Axioms of probability (Cont)

Math 477

September 8, 2014

1 Some elementary propositions

Some consequences of the axioms of probilities are as followed:(Note that they are "natural" in the sense that one would expect them to be true anway, but we can give rigorous proofs based on the axioms)

1.

$$P(E^c) = 1 - P(E).$$

Proof: Since $E^c \cup E = S$, $E^c \cap E = \emptyset$,

$$1 = P(S) = P(E^c \cup E) = P(E^c) + P(E).$$

2. If $E \subseteq F$ then $P(E) \leq P(F)$.

Proof: We use the notation $F \backslash E := F \cap E^c$ to denote the members of F that is not in E. Then $E \cap F \backslash E = \emptyset, E \cup F \backslash E = F$ and $P(F \backslash E) \geq 0$. Thus

$$P(F) = P(E) + P(F \backslash E) \geq P(E).$$

3. The (simple) inclusion - exclusion principle:

$$P(E \cup F) = P(E) + P(F) - P(EF).$$

This principle generalizes axiom 3. Here we do not require E and F to be mutually exclusive. If they are, then $P(EF) = 0$ and we recover axiom 3.

Proof: Note that $F \backslash EF = F \backslash E$ and thus $E \cap (F \backslash EF) = \emptyset, E \cup (F \backslash EF) = E \cup (F \backslash E) = E \cup F$. Therefore,

$$P(E \cup F) = P(E \cup (F \backslash EF)) = P(E) + P(F \backslash EF).$$

Again recognize that $(F\backslash EF) \cap EF = \emptyset$ and $EF \cup (F\backslash EF) = F$. Thus

$$P(F) = P(F\backslash EF) + P(EF).$$

Putting these together we get the proposition.

2 The general inclusion exclusion principle

Generally we have

$$\begin{aligned} P(E_1 \cup E_2 \cup \cdots \cup E_n) &= \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) \\ &+ (-1)^{r+1} \sum_{i_1<i_2<\cdots<i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r}) + \cdots \\ &+ (-1)^{n+1} P(E_1E_2\cdots E_n), \end{aligned}$$

or more compactly

$$P(\cup_{i=1}^{n} E_i) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<i_2<\cdots<i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r}).$$

Remark: The summation $\sum_{i_1<i_2<\cdots<i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r})$ is taken over all possible $\binom{n}{r}$ subsets of size r of the set $\{1, 2, \cdots, n\}$. The notation $i_1 < i_2 < \cdots < i_r$ is to make sure we *do not repeat* the summation over a subset after we have selected it, which serves the same purpose as a combination. For example, if $n = 5, r = 3$, then we only sum over the subset $\{1, 2, 4\}$ but not the subset $\{4, 1, 2\}$ or $\{2, 4, 1\}$ etc., as dictated by the requirement that $1 < 2 < 4$ or by the understanding that we are dealing with a combination here.

Proof: We proceed by induction. The case $n = 2$ as been proven above for the simple principle. Generally suppose the principle is true for n. We show it is true for $n + 1$. But note that we can write $\cup_{i=1}^{n+1} E_i = \cup_{i=1}^{n} E_i \cup E_{i+1}$. Thus

$$\begin{aligned} P(\cup_{i=1}^{n+1} E_i) &= P(\cup_{i=1}^{n} E_i) + P(E_{i+1}) - P(\cup_{i=1}^{n} E_i \cap E_{n+1}) \\ &= P(\cup_{i=1}^{n} E_i) + P(E_{i+1}) - P(\cup_{i=1}^{n} (E_i \cap E_{n+1})) \end{aligned}$$

by the base case $n = 2$. Apply the induction hypothesis on $P(\cup_{i=1}^n E_i)$ we have

$$\begin{aligned} P(E_1 \cup E_2 \cup \cdots \cup E_n) &= \sum_{i=1}^n E_i - \sum_{i_1 < i_2 \le n} P(E_{i_1} E_{i_2}) \\ &+ (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r \le n} P(E_{i_1} E_{i_2} \cdots E_{i_r}) + \cdots \\ &+ (-1)^{n+1} P(E_1 E_2 \cdots E_n), \end{aligned}$$

