Independence, Random Variables

Thursday, September 10, 2015
1:59 PM

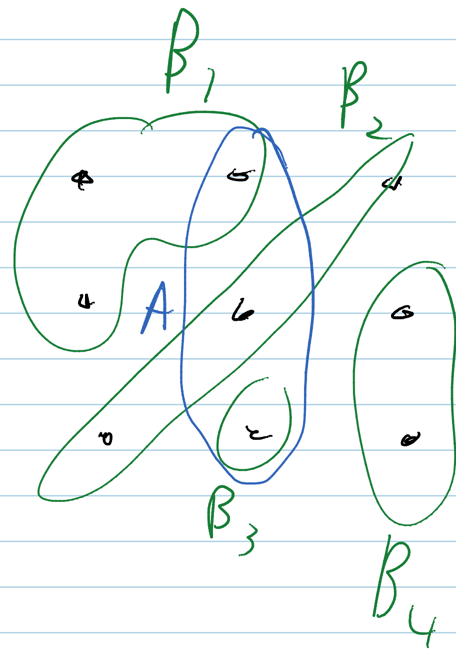
Reading:

- Karlin and Taylor Secs. 1.1-1.3

Homework 1 posted, due Friday, October 2 at 5 PM.

- homepages.rpi.edu/~kramep/Stoch/stoch2015.html

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 which simplifies the calculation. The law of total probability also remains valid if the number of elements in the partition is infinite (but countable).

Independence (definition!):

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:

$$P(A|B) = P(A)$$

$$\Updownarrow$$

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

$$\Updownarrow$$

$$P(A\cap B) = P(A)P(B)$$

$$\Updownarrow$$

$$P(B|A) = P(B)$$

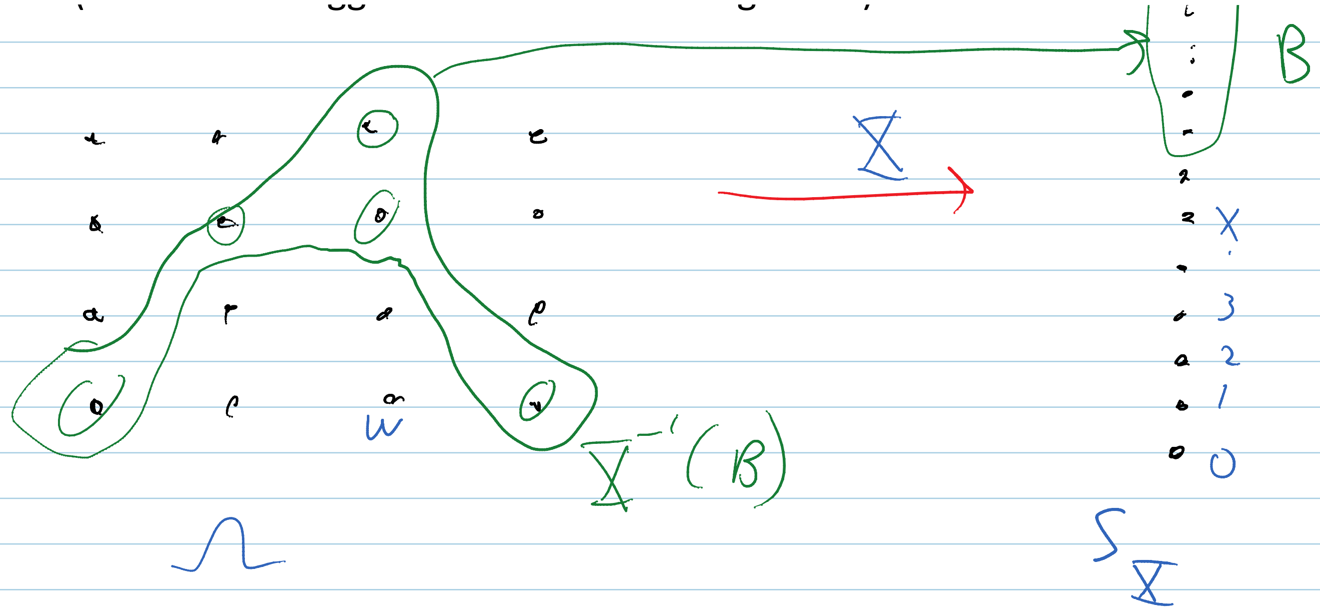
Random Variables

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. $X\colon\ \Omega \to S_X$

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).

B



Example: X is the number of migrants that arrive on day 3.

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 $X = (X_1, X_2, \dots, X_{30})$ 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 managable to work with/compute with.

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

First of all, there's 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 just 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:

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

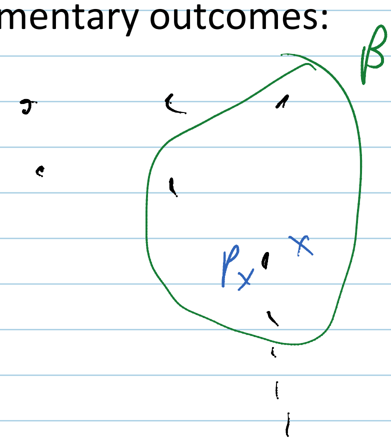
Probability distribution P_X is still a measure, which can be somewhat awkward for computations. Desirable to reformulate in terms of functions.

- For finite-dimensional state spaces, one can always associate a cumulative distribution function (CDF) 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
    i. discrete
    ii. absolutely continuous

First we'll talk about random variables that have discrete (finite or countably infinite) state space S_X

Then 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:

$p_x \geq 0$

$$\sum_{x \in S_X} p_x = 1$$

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.

