Lecture 4: Notions of independence and applications

Rajat Mittal

IIT Kanpur

First, let's revisit some of the concepts introduced in conditional probability. There are many situations where an event affects the probability of another event. Your chances of winning a lottery might severely decrease or increase given that the winning number is even. Your chances of coming to the class might depend on whether your friend is in the class or not.

To capture these situations, we defined *conditional probability.* Given two events A, B, the conditional probability of A given B is defined by,

$$P\left(\frac{A}{B}\right) := \frac{P(A \cap B)}{P(B)}.$$

We also defined the related concept of independent events.

1 Independent events

Two events A, B are said to be independent if,

$$P(A \cap B) = P(A)P(B).$$

Again, A and B being independent means, occurrence of A has no effect on occurrence of B. We keep this definition because it takes care of the case $P(B) = 0$, as compared to the equivalent (almost) definition, $P(A|B) = P(A)$.

Note 1. Both these definitions are pretty intuitive.

If there is no relation between A and B (independent), we expect the probability of $P(A \cap B)$ to be the product of the two individual probabilities. If the probability is lower, they are said to be *negatively correlated*,

$$P(A \cap B) < P(A)P(B).$$

Similarly, two events are said to be *positively correlated* if,

$$P(A \cap B) > P(A)P(B).$$

Intuitively, negative correlation means event A reduces the probability of event B. On the same lines, positive correlation means occurrence of A increases the chance of event B.

Exercise 1. What is the relationship between independence and conditional probability of two events?

Many a times it might be clear from the context that two events are independent. For example,

- If we toss two coins, we get a head on first toss is independent of getting a tails on the second toss. More generally, any event *totally dependent* on first toss will be independent of any other event *totally dependent* on second toss. This is sometimes clearly stated by stating, a coin is tossed twice independently.
- If we pick two cards from a deck sequentially, with replacement, the event that we get hearts in first pick is independent to that we get ace on the second pick. If we don't put the first card back before picking the second, without replacement, they might not be independent event.
- As mentioned before, two disjoint events (non-zero probability) are always dependent.

In some cases, it might not be this clear. The best way is to check the definition of independence. Suppose you toss two coins, let A be the event that *the first toss is head* and B be the event that *the two outcomes are same.* Calculating,

$$P(A) = 1/2,\ P(B) = 1/2,\ P(A \cap B) = 1/4.$$

So, the two events are independent.

Let's look at another scenario for independence in probability. There was a survey conducted by the Health dept. in a hospital (with Asthma and Diabetes patients), it found that people who had diabetes did not have Asthma with higher probability as compared to the general population. This suggests that people who have Asthma have less likelihood of getting Diabetes.

It turns out that even if having Asthma and Diabetes are independent of each other, there will be a negative relation between Asthma and Diabetic patients in a hospital.

In other words, suppose A, B are two independent events. They will not be mutually independent if we consider the conditional probabilities given $A \cup B$. That is, events $\frac{A}{A \cup B}$ and $\frac{B}{A \cup B}$ will