Indian Institute of Technology Bombay
Department of Electrical Engineering

Handout 4 EE 325 Probability and Random Processes
Lecture Notes 2 July 24, 2014

1 Classical Probability

The word *classical* in the title is a slight misnomer. The intention here is to differentiate it from the modern version (or axiomatic version). Classical probability is mostly the one that you learned already. It is about counting and computing the frequency of occurrence, or more generally computing ratios for real-valued variables. Though there are some reservations about its universal applicability, the classical version is very useful and the time spend there is of immense value. We can adapt and upgrade almost all that we learn here to the axiomatic framework of modern probability.

The following notations are reserved for the rest of this course.

$\mathbb{N}$ – natural numbers, $\mathbb{Z}$ – integers, $\mathbb{R}$ – real numbers,
$\mathbb{C}$ – complex numbers, $\mathbb{Q}$ – rational numbers

Before we start, two words of caution. The purpose of this chapter is not to prove anything in a rigorous fashion. So we will redefine many things that we learn now in a more rigorous way later. Several phrases are used in a loose sense, for example 'associate' or 'uniformly' at random. The later has a precise meaning, but we will reach there only after several lectures. For the time being, take it in their literal sense from the context. For example, 'uniformly' at random corresponds to some fair way of choosing among the outcomes.

2 Experiments, Outcomes and Events

Probability theory is concerned with experiments, whether physically conceivable or not. On a finer scale, we worry about

1. the possible outcomes of an experiment
2. interesting events that we wish to enquire about
3. degrees of assurance on various possibilities

Here are some simple examples.

1. A coin is tossed, we are interested in knowing whether a HEAD occurred.
2. A die is rolled twice, we wish to know whether the sum of faces is 7 or 11.
3. A bag has 3 blue balls and 2 red balls, we wish to know whether it will be red, if one ball is picked without looking into the bag.

In order to clearly articulate experiments like the ones listed above, we need at least three entities.

1. Outcomes of the experiment, or observations (to be denoted as Ω).
2. Events of interest (denoted by A or A_i, where $i \in \mathbb{N}$).
3. A measure that we can associate to each event (denoted by $P(A)$).

The first entity is referred to as **Sample Space**, which is the set of all outcomes of the experiment. The elements of the sample space are also known as **sample points**. Similarly, the events of interest belongs to the so-called **Event Space**, or the set of all *interesting* events. We are all familiar with the third entity, where the symbol P stands for **probability** or probability measure.

While the first two quantities are natural and unambiguous, the third one needs careful consideration.

2.1 Frequency Interpretation

In the classical interpretation, the frequency of occurrence of an event is assigned as the probability of the said event. In particular, for any **event** A, we will assign a probability as

$$P(A) = \frac{\#\text{outcomes favoring } A}{\text{Total number of outcomes}}. \tag{1}$$

Here 'favoring' is used in the sense, 'those outcomes which will lead to the said event A'. We will write this as,

$$P(A) = \frac{|A|}{|\Omega|}, \tag{2}$$

where $|A|$ counts the number of outcomes in the set A. Keep in mind, it is the number of **outcomes** that we count in assigning the probability to an **event** A.

While the theory is simple enough to comprehend, this may lead to inconsistencies, necessitating a more foolproof approach. We will develop that framework later, but let us first go through the frequency interpretation for some traditional examples.

Example 1 *Consider throwing a balanced die. We can articulate this as*

1. *Outcomes* $\Omega = \{1, 2, 3, 4, 5, 6\}$.
2. *Event-Space* $\mathcal{P}(\Omega)$*, let us take it as the set of all subsets of* Ω*). An example event of interest* $\{1, 6\}$*. This is equivalent to asking: is the outcome* 1 *or* 6*?*
3. *A probability measure*
$$P(A) = \frac{|A|}{|\Omega|}$$
where $|A|$ *denotes cardinality of the set* A *(to be read as card(A)).*

■

Notice that the third entity, i.e. the probability measure, to an extend summarizes our past or prior knowledge about the experiment. In the actual experiment, there maybe other exogenous factors including

- the softness of the hand

- the surface on which the die fell
- you prayed for a six or one!

But none of those listed is our concern, we simply strip them out of consideration and take a *naive* view, which in turn gives us the power to generalize.

