Indian Institute of Technology Bombay
Department of Electrical Engineering

Handout 7 EE 325 Probability and Random Processes
Lecture Notes 4 August 4, 2014

1 Properties of Probability

Based on the three probability axioms, we can derive many properties of the probability measure.

Many of you are already familiar with an alternate notation for intersection of sets, widely written as a simple product form, i.e. $A \bigcap B$ and AB mean the same, whenever A and B are clear from the context. Let us start by proving a general law for sets, which will turn out to be useful later.

$$\left(\bigcup_{i=1}^{n} A_i\right) \bigcap B = \bigcup_{i=1}^{n} (A_i \bigcap B) \tag{1}$$

The proof of this is very simple, but is listed here to refresh your ideas. Denote $A_i \bigcap B$ by $\hat{A}_i$ for convenience. Then, for any element x,

$$\begin{aligned} x \in \left(\bigcup_{i=1}^{n} A_i\right) \bigcap B &\Leftrightarrow x \in \bigcup_{i=1}^{n} A_i \text{ and } x \in B \\ &\Leftrightarrow x \in A_i \text{ for some } i \text{ and } x \in B \\ &\Leftrightarrow x \in (A_i \bigcap B) \text{ for some } i \\ &\Leftrightarrow x \in \hat{A}_i \text{ for some } i \\ &\Leftrightarrow x \in \bigcup_{i=1}^{n} \hat{A}_i \end{aligned}$$

Property 1 *If $B \subset A$, then $P(B) \leq P(A)$.*

Proof: We should use the probability axioms to obtain this property, particularly Axiom II, which deals with mutually disjoint sets. Can we write A as the union of mutually disjoint sets, at the same involving the subset B.

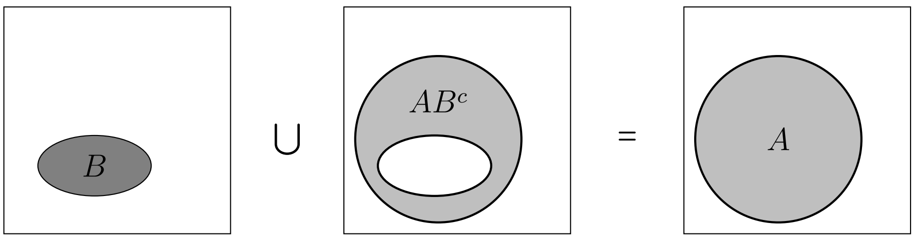

Figure 1: $A = B \bigcup AB^c$ when $B \subset A$

Indeed one can, as shown in Figure 1. Using the fact that $B \subset A$,

$$A = B \bigcup (A \setminus B) = B \bigcup AB^C$$

Applying axiom II,

$$\begin{aligned} P(A) &= P(B \bigcup AB^c) \\ &= P(B) + P(AB^c) \\ &\geq P(B). \end{aligned}$$

where the last step follows from axiom I, i.e. $P(\cdot) \geq 0$. ■

Using Venn diagrams as above is a good method to verify set properties when there are a small number of participants. As the operations involve many sets, a very useful technique is to use mathematical induction, which we illustrate by the next two properties.

Property 2

$$P(A \bigcup B) = P(A) + P(B) - P(AB) \tag{2}$$

Proof: Notice that any set A can be partitioned as $A = AB \cup AB^c$, where B is any other set. Furthermore, the partitions are mutually disjoint, thus

$$P(A) = P(AB^c) + P(AB).$$

We can also partitioning $A \cup B$ into three disjoint parts in a convenient way to write

$$A \bigcup B = AB^c \bigcup BA^c \bigcup AB. \tag{3}$$

Applying axiom II

$$\begin{aligned} P(A \bigcup B) &= P(AB^c) + P(BA^c) + P(AB) \\ &= P(AB^c) + P(BA^c) + P(AB) + P(AB) - P(AB) \\ &= P(A) + P(B) - P(AB), \end{aligned}$$

where the last equation used (3). ■

We now extend this to any number of sets. Instead of learning the analytic details first, let us look at an example. This is popularly known as the '*coincidences problem*.

Example 1 *There were M boys M girls in the senior secondary graduating batch in a school. Suppose the school pairs each boy with a girl to do lab experiments. They graduated, many boys went to IITs, girls to Medical Schools etc. After* 10*years they re-assembled for the alumni meet at the school. M tables numbered $1, \cdots, M$, each with a pair of chairs, were arranged for the function. The ladies went first and each lady occupied a table. If the gents now walk in at a random order and occupy the first vacant seat, what is the probability that at least one of the table has an actual pair from the school days.*

There are many variants to this problem, the above one may appear silly, but it should help you get interested. Without loss of generality, let us associate the table number with the lady sitting there. We will do a counting argument as all combinations are equally likely. Clearly there are $M!$ ways of men seatings. Let $A_i, 1 \leq i \leq M$ be the event that there is a match at table i. The quantity of interest is $P(\bigcup A_i)$.

