6.262: Discrete Stochastic Processes 2/2/11

Lecture 1: Introduction and Probability review

Outline:

- Probability in the real world
- Probability as a branch of mathematics
- Discrete stochastic processes
- Processes to be studied
- When, where, and how is this useful?
- The axioms of probability theory
- Independent events and experiments
- Random variables

Probability in the real world

Games of chance started in ancient civilizations. Along with a propensity to gamble, people have an intuitive sense of likelihood and average behavior.

Games of chance are essentially repeatable, and experimental verification of likelihoods is essentially possible.

Most of life's decisions involve uncertainty. Wise people learn to associate some sense of likelihood with uncertain possibilities.

Probability is most useful when repeatability under essentially the same conditions occurs.

Essentially − essentially − essentially ??? In trying to be precise, many problems emerge.

In flipping a coin, the outcome depends on initial velocity and orientation, the coin surfaces, and the ground surface. Nothing is random here. Subsequent tosses are also related through the coin and the flipper.

Important questions involving uncertainty are far harder to make sense of. What is the probability of another catastrophic oil spill in the coming year? What is the probability that Google stock will double in 5 years?

Probability as a branch of mathematics

Despite discomfort with what probability 'means,' people have felt comfortable using combinatorics and symmetry to create probabilities for events in all areas of science and life.

Going one step further, standard models are created where events have probabilities and there are sensible rules for working with these probabilities.

Students are given a well-specified model and calculate various quantities. Heads and tails are equiprobable and subsequent tosses are 'independent.'

Everyone is happy. Students compute; professors write papers; business and government leaders obtain questionable models and data on which they can blame failures.

The use of probability models has 2 major problems:

First, how do you make a probability model for a real world problem?

Partial answer: Learn about estimation and decisions within standard models. Then learn a great deal about the real-world problem. Then use common sense and tread lightly.

Better answer: Try oversimplified models first. Use the mathematics of those simple models to help understand the real problem. Then in multiple stages, add to and modify the models to understand the original problem better.

Usually no model is perfect (look at coin tossing).

Alfred North Whitehead: Seek simplicity and distrust it.

Second problem: How do you make a probability model that has no hidden paradoxes?

Everyone's answer: Follow Kolmogorov's axioms of probability.

Kolmogorov did this in 1933, finally putting probability on a firm mathematical foundation and opening the field to steady progress.

These axioms essentially say that probability theory is a branch of measure theory.

These axioms are needed here to avoid paradoxes, but for the topics treated, measure theory is not needed and will not be used.

Discrete stochastic processes

A stochastic process is a special type of probability model in which the sample points represent functions in time.

It often can be viewed as a sequence of random variables evolving in time. Often there is a continuum of random variables, one for each real valued instant of time.

A discrete stochastic process is a stochastic process where either the random variables are discrete in time or the set of possible sample values is discrete.

It is not important to define which stochastic processes are discrete precisely.

Processes to be studied

Counting processes — Each sample point is a sequence of 'arrival' times. Special cases are Poisson processes (chap. 2) and Renewal processes (chap. 4).

Markov processes — The future state depends on the past only through the present. Special cases are Finite Markov chains (chap. 3), countable Markov chains (chap. 5) and Markov processes with countable state spaces (chap. 6).

Random Walks and martingales (chap. 7)

We will study various mixtures of these, particularly standard models for many applications. See table of contents (or text itself) for more detail.

When, where, and how is this useful?

Broad answer: Probability and stochastic processes are an important adjunct to rational thought about all human and scientific endeavor.

Narrow answer: Probability and stochastic processes are essential components of the following areas:

Communication systems and networks; computer systems; Queueing in all areas; risk management in all areas; catastrophe management; failures in all types of systems; operations research; biology; medicine; optical systems; control systems; etc.

The axioms of probability theory

Probability models have 3 components: a sample space Ω, which is an arbitrary set of sample points; a collection of events, each of which is a subset of Ω; and a probability measure, which assigns a probability (a number in [0, 1]) to each event. The collection of events satisfies the following axioms:

1. Ω is an event.

2. If $A_1, A_2, \ldots$, are events, then $\bigcup_{n=1}^{\infty} A_n$ is an event.

3. If A is an event, the complement A^c is an event.

Not all subsets need be events. Usually each sample point is taken to be a singleton event. Then non-events are weird, often necessary, but usually ignorable.

The empty set ϕ is Ω^c, so is an event.

If all sample points are singleton events, then all finite and countable sets are events (i.e., they are finite and countable unions of singleton sets).

From deMorgan's law,

$$\left[\bigcup_n A_n\right]^c = \bigcap_n A_n^c.$$

so countable intersections of events are events. All combinations of intersections and unions of events are also events.

The probability measure on events satisfies the following axioms:

1. $\Pr\{\Omega\} = 1$.