Example 2 *Two balanced dice are rolled.*

1. $\Omega = \{(i,j) : 1 \le i \le 6, 1 \le j \le 6\}$
2. *Event Space* $\mathcal{P}(\Omega)$*. Example* $A = \{(i,j) : i + j = 10\}$.
3. $$P(A) = \frac{|A|}{|\Omega|}.$$

We can evaluate $P(A)$ *for the given* A *to be*

$$P(A) = \frac{3}{36} = \frac{1}{12}.$$

Example 3 *A fair coin is tossed* n *times.*

1. $\Omega = \{(\omega_1, \omega_2, \cdots, \omega_n) : \omega_i = \{H, T\}\}$.
2. *An example event* $A = \{(\omega_1, \cdots, \omega_n) : \sum_{i=1}^{n} \mathbb{1}_{\{\omega_i = H\}} = k\}$.
3. *For the event* A*, we can assign a measure*

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{n}{k}}{2^n}.$$

To show that classical probability is more than simple *counting*, let us consider the following example (from Hajek, Lecture-notes, UIUC, see ee325 website).

Example 4 *A traffic light repeats in cycles of length* $75s$*. The respective duration of green, orange and red are* $30s$*,* $5s$ *and* $40s$*. Suppose you do not have a watch, and drive into this traffic intersection and observe the lights.*

1. $\Omega = \{Red, Green, Orange\}$.
2. $A = Red$
3. *Probability of* A *can be assigned as*

$$P(Red) = \frac{40}{75} = \frac{8}{15}.$$

The above example also contains some not-so-true assumptions, that the traffic signal is not connected to the arrivals of vehicles and other road conditions.

Exercise 1 *Find the probabilities of observing Green, and that of Orange in Example 4.*

2.2 Bayes' Rule

Since probability is concerned with events, one can start thinking of what constitutes events. Of course events are those interesting sets on which we like to ask questions. For example, for an event $A = \{a, b, c, d\}$, the question 'whether A occurred?', is like asking if either one of a, b, c or d occurred, where a, b, c, d are among the possible outcomes. This is one difference about outcomes and events. Outcomes are mutually exclusive, i.e one of the outcomes do happen, and only one. Many events can happen simultaneously, in particular if $A_1 \subset A_2$, then A_1 happened will imply that A_2 also did happen (the reverse is not true.)

For our probability measure to be meaningful we should be able to tackle the set-operations of union ($\bigcup$), intersection ($\bigcap$), complement A^c etc. Frequency interpretation gives natural answers to such questions when there are only finitely many sets involved in the operations. Let us first look at the *union*.

$$P(A\bigcup B) = \frac{|A \cup B|}{|\Omega|} \tag{3}$$

$$= \frac{|A|}{|\Omega|} + \frac{|B|}{|\Omega|} - \frac{|A \cap B|}{|\Omega|} \tag{4}$$

$$= P(A) + P(B) - P(A\bigcap B). \tag{5}$$

In particular, when A and B are disjoint sets (i.e. $A \cap B = \varnothing$) then

$$P(A\bigcup B) = P(A) + P(B).$$

Let us now define a frequency interpretation of conditional probability. Let A and B are events associated with some sample space Ω.

$$P(A|B) \triangleq \frac{\text{No. of outcomes favoring } A \text{ and } B}{\text{No. of outcomes favoring } B} \tag{6}$$

$$= \frac{|A \cap B|}{|B|} \tag{7}$$

$$= \frac{\frac{|A \cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}} \tag{8}$$

$$= \frac{P(A \cap B)}{P(B)}. \tag{9}$$

Note: I accept that there is an imprecision in the above statement, i.e. 'events associated with some sample space Ω', what does it mean?. This will get eminently clear as we go on, but for the time being take the last equation above as the definition of conditional probability.

It is time to revisit our open-the-boxes example from the previous notes. In turns out that we can model the full exercise by the flip of a coin. That is, the experiment of switching the box and winning is equivalent to tossing a biased coin with $P(H) = \frac{2}{3}$.

Example 5 *We mentioned that 'switching' the box gives a better chance of winning in finding the ALICE problem. Let us find this probability, denoted as $P(W)$. Assume that if ALICE in not in Box I, the host randomly opens one of the other boxes to show a Leopard. If a leopard is in Box I, then the host opens the box with the other leopard.*

Solution: We will break the event of winning to two parts. Winning after getting door 2 opened and otherwise. Notice that our first choice is always Box 1. We will denote by D_{2O} the opening of door 2. Since D_{20} and D_{20}^c are disjoint sets,

$$P(W) = P(W\bigcap D_{2O}) + P(W\bigcap D_{20}^c) \tag{10}$$

What is the probability of $W \bigcap D_{2O}$? It is asking about both door 2 opening and win happening together. Consider $\Omega_1 = \{LLP, LPL, PLL\}$, we know that each outcome here has probability $\frac{1}{3}$. Of this, only the event LLP has both D_{20} and W happening. Hence $P(W \bigcap D_{2O}) = \frac{1}{3}$. Same is the case where door 3 gets opened, and thus,

$$P(W) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}, \tag{11}$$

the answer we got in class.

Note: Please note that we used a rule called Bayes Rule II in deriving this result. This rule will be covered in detail in the later chapters. ■

The 3 Box question can also reveal some intriguing aspects with respect to conditional probability. This explanation can clear any more doubts on this question.