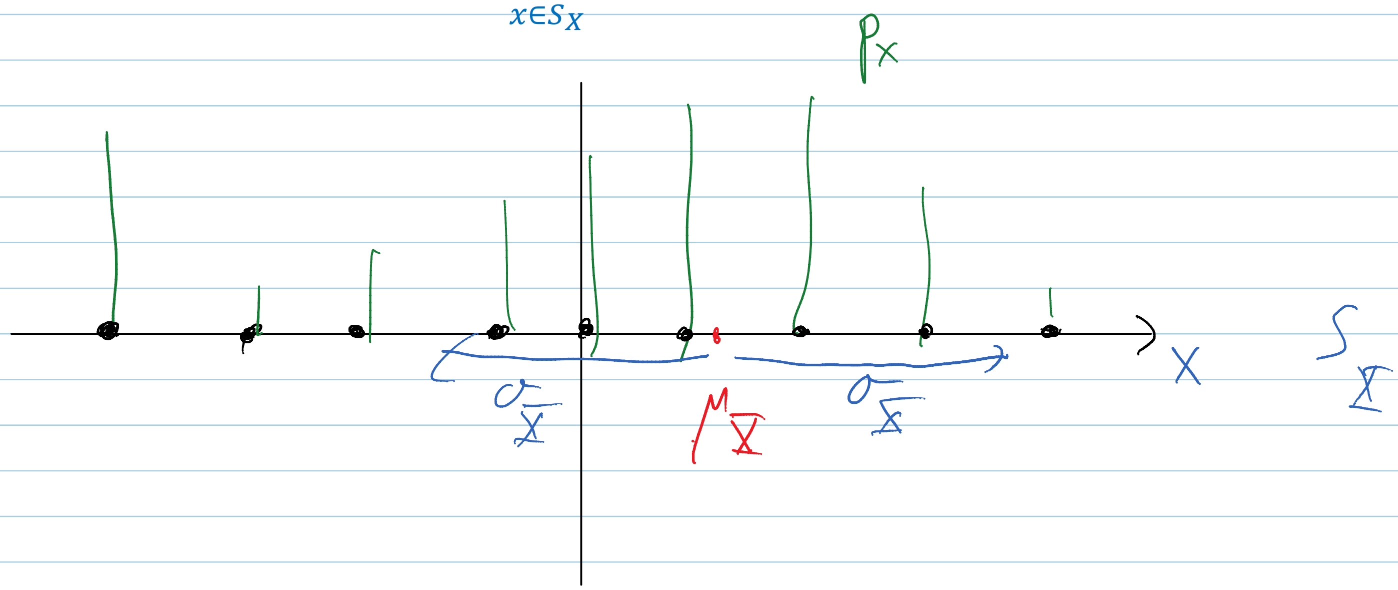
Examples of discrete random variables associated to the migration model:

- the number of migrants arriving on a given day ($S_X = \{0,1,2, \dots\} = \mathbb{Z}_{\geq 0}$)
- 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 representations 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

Standard deviation σ_X, variance: σ_X^2

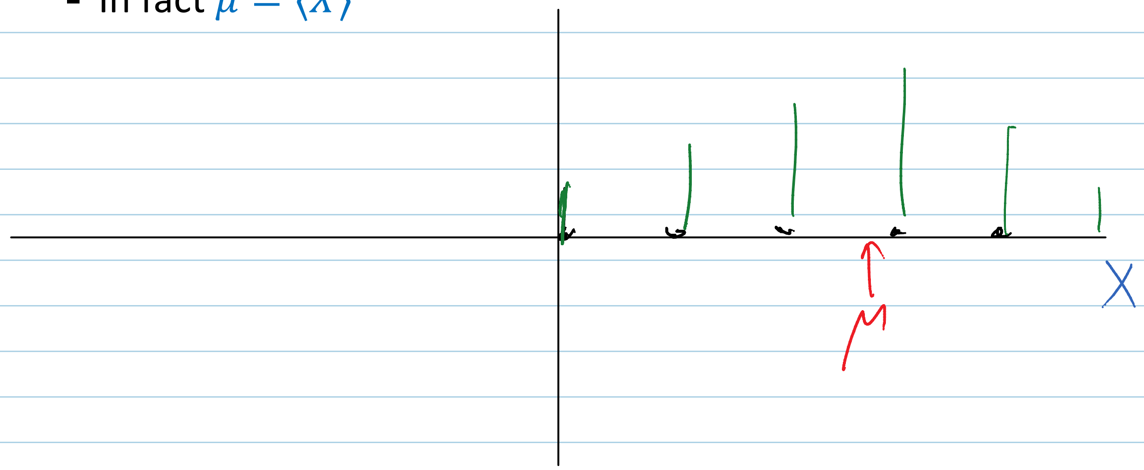
$$Var\,(X) = \sigma_X^2 \equiv \mathbb{E}((X - \mathbb{E}X)^2) = \sum_{x \in S_X} (x - \mu_X)^2 p_x$$

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 $\{1,2,\ldots,M\}$
   - $p_x = \frac{1}{M}$ for all $x \in \{1,\ldots,M\}$
2. Poisson distribution on the state space $\mathbb{Z}_{\geq 0} = \{0,1,2,\ldots\}$
   - $p_x = \frac{e^{-\mu}\mu^x}{x!}$ for $x \in \mathbb{Z}_{\geq 0}$
   - 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 independently, 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 assumptoins are good
  - cars entering an expressway, demand arriving at a server, signal arriving at a neuron.

3. Geometric distribution

$p_x = p(1-p)^x$ for $x \in S_X = \{0,1,2,\ldots.\}$

with parameter $0<p<1$.

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.

The geometric distribution can be seen to have a ``memoryless property'' which we will explain in the context of the exponential distribution and illustrate via the Bernoulli process in upcoming lectures.

Also, $\langle X \rangle = \frac{1}{p}$