**ORIE 5530: Modeling Under Uncertainty**

# Lecture 3 (Basic Probability)

*Professor Mark S. Squillante,*
*Adapted from Professor Itai Gurvich's original notes*

So we said an event is a subset of the sample space. In essence, an event is a collection of possible outcomes. We say that an event happened if one of the outcomes in it happened. If $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, then $A = \{2, 4, 6\}$ corresponds to the event that we got an even outcome – one of the numbers $2, 4, 6$ was the outcome of the experiment.

So you could be slightly formal and define $\Omega$ to be the collection of all subsets of the sample space $\mathcal{S}$. For example if $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, $\Omega$ includes individual-outcome events like $\{1\}$ (this is the event the outcome was 1 on the die), it includes events of the form $\{3, 5\}$ (that we got 3 or 5), etc.; and all possible combinations up to and including the event $\{1, 2, 3, 4, 5, 6\}$, which is the whole space.

## 3.1 Probability on events

Now, once you listed the possible events you should be able to assign probabilities to them. There is a probability 1 that I get some number in the set $[6] := \{1, \ldots, 6\}$ when I roll the dice, so $\mathbb{P}(\mathcal{S}) = 1$. The event $A = \{2, 4, 6\} =$ "even outcome" has probability $1/2$ (if the die is fair, etc.).

It also is clear that a probability (likelihood) will always be a real number between 0 and 1.

Finally, it is intuitive to us (and correctly so) that the probability of $\{2, 4, 6\}$ in the roll of a die is the sum of the probabilities of getting 2, 4 and 6. This is because these are exclusive events (if we get 2 on the roll of a dice, that's it. We did not get a 4 or a 6).

So we want to be able to say that if two events are exclusive (meaning they cannot happen together – $A$ and $B$ are exclusive means $A \bigcap B = \emptyset$) then the probability that one (or both) of them happens should be the sum of the probabilities.

In summary, a probability is a function from $\Omega$ to $[0, 1]$ that has the properties:

i. $0 \leq \mathbb{P}(A) \leq 1$ for any event $A$;

ii. $\mathbb{P}(\mathcal{S}) = 1$; and

iii. For any collection of mutually exclusive events $A_1, A_2, \ldots, A_n$ (i.e., $A_i \bigcap A_j = \emptyset$, $\forall i, j$),

$$\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i).$$

An event $A$ and its complement $A^c$ ("not A") are obviously exclusive and their union is $\mathcal{S}$ so that $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$ (hence, $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$).

If two events $A$ and $B$ are not exclusive then

$$\mathbb{P}(A \bigcup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \bigcap B)$$

so that we always have the inequality

$$\mathbb{P}(A \bigcup B) \leq \mathbb{P}(A) + \mathbb{P}(B).$$

## 3.2 Conditional probability on events

The basic identity here is that (we use | for conditioning)

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \bigcap A)}{\mathbb{P}(A)}. \tag{*}$$

Notice that this makes sense only if $\mathbb{P}(A) > 0$. (This make intuitive sense – If $A$ is an event that has 0 probability, it will never happen so we will never care about what is the likelihood of $B$ given $A$). Observe that $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$ unless $\mathbb{P}(A) = \mathbb{P}(B)$.

Now, from here follows a characterization of what it means for two events to be independent. Conceptually, you want to say that $B$ and $A$ are independent if knowing that $A$ happened does not change the likelihood I assign to $B$. That is,

$$\mathbb{P}(B|A) = \mathbb{P}(B).$$

If you plug this to the identity above (replace the left hand side $\mathbb{P}(B|A)$ with $\mathbb{P}(B)$) you get $\mathbb{P}(B) = \frac{\mathbb{P}(B \bigcap A)}{\mathbb{P}(A)}$ or, as we mathematically think of independence, we have

$$\mathbb{P}(B \bigcap A) = \mathbb{P}(B)\mathbb{P}(A). \tag{independence}$$

If we have multiple events, say, $A_1, \ldots, A_m$, we say that they are all independent of each other if any sub-collection satisfies the above. That is, we definitely need that $\mathbb{P}(A_i \bigcap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ for any $i \neq j$, but we also need $\mathbb{P}(A_i \bigcap A_j \bigcap A_l) = \mathbb{P}(A_i)\mathbb{P}(A_j)\mathbb{P}(A_l)$ for any $i \neq j \neq l$, etc. In general, for any $n$ and any sub-collection $A_{i_1}, \ldots, A_{i_n}$ of size $n$ with $i_1 \neq i_2 \neq \ldots \neq i_n$, we need that

$$\mathbb{P}(\bigcap_{k=1}^{n} A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \ldots \mathbb{P}(A_{i_{n-1}})\mathbb{P}(A_{i_n}).$$

**Useful fact:** If you take a partition of the world into $B$ and $B^c$ (not B), it is obvoius you can write a set $A$ as the union of its intersection with $B$ (what is in B and in A) and an intersection with $B^c$ (what is in $A$ but not in $B$). That is,

$$A = (A \bigcap B) \bigcup (A \bigcap B^c).$$

Because $B$ and $B^c$ are exclusive so are $(A \bigcap B)$ and $(A \bigcap B^c)$ and we have by the basic rule (iii) of probability that

$$\mathbb{P}(A) = \mathbb{P}(A \bigcap B) + \mathbb{P}(A \bigcap B^c).$$

Using (*) we can re-write this as

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c). \tag{**}$$

More generally if $B_1, \ldots, B_m$ is some way to divide the world of outcomes (meaning it is a *partition* of $\mathcal{S}$: $B_i \bigcap B_j = \emptyset$ for all $i \neq j$ and $\bigcup_{i=1}^m B_i = \mathcal{S}$) then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \bigcap B_i).$$

We used this fact in an example we had in class about the conditional likelihood of two boys.

Notice that because of the basic rule (*) we can re-write this as

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Replacing $\mathbb{P}(B \bigcap A) = \mathbb{P}(A|B)\mathbb{P}(B)$ and replacing $\mathbb{P}(A)$ in the denominator using (**) we get the simplest version of Bayes rule:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}.$$

The beauty of Bayes rule is in how it reverses the conditioning. We are interested in $B$ conditional on $A$ and we evaluate this by conditioning $A$ and $B$ and $B^c$.

More generally, if $B_1, \ldots, B_n$ are a partition of the set $\mathcal{S}$, then for each $j = 1, \ldots, n$ we have

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}.$$