Example 6 *Compute the probability of winning given that D_{2O} happened.*

Solution: $P(W|D_{20})$ is looking at the fraction of winning events which also has D_{20}, as compared to the fraction of W or W^c which also has D_{2O}. The latter is nothing but the probability of D_{2O} itself. It is clear that the fraction of D_{2O} is $\frac{1}{2}$. The fraction of both D_{2O} AND W is $\frac{1}{3}$ (i.e. LLP happens). Thus

$$P(W|D_{20}) = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}. \tag{12}$$

We can verify the formula for winning using Bayes' Formula as

$$\begin{aligned}
P(W) &= P(W \bigcap D_{2O}) + P(W \bigcap D_{2O}^c) && (13)\\
&= P(D_{2O})P(W|D_{20}) + P(D_{2O}^c)P(W|D_{20}^c) && (14)\\
&= \frac{1}{2} \times \frac{2}{3} + \frac{1}{2} \times \frac{2}{3} && (15)\\
&= \frac{2}{3}. && (16)
\end{aligned}$$

■

Exercise 2 *Use the above exercise to argue that in finding the LEOPARD problem, knowing which other box has a leopard does not change the probability of a leopard being in the first box.*

3 Bertrand's Paradox

If we can assign probabilities for each event of interest in the manner shown in last section, then there is little need for an alternate approach. However, our probability assignment strategy may fail, or lead to inconsistencies, as exemplified by the so-called Bertrand's Paradox. In this paradox, three *seemingly* similar experiments will end up giving dramatically different answers, and to make it worse, each one is true. Betrand's paradox is about circles and chords, recall that a line which connects any two points of the circle is known as a chord.

> Consider a circle or radius r centered at the origin. Suppose a cord AB is chosen at *random*, what is the probability that $l(AB) > \sqrt{3}r$, where $l(AB)$ denotes the length of the chord AB.

Before we start solving, realize that $\sqrt{3}r$ is nothing but the side of an equilateral triangle inscribed in the circle, see Figure 1. The sides are calculated using

$$\cos 30 = \frac{\sqrt{3}}{2} \text{ and } \sin 30 = \frac{1}{2}.$$

We will provide three different methods, each one *appearing* to solve this problem.

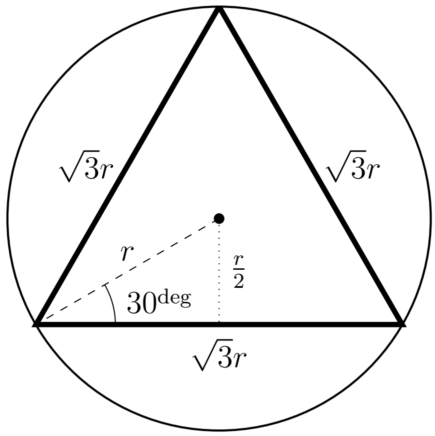

Figure 1: Inscribed equi-lateral triangle has side $\sqrt{3}r$

3.1 Method I: Random End-points

Let us first pick a point A on the circle. We will form the chord AB by randomly picking the second point B on the circle. For example, A is chosen as in Figure 2. Now, if the

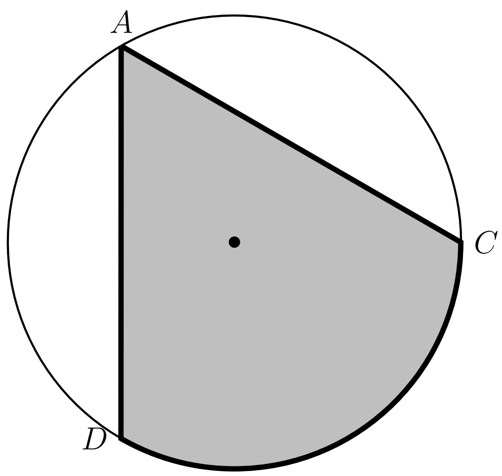

Figure 2: End points at random

second point lies anywhere in the segment CD of the circle, the corresponding chord will lie in the shaded region, which in turn implies that $l(AB) \geq \sqrt{3}r$. However, CD is exactly $\frac{1}{3}$ of the circumference. Since B is chosen uniformly on the circumference,

$$P(l(AB) > \sqrt{3}r) = \frac{1}{3},$$

when A is chosen as shown. However, if we change A, the problem stays the same and our computations are independent of the initial choice of A. Thus $\frac{1}{3}$ is the probability that we look for. ■