Suppose the first table got an actual pair, the rest of the men can now sit in $(M-1)!$ ways.

$$P(A_1) = \frac{(M-1)!}{M!} = \frac{1}{M}.$$

We also know that

$$
\begin{aligned}
P(A_1 \bigcup A_2) &= P(A_1) + P(A_2) - P(A_1, A_2) \\
&= \frac{1}{M} + \frac{1}{M} - \frac{(M-2)!}{M!} \\
&= \frac{1}{M} + \frac{1}{M} - \frac{1}{M(M-1)}.
\end{aligned}
$$

We can generalize this ides, we will add all those options where we can ensure that the i^{th} table has a pair. The trouble is that we are doing some double countings!. As an example, imagine a case where the first two tables have matching pairs, this will be counted in the $(M-1)!$ combinations with the first matching, as well as for the $(M-1)!$ matchings in the second table, we should only be counting it once. A pair of matching at the first two tables will have $(M-2)!$ combinations for others. However, in this last event, we will double counting all the cases where there were three or more matchings. More specifically any sequence with the first three tables matching will be counted into two separate pairs, and we will have to deduct one of those. If you have any doubt, just run through the question with two tables as we illustrated above. Putting it all together,

$$
\begin{aligned}
P(\bigcup A_i) &= \frac{1}{M!}\left(M(M-1)! - \binom{M}{2}(M-2)! + \binom{M}{3}(M-3)! - \cdots\right) && (4) \\
&= \sum_{j=1}^{N} (-1)^{j+1} \binom{M}{j} \frac{(M-j)!}{M!} && (5) \\
&= \sum_{j=1}^{N} (-1)^{j+1} \frac{1}{j!}\ . && (6)
\end{aligned}
$$

Notice that the right hand side quickly tends to $1 - e^{-1} \approx 0.62$ with M. Let us now generalize the idea in the example to arbitrary sets.

We introduce some notation first. When we write

$$
\sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} P(A_{i_1} A_{i_2}, \cdots, A_{i_k})
$$

the sum is over all possible $k-$tuples out of the given n sets. In particular, there are $\binom{n}{k}$ terms in the above summation. Think that there are n balls and you are picking k from them. Each different pick corresponds to a term in the above summation. For example, when $k = 2$ this corresponds to summation over all possible pairs. Those who like to think in terms of matrices, here is a simple illustration for $n = 5$ and $k = 2$.

$$
\begin{bmatrix}
 & A_1 & A_2 & A_3 & A_4 & A_5 \\
A_1 & 0 & 1 & 1 & 1 & 1 \\
A_2 & 0 & 0 & 1 & 1 & 1 \\
A_3 & 0 & 0 & 0 & 1 & 1 \\
A_4 & 0 & 0 & 0 & 0 & 1 \\
A_5 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

In here, where ever you find an entry 1, take the intersection of the corresponding row-head[1] and column-head and accumulate the probabilities. Once we understand this notation, we can write an important property, also known as **inclusion-exclusion formula**.

[1]first entry in that row

Property 3

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_iA_j) + \sum_{1\le i<j<k\le n} P(A_iA_jA_k) + \cdots + (-1)^{n+1}P(A_1A_2,\cdots,A_n) \quad (7)$$

Proof: Notice that we can concisely write this as,

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{j=1}^{n}(-1)^{j+1} \sum_{1\le i_1<\cdots<i_j\le n} P(A_{i_1},\cdots,A_{i_j}).$$

Expand a few terms and see for yourself that this is true. For example, $n = 3$ will give

$$P(A_1 \bigcup A_2 \bigcup A_3) = P(A_1) + P(A_2) + P(A_3) - P(AB) - P(AC) - P(BC) + P(ABC),$$

a result which can be derived by repeatedly applying (2). While large n looks a little messy, we can tackle this with the principle of induction[2]. The principle is simple, suppose a property which is true for a given n can be shown to hold for $n+1$. Then, essentially we are done by demonstrating the said property for the minimum possible value of n. Here, the base case is $n = 2$, which was already shown in (2). So assume that the inclusion-exclusion theorem is true for some $n \ge 2$. Denoting $A_i \bigcap A_{n+1}$ as $\hat{A}_i$ and inducting,

$$\begin{aligned}
P(\bigcup_{i=1}^{n+1} A_i) &= P(\bigcup_{i=1}^{n} A_i \bigcup A_{n+1}) && (8)\\
&= P(\bigcup_{i=1}^{n} A_i) + P(A_{n+1}) - P(\bigcup_{i=1}^{n} A_i \bigcap A_{n+1}) && (9)\\
&= P(\bigcup_{i=1}^{n} A_i) + P(A_{n+1}) - P(\bigcup_{i=1}^{n} \hat{A}_i) && (10)\\
&= \sum_{i=1}^{n} P(A_i) + \sum_{j\ge 2}(-1)^{j+1} \sum_{1\le i_1<\cdots<i_j\le n} P(A_{i_1},\cdots,A_{i_j}) + P(A_{n+1}) - P(\bigcup_{i=1}^{n} \hat{A}_i) && (11)\\
&= \sum_{i=1}^{n+1} P(A_i) + \sum_{j\ge 2}(-1)^{j+1} \sum_{1\le i_1<\cdots<i_j\le n} P(A_{i_1},\cdots,A_{i_j}) - P(\bigcup_{i=1}^{n} \hat{A}_i) && (12)
\end{aligned}$$

