $T > s$. Then the probability that the incident still has not occurred at a later time t is the same as just recomputing probabilities by starting over from the last observed time s . Exponentially distributed random variables will arise naturally in stochastic processes where the time to wait for an incident does not depend on the past. It turns out that, up to weird pathological examples,

- exponential distribution is the only continuous random variable that has the memoryless property
- geometric distribution is the only discrete random variables that has the memoryless property