2. If A is an event, then $\Pr\{A\} \geq 0$.

3. If $A_1, A_2, \ldots$ are disjoint events, then

$$\Pr\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \sum_{n=1}^{\infty} \Pr\{A_n\} = \lim_{m \to \infty} \sum_{n=1}^{m} \Pr\{A_n\}$$

It's surprising that this is all that is needed to avoid paradoxes. A few simple consequences are

$$
\begin{aligned}
\Pr\{\phi\} &= 0 \\
\Pr\{A^c\} &= 1 - \Pr\{A\} \\
\Pr\{A\} &\leq \Pr\{B\} \leq 1 \qquad \text{for } A \subseteq B
\end{aligned}
$$

Another consequence is the union bound,

$$\Pr\left\{\bigcup_n A_n\right\} \le \sum_n \Pr\{A_n\}; \qquad \text{finite or countable } n$$
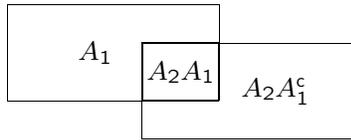
$$A_1 \bigcup A_2 = A_1 \bigcup A_2 A_1^c$$
$$\Pr\{A_1 \bigcup A_2\} = \Pr\{A_1\} + \Pr\{A_2 A_1^c\}$$
$$\le \Pr\{A_1\} + \Pr\{A_2\}$$

These axioms probably look ho-hum, and we ignore them much of the time. They are often needed for infinite sums and limits.

As in elementary probability courses, we emphasize random variables and expectations. The axioms, however, say that events and probabilities of events are the fundamental quantities.

Independent events and experiments

Two events A_1 and A_2 are <u>independent</u> if $\Pr\{A_1 A_2\} = \Pr\{A_1\}\Pr\{A_2\}$.

Given two probability models, a combined model can be defined in which, first, the sample space Ω is the Cartesian product $\Omega_1 \times \Omega_2$, and, second, for every event A in model 1 and B in model 2, $\Pr\{AB\} = \Pr\{A\}\Pr\{B\}$.

The two original models are then said to be independent in the combined model. We won't try to develop notation for this.

If the axioms are satisfied in each separately, they can be satisfied in the combined model, so complex models can be formed from simpler models.

Random variables (rv's)

Def: A rv X (or $X(\omega)$) is a function from Ω to $\mathbb{R}$. This function must satisfy the constraint that $\{\omega : X(\omega) \le a)\}$ is an event for all $a \in \mathbb{R}$. Also, if $X_1, X_2, \ldots X_n$ are each rv's, then $\{\omega : X_1(\omega) \le a_1; \ldots, X_n(\omega) \le a_n\}$ is an event for all $a_1, \ldots, a_n$ each in $\mathbb{R}$.

Every rv X has a distribution function $\mathsf{F}_X(x) = \mathsf{Pr}\{X \le x\}$. It's a non-decreasing function from 0 to 1.
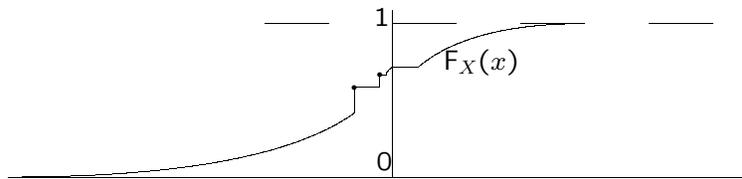
If X maps only into a finite or countable set of values, it is <u>discrete</u> and has a probability mass function (PMF) where $\mathsf{p}_X(x) = \mathsf{Pr}\{X = x\}$.

If $dF_X(x)/dx$ exists and is finite for all x, then X is continuous and has a density, $\mathsf{f}_X(x) = d\mathsf{F}_X(x)/dx$.

If X has discrete and continuous components, it's sometimes useful to view it as a density with impulses.

In general, $\mathsf{F}_X(x) = \mathsf{Pr}\{X \le x\}$ always exists. Because $X = x$ is included in $X \le x$, we see that if $\mathsf{F}_X(x)$ has a jump at x, then $\mathsf{F}_X(x)$ is the value at the top of the jump.

Theoretical nit-pick: $F_X(x)$ **must be continuous from the right, i.e.,** $\lim_{k \to \infty} F_X(x + 1/k) = F_X(x)$.

This seems obvious, since for a discontinuity at x, $F_X(x)$ **is the value at the top (right) of the jump.**

Proof: Let $A_k = \{\omega : X(\omega) > x + \frac{1}{k}\}$. **Then** $A_{k-1} \subseteq A_k$ **for each** $k > 1$.

$$\{\omega : X(\omega) > x\} \;=\; \bigcup_{k=1}^{\infty} A_k$$

$$\Pr\{X > x\} = \Pr\Big\{\bigcup_k A_k\Big\} \;=\; \lim_k \Pr\{A_k\} = \lim_k \Pr\Big\{X > x + \frac{1}{k}\Big\}$$

Center step: let $B_1 = A_1$; $B_k = A_k - A_{k-1}$ **for** $k > 1$. **Then** $\{B_k; k \geq 1\}$ **are disjoint.**

$$\Pr\Big\{\bigcup_k A_k\Big\} \;=\; \Pr\Big\{\bigcup_k B_k\Big\} = \sum_{k=1}^{\infty} \Pr\{B_k\}$$

$$= \lim_{k \to \infty} \sum_{m=1}^{k} B_m = \lim_{k \to \infty} A_k$$

6.262 Discrete Stochastic Processes
Spring 2011