Apply the induction hypothesis on $P(\cup_{i=1}^n (E_i \cap E_{n+1})$ we have

$$\begin{aligned} -P(\cup_{i=1}^n (E_i \cap E_{n+1}) &= -\sum_{i=1}^n P(E_i \cap E_{n+1}) + \sum_{i_1 < i_2 \le n} P(E_{i_1} E_{i_2} E_{n+1}) \\ &+ (-1)^{r+2} \sum_{i_1 < i_2 < \cdots < i_r \le n} P(E_{i_1} E_{i_2} \cdots E_{i_r} E_{n+1}) + \cdots \\ &+ (-1)^{n+2} P(E_1 E_2 \cdots E_n E_{n+1}), \end{aligned}$$

(note the sign change on the RHS).

But then you see all the signs match up right, for example

$$\begin{aligned} -\sum_{i_1 < i_2 \le n} P(E_{i_1} E_{i_2} E_{i_3}) - \sum_{i=1}^n P(E_i \cap E_{n+1}) &= -\sum_{i_1 < i_2 \le n+1} P(E_{i_1} E_{i_2}) \\ \sum_{i_1 < i_2 < i_3 \le n} P(E_{i_1} E_{i_2}) + \sum_{i_1 < i_2 \le n} P(E_{i_1} E_{i_2} E_{n+1}) &= \sum_{i_1 < i_2 < i_3 \le n+1} P(E_{i_1} E_{i_2} E_{i_3}) \cdots \end{aligned}$$

so we have verified the formula for the case $n + 1$.

2.1 Consequence - Bounds for probability of union of events

A rather useful consequence of the general inclusion - exclusion principle is to provide bounds for union of events. Specifically:

$$\begin{aligned} P(\cup_{i=1}^n E_i) &\le \sum_{i=1}^n P(E_i) \\ P(\cup_{i=1}^n E_i) &\ge \sum_{i=1}^n P(E_i) - \sum_{i_1 < i_2 \le n} P(E_{i_1} E_{i_2}) \\ P(\cup_{i=1}^n E_i) &\le \sum_{i=1}^n P(E_i) - \sum_{i_1 < i_2 \le n} P(E_{i_1} E_{i_2}) + \sum_{i_1 < i_2 < \cdots < i_3 \le n} P(E_{i_1} E_{i_2} E_{i_3}) \cdots \end{aligned}$$

This is very similar to the oscillating series you learned in Calculus 2, for example $x = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n}$. There you showed that $x \leq 1, x \geq 1-\frac{1}{2}, x \leq 1-\frac{1}{2}+\frac{1}{3}$ and so forth. The only difference is here we deal with a finite sum; but one can imagine that similar bounds hold for the expression $P(\cup_{i=1}^{\infty} E_i)$, from which we indeed get an infinite oscillating sum.

One can prove these bounds by induction and use the inclusion-exclusion principle for the base case. We'll prove the first 2 inequalities as examples.

Ex1: Prove

$$P(\cup_{i=1}^{n} E_i) \leq \sum_{i=1}^{n} P(E_i).$$

For $n = 2$, since $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1E_2)$ the inequality holds. Assume the inequality holds for n. Then

$$P(\cup_{i=1}^{n+1} E_i) \leq P(\cup_{i=1}^{n} E_i) + P(E_{n+1}) \leq \sum_{i=1}^{n+1} P(E_i).$$

■

Ex2: Prove

$$P(\cup_{i=1}^{n} E_i) \quad \geq \quad \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2\leq n} P(E_{i_1}E_{i_2}).$$

For $n = 3$, since

$$P(\cup_{i=1}^{3} E_i) = \sum_{i=1}^{3} P(E_i) - \sum_{i_1<i_2\leq 3} P(E_{i_1}E_{i_2}) + P(E_1E_2E_3),$$

the inequality holds.

