Weekly Notes for EE2012 2014/15 – Week 1

T. J. Lim

January 12, 2015

Book sections covered in these two weeks: 1.1 – 1.5, 2.1 – 2.3.

1 Deterministic Versus Probabilistic Models

We have all been familiar with models since we were little. Model cars, planes, dinosaurs and people intrigue young children because they contain enough detail and so are "real enough" for them to study and examine[1]. As we grow up, we encounter other models, e.g. a model human skeleton, brain or other body part in a biology class, a model molecule in chemistry, a model solar system in physics. These are *physical models*, and they are all simplified depictions of reality, which contain enough detail for the purpose at hand.

In secondary school (or high school), we started studying physics using mathematical models, e.g. Newton's gravitational model, Ohm's Law, and Boyle's Law. These relate one quantity (mass, current, pressure) to another (force, voltage, volume) through a fixed functional relationship. These are *deterministic* mathematical models, which we know to be idealized but which have been proven by countless experiments to be very accurate at a macroscopic scale.

Finally, in many instances, uncertainty comes into the picture either due to an inadequate knowledge of the underlying natural processes, or due to an inherent randomness that is fundamental and unchangeable even with divine knowledge. Examples can be drawn from gambling, medicine, quantum physics, the Internet, mobile phones, and so on. To study uncertainy, we need a *probabilistic* mathematical model of the system, and this is the focus of this course.

In order to build up a probability model, certain fundamental concepts need to be carefully understood.

2 Fundamental Concepts in Probability Theory

2.1 Relative Frequency

When we roll a die many times, we intuitively believe that the number "1" should show up in about one in six tosses. We will then call such a die a "fair die".

[1]Look, mommy! This dinosaur has such tiny arms!

Suppose we denote the event "The face with one dot is on top" by A, we define the **relative frequency** of A in n rolls of the die as

$$f_A(n) = \frac{n_A}{n} \tag{1}$$

where n_A is the number of times event A occurs in n rolls.

Our intuition is that when n is large, $f_A(n) \to 1/6$, or in other words

$$\lim_{n\to\infty} f_A(n) = \frac{1}{6}. \tag{2}$$

It is natural to think of long-term relative frequency as the "probability" or "likelihood" of an event A, i.e. $\Pr[A] = \lim_{n\to\infty} f_A(n)$.

2.2 Belief

A problem arises when we consider events that arise from situations that cannot be repeated e.g. "It is raining outside LT7 at 10.00 am tomorrow", or "Roger Federer wins the 2015 Australian Open". In these cases, if probability is defined as long-term relative frequency, then there can be no way to assign a probability or likelihood value to these events. But still we have certain **beliefs** about the likelihood of rain or Roger Federer winning a tournament. Can we not numerically express these beliefs?

The modern theory of probability tackles this problem by using an axiomatic approach to allow both long-term relative frequency and belief to be admitted as interpretations of probability. We will expand on this point later.

2.3 Experiments, Sample Spaces and Events

A probabilistic scenario can always be described by an **experiment**, defined as a set of procedures and observations. An experiment can be as simple as

> Toss a coin two times, note the sequence of heads and tails obtained.

Or it may be rather grandiose, such as

> Lay down the physical laws of the universe, set the universe in motion starting from the birth of the universe, and observe whether it is raining on January 12th, 2015, outside LT7.

In all cases, an experiment produces one and only one **outcome** (observation) each time it is performed. Prior to performing the experiment, it is impossible to predict with certainty the outcome that will be produced. Hence the outcome is said to be random.

However, in spite of being unable to predict with perfect accuracy the outcome that will be seen when the experiment is performed, we can usually list all possible values the outcome can take. The set of all possible (or potential) outcomes is called the **sample space**, denoted S.

We saw many examples of sample spaces in the problem sets and in the lectures, and you should make sure you understand the idea.

An **event** usually has physical meaning, e.g. "Height of a randomly selected student is above 1.6 metres", or "Rainfall today is less that 10 mm". Every event can be translated into a subset of the sample space S, as follows: If an outcome $\zeta \in A$, then we say that the event A has occurred, otherwise A did not occur. For instance, if the experiment is "Pick a student at random from within NUS, and measure his/her height", and the event is $A =$"Height is above 1.6 metres", then $S = (0.5, 2.5)$ metres should be a safe range, and A corresponds to $(1.6, 2.5)$ metres because if and only if ζ (height of the randomly picked student) is in A will A have occurred.