3.2 Method II: Random Midpoint

In this method, a point inside the circle is chosen at random. Observe that any point P other than the origin uniquely corresponds to a chord with P as the mid-point. There are many possibilities when P is indeed the circle center. However we are free to draw any chord as we please, once the origin is chosen. We will learn later that the probability of getting origin in this experiment is anyway zero. Consider Figure 3, and notice that when the chosen point falls inside the shaded region, the corresponding chord will have length greater than $\sqrt{3}r$. This is because the inner circle is the locus of all the center-of-chords with length $\sqrt{3}r$. Since the mid-points are uniformly chosen, the chance that a point in the shaded region is chosen is proportional to the area of the shaded region. Thus,

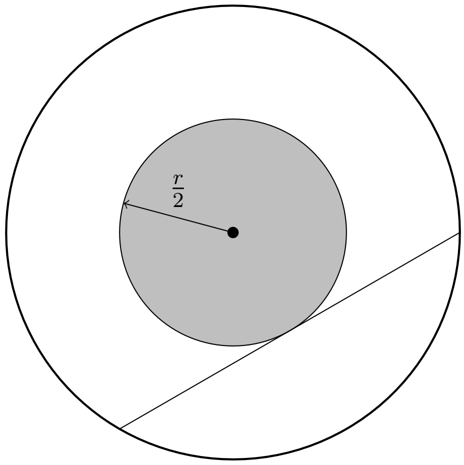

Figure 3: The midpoints of chords are uniformly chosen

$$P(l(AB) \geq \sqrt{3}r) = \frac{\pi(r/2)^2}{\pi r^2} = \frac{1}{4}.$$

■

3.3 Method III

Another way of finding a solution to our problem is to consider random radial lines, i.e. an angle is uniformly chosen in the interval $[0, 360]$ and a radial line is drawn from origin at this angle to the positive horizontal axis. Figure 4 illustrates this, where the radial is drawn at an angle -75^{deg}. Once the radial line is chosen, a point is uniformly picked on this radial line. The picked point then acts as the center of a chord for our original circle (notice that this chord is normal to the radial line, see Figure).

Given a radial line, if a point picked is within a distance of $\frac{r}{2}$ from the center, then it will fall inside the boundary of the inner-circle. Like in the last problem, a point in the inner circle will lead to a desired chord.

Thus the probability we seek is the probability of choosing a uniform value in the first half of the radial line.

$$P(l(AB) \geq \sqrt{3}r) = \frac{1}{2}.$$

■

So the three methods gave completely different answers and leave us with the question, 'which one shall we believe?'. The key to this lies in the fact that the three methods are

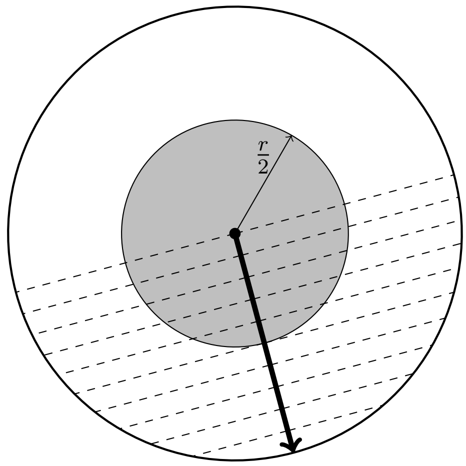

Figure 4: First choosing a radial and then a point on it

attempting to find three different quantities, none of which we can unequivocally call as the probability of $\{l(AB) > \sqrt{3}r\}$. In particular, the association with a probability measure is part of the problem definition itself, and we can define many such measures on a given sample space with the same events of interest. In order to throw some more light, let us consider the event E_m that the midpoint of a chord is chosen inside the inner circle of radius $\frac{r}{2}$, and compute $P(E_m)$ for the three methods. Clearly $P(E_m) = \frac{1}{3}$ for Method-I, implying that a lot more centers are possibly chosen in the inner circle than Method II. Method II will have a uniform distribution of points since we chose it that way. Thus about a quarter of the points is expected to be inside the inner circle using Method-II. On the other hand, Method III is expected to pack $\frac{1}{2}$ of the points inside the inner circle. Conversely, Method I and III will choose possibly less number points in the outer ring, leading to a diluted density of mid-points there.

The above paradox is carefully tackled in the axiomatic approach by including the probability association as a part of the problem definition itself. More details are included in the coming Section.

Exercise 3 *Some one stays behind a curtain, tosses a fair coin three times and then shouts loudly from behind. "Two HEADs are there" (at least). What is the probability that all results are HEAD?*

Exercise 4 *Some one stays behind a curtain, tosses a fair coin three times and then shouts loudly from behind. "The first two results are HEADs". What is the probability that all results are HEAD?*

Exercise 5 *A bag contains* 8 *red balls and* 6 *blue balls. If* 5 *balls are picked at* random*, find the probability that* $2\#\{RED\} + \#\{BLUE\} = 8$*, where the notation* $\#\{RED\}$ *counts the number of RED balls picked.*