Notice that the first term on the RHS is part of the desired term. Let us look the sum of the last two terms, say terms II and III.

$$II + III = \sum_{j\ge 2}(-1)^{j+1} \sum_{1\le i_1<\cdots<i_j\le n} P(A_{i_1},\cdots,A_{i_j}) + \sum_{j\ge 1}(-1)^{j} \sum_{1\le i_1<\cdots<i_j\le n} P(\hat{A}_{i_1},\cdots,\hat{A}_{i_j}) \quad (13)$$

While summing over j, the first term of II will have $j = 2$, while that of III will have $j = 1$, we can combine these terms to obtain

$$\begin{aligned}
- \sum_{1\le i_1<i_2\le n} P(A_{i_1}A_{i_2}) - \sum_{1\le i_1\le n} P(\hat{A}_i) &= - \sum_{1\le i_1<i_2\le n} P(A_{i_1}A_{i_2}) - \sum_{1\le i_1\le n} P(A_iA_{n+1}) && (14)\\
&= - \sum_{1\le i_1<i_2\le n+1} P(A_{i_1}A_{i_2}), && (15)
\end{aligned}$$

where the last summation runs up to $n+1$. This gives the second term in the desired expression. Similarly, combining term by term we will get each of the parts in the formula. After pairwise combining, one entry remains, which is

$$(-1)^n P(A_1,\cdots,A_{n+1}) = (-1)^{n+2} P(A_1,\cdots,A_{n+1}),$$

thus verifying that the formula (7) is true for $n+1$. ∎

[2]which dates back to the BCs: perhaps the Greeks used it first

Exercise 1 *Verify the inclusion-exclusion formula for terms containing* $3-$*tuples (i.e. the third term in the expansion* (7)*).*

Example 2 *Show that* $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$.

Solution: While this can be shown based on the principles above, an easier way is to use axiom-II. In particular,

$$\bigcup_{i=1}^{n} A_i = A_1 \bigcup A_1^c A_2 \bigcup A_1^c A_2^c A_3 \bigcup \ldots \bigcup \left(\bigcap_{i=1}^{n-1} A_i^c A_n \right).$$

By applying axiom II, it is clear that

$$\begin{aligned} P(\bigcup_{i=1}^{n} A_i) &\leq P(A_1) + P(A_1^c A_2) + P(A_1^c A_2^c A_3) + \ldots + P(\bigcap_{i=1}^{n-1} A_i^c A_n) \\ &\leq \sum_{i=1}^{n} P(A_i) \end{aligned}$$

■

Exercise 2 *Can you extend the formula in Example 2 to countable unions?*

2 Constructing Probability Spaces

We will use our axioms and set properties to paint a larger picture now, i.e. to construct probability spaces in an unambiguous manner. The axioms and properties are kind of the building blocks and glue, magnificent structures can be constructed by proper placement. Here is an algorithmic view of what we are about to do.

1. Have the set of axioms ready as a toolbox.
2. Take an initial seed C of events. The seed is some set of events for which one can unambiguously associate probability values. Many a times this set is a very natural or intuitive choice.
3. Apply the axioms to extend the probability measure defined on the seed to all countable unions and complements of these events.

In the above, the probability values are specified (associated) so as to truthfully model a system we wish to examine. We do not overly worry about the statistical learning and inference done in the past to arrive at a meaningful probability association. Again, what we suggested above is a recipe to construct a probability space. The taste of our preparation will depend on the ingradients used. In particular, a proper choice of the raw ingradient, i.e. seed sets will turn out of atmost importance to get a consistent definition which is stable over allowed operations.

The probability space construction that we describe is similar to building some shape with LEGOs. In LEGO, one needs to be alert about a few things. Firstly, if the LEGO has too many blocks, all the time will be spend in finding the correct pieces, can be a big problem when the blocks are *countably infinite* or more. In fact, the idea of the seed set was to tackle this. We kind of choose the *blocks* which appear *useful* in our construction and start with those. The second problem occurs when some necessary shape cannot be made out of the seed set, in other words, an inadequate seed. So our seed set has to be sufficiently large to accommodate our designs, but not too large that we fail to manage it effectively. One may feel that what we talk is more *management* than mathematics, with lot of dangling English words. But this is the real story behind it, and mathematics plugs the gaps and presents these in a coherent manner.