Assume the inequality holds for n. Then

$$P(\cup_{i=1}^{n+1} E_i) \quad \geq \quad P(\cup_{i=1}^{n} E_i) + P(E_{n+1}) - P(\big(\cup_{i=1}^{n} E_i\big)E_{n+1}).$$

We have

$$P(\cup_{i=1}^{n} E_i) \geq \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2\leq n} P(E_{i_1}E_{i_2})$$

by the induction hypothesis,

and

$$
\begin{aligned}
P\big(\big(\cup_{i=1}^{n} E_i\big)E_{n+1}\big) &= P\big(\cup_{i=1}^{n}\big(E_iE_{n+1}\big)\big) \\
&\leq \sum_{i=1}^{n} P(E_iE_{n+1}),
\end{aligned}
$$

by the first inequality we proved. Putting these two facts together we get that the second inequality holds for the case $n+1$. ■

2.2 Interpretation of the bounds

These bounds for the probability of the union of the events have interpretations that corresponds to our intuition. For example, the first one says that if we have n event (say $n = 10$) each has *very small probability* of happening (relative to n, say $P(E_i) = 10^{-4}, i = 1, \cdots, 10$), then *the probability of any of them happening is also small* (since $P(\cup_{i=1}^{10} E_i) \leq 10^{-3}$ by the first inequality). In other words, the rarity of individual events translate to the rarity of the collection of events as long as there are not too many of them.

The second one says that if we have n event, again say $n = 10$, such that *the sum of their probability is large* (say $\sum_{i=1}^{10} E_i = 1$). Then as long as, the probability of *any of two of them happening together is small* say $P(E_iE_j) = 10^{-4}$, then *the probability of any of them happening is also large* (more than $1 - \frac{45}{10^4}$).

3 Sample spaces having equally likely outcomes

3.1 Assigning probabilities

Consider the experiment of tossing a die. Suppose we have no reason to suspect that the die is loaded, then we must conclude that each outcome is equally likely. That is

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}).$$

Since these outcomes are mutualyl exclusive and their union is the sample space S, by the axioms of probability, we have

$$1 = P(S) = \sum_{i=1}^{6} P(\{i\}) = 6P(\{i\}), i = 1, \cdots, 6.$$

Thus it is *reasonable* to assign

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = \frac{1}{6}.$$

This is the basis for assigning $\frac{1}{2} = P(\{H\}) = P(\{T\})$ if we believe a coin is fair or $\frac{1}{52}$ to be the probability of drawing a specific card in a deck of card, for example.

More abstractly, for events that are composed of elementary outcomes, as long as we assume each out come is equally likely, we can conclude

$$P(E) = \frac{n(E)}{n(S)},$$

where $n(E)$ denotes the number of outcomes in event E and $n(S)$ denotes the total number of outcomes in the experiment. You'll quickly see how the combinatorics techniques we develop in the previous lecture comes to play in computing these probabilities.

3.2 A basic example

Ex1: Suppose we toss a coin twice. Compute the probabilities of : a) The first toss is H b) Both tosses are T c) We get a head and a tail, in either order.

Note: Intuitively, you'll give the answer as a) 1/2 b) 1/4 c) 1/2. When asked why, you'll probably say something about the coin being fair, and say for question b) the two tosses are independent thus the probability of 2 tails is $1/2 \times 1/2 = 1/4$. I hope you see that these answers lack rigorous justification. Plus we did not discuss the notion of independence. And the pseudo answer given above is a perfect example of mixing the order of arguments. After we cover independence of events, you'll see that BECAUSE that the probability of both tosses are T is 1/4 (and similar computations) we can CONCLUDE that the two tosses are independent, not the other way around.

The rigorous ways to answer these question is to consider the sample space of tossing a coin twice. It is

$$S = \Big\{\{HT\}, \{TH\}, \{TT\}, \{HH\}\Big\}$$

and we assume each outcome is equally likely (again, on the basis of our belief about the coin being fair, thus no outcome should be more likely than other - You may even say that when we say a coin is fair, what we have in mind is exactly equally outcome.) Thus

$$P(\{HT\}) = P(\{TH\}) = P(\{TT\}) = P(\{HH\}) = \frac{1}{4}.$$

Now we can easily answer

$$\begin{aligned} P(\text{First toss is head}) &= P(\{HT\} \cup \{HH\}) = \frac{2}{4} = \frac{1}{4} \\ P(\{TT\}) &= \frac{1}{4} \\ P(\text{Head and tail, either order}) &= P(\{HT\} \cup \{TH\}) = \frac{2}{4} = \frac{1}{2}. \end{aligned}$$

3.3 Switching sample space for equally likely events

Consider the following question: Suppose we toss a coin 4 times and all outcomes are equally likely. What is the probability that the third coin is H?