2.4 Event Class or Field

While an event is necessarily a subset of S, we would also like to define in any probabilistic model the types of events that we can or should handle. Conceptually, the set or collection of all events of interest is called the **class or field of events**, denoted $\mathcal{F}$.

For instance, consider the experiment "Draw a card from a deck of 52 cards, note the card drawn". The sample space will comprise elements of the form (2,Heart), (5, Diamond), etc. Suppose we are only interested in the suits (Hearts, Diamonds, Spades, Clubs). Can we say that the field of events is

$$\mathcal{F} = \{\{(\times, \text{Hearts})\}, \{(\times, \text{Diamonds})\}, \{(\times, \text{Spades})\}, \{(\times, \text{Clubs})\}\}?$$

(In the above, $\times$ represents all thirteen possible values.) Actually, no. This would not be sensible because $\mathcal{F}$ must include *all* events of interest. It is not sensible to exclude the events "Hearts or Spades", "Card has the number 2", etc. from the set of events, is it? To ensure the inclusion of all such events, we define the following rules regarding what types of sets are admissible as event classes.

The rules governing the construction of $\mathcal{F}$ are (with A and B being subsets of S):

1. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
2. If $A, B \in \mathcal{F}$, then $A \cup B$ and $A \cap B$ are in $\mathcal{F}$.
3. The impossible event $\emptyset$ and certain event S must be in $\mathcal{F}$.

These rules in plain English ensures that (i) if we are interested in the probability of event A occurring, we must also be interested in the probability of it *not* occurring; (ii) if we are interested in the probabilities of A and B individually, we should also be interested in the probability of A *or* B occurring, and of A *and* B occurring; and (iii) we must always be able to state the probability of the impossible and the certain events. The rules also make $\mathcal{F}$ a "Borel field" in mathematical language.

3 Axioms of Probability

Consider a probability mapping

$$P : \mathcal{F} \to \mathbb{R} \tag{3}$$

where $\mathcal{F}$ is the field of events just described. The function $P(\cdot)$ takes as its argument (input) an event in $\mathcal{F}$ and outputs a probability value that is a real number. In this abstraction, there is no need to become entangled with the problem of non-repeatability of certain experiments (such as setting the universe in motion). However, to be consistent with the nice intuitive notion of long-term relative frequency as a measure of probability in repeatable scenarios, we dictate that the probability function obeys the following **axioms**:

1. $P(A) \geq 0$.
2. $P(S) = 1$.
3. If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.

These three axioms are natural extensions of the core properties of relative frequency. Being axioms, they cannot be proven. Rather, they can be seen as setting the rules of the game.

From the three Axioms of Probability, we derive all useful results[2] related to probability, which are now applied in far-flung domains ranging from detection and estimation, through communication networks and control systems, to insurance and financial derivatives.

3.1 Corollaries

Corollaries are results that easily follow from a more basic set of results, in this case the Axioms of Probability. Some of these corollaries may seem rather common-sensical to you, but nonetheless, you should know that they are obtainable from the Axioms of Probability.

1. $P[A^c] = 1 - P[A]$: This follows from the fact that A and A^c are mutually exclusive, that $A \cup A^c = S$, and that $P[S] = 1$.
2. $P[A] \leq 1]$: This follows from Corollary 1 by noting that $P[A^c] \geq 0$ from Axiom I.
3. $P[\emptyset] = 0$: This follows from $\emptyset \cup S = S$, and that $\emptyset \cap S = \emptyset$, so from Axiom III, $P[\emptyset] + P[S] = 1$ but since $P[S] = 1$, $P[\emptyset] = 0$.
4. $P\left[\bigcup_{i=1}^{k} A_k\right] = \sum_{i=1}^{k} P[A_k]$ if $A_1, \ldots, A_k$ are pairwise mutually exclusive. This is an extension of Axiom III to more than two events.

[2] And some that are not so useful!

5. $P[A \cup B] = P[A] + P[B] - P[A \cap B]$: This was proven in class with the aid of a Venn diagram.

6. $A \subseteq B \Rightarrow P[A] \leq P[B]$: Since $P[B] = P[A] + P[B \cap A^c]$ and $P[B \cap A^c] \geq 0$, the result follows.

4 Diagnostic Questions

1. Think of an example of randomness. Identify the experiment, outcome, sample space, a few example events, and then assign sensible probability values to these events, based on your personal experience.

2. Let $S = \{a, b, c, d\}$, $A = \{a, b\}$, $B = \{c, d\}$ and $C = \{b, c\}$. Is it valid to have $P[A \cup C] = 0.1$ and $P[B] = 0.5$? Explain your answer.