There are 2 ways to answer. You could divide the number of outcomes with H in the 3rd toss (8 total) by the number of total outcomes (16) to get 1/2, a rather unsurprising answer ! Or you could say there is only two possibilities for the 3rd toss, H or T and these are equally likely, so the probability must be 1/2. Of course the 2nd answer is correct, but note that implicitly there you have *changed the sample space*. Your sample space consists only of the outcomes of the 3rd toss whereas the original sample space has the outcomes of all 4 tosses. This technique of swithching sample space (which many of us would employ without realizing it) is of course valid, in computing probability of equally likely event. The important conclusion to derive here is that suppose we run n experiments and all of the outcomes are equally likely, then the outcomes of a particular experiments among the n are also equally likely.

Here's another example where we employ this technique in a less obvious manner. Suppose there are n people in the room and they toss their hats in a box, mix those up and randomly draw out the hats. What is the probability that none of them get their correct hat?

Ans: We will give the answer to this question later. For now, we're interested in computing the following: let E_i be the event that the i-th person got his correct hat. We want to compute $P(E_i), P(E_iE_j), P(E_iE_jE_k)\cdots$.

When we say people randomly draw out the hats, it also means all outcomes are equally likely.

Observe that it is equally likely for the ith person to choose any hats among the n. Thus $P(E_i) = \frac{1}{n}$. There are different ways to compute $P(E_i)$, with different interpretations. Here's another good observation: $P(E_i) = P(E_j)$ because there should be no difference in the chance of the ith or the jth person getting their correct hats. However, we DO NOT know that $\sum_{i=1}^{n} P(E_i) = 1$. So we CANNOT draw any conclusion about $P(E_i)$ from this observation.

To compute $P(E_iE_j)$ note that there are $\binom{n}{2}2! = n(n-1)$ ways for 2 person to select 2 hats, and each of these is equally likely. Thus $P(E_iE_j) = \frac{1}{n(n-1)}$.

Again observe that $P(E_iE_j) = P(E_jE_k)$ and there are $\binom{n}{2}$ pairs of these. But note how $P(E_iE_j) \neq \frac{1}{\binom{n}{2}} = \frac{2}{n(n-1)}$. It is because we DO NOT have $\sum_{i_1<i_2} P(E_{i_1 i_2}) = 1$.

Similarly you can see that $P(E_iE_jE_k) = \frac{1}{n(n-1)(n-2)}$ etc.

Now we can answer the original question. Note that

$$P(\text{no one get their correct hat}) = P(E_1^cE_2^c \cdots E_n^c) = 1 - P(E_1 \cup E_2 \cup \cdots \cup E_n).$$

By the inclusion exclusion principle

$$\begin{aligned} P(E_1 \cup E_2 \cup \cdots \cup E_n) &= \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) \\ &\quad + \sum_{i_1<i_2<i_3} P(E_{i_1}E_{i_2}E_{i_3}) + \cdots + (-1)^{n+1}P(E_1E_2\cdots E_n) \\ &= 1 - \binom{n}{2}\frac{1}{2!\binom{n}{2}} + \binom{n}{3}\frac{1}{3!\binom{n}{3}} + \cdots + (-1)^{n+1}\frac{1}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} + \cdots + (-1)^{n+1}\frac{1}{n!}, \end{aligned}$$

where we have used the observation that $P(E_iE_j) = P(E_jE_k)$ etc. Thus

$$P(\text{no one get their correct hat}) = \sum_{k=2}^{n} (-1)^k \frac{1}{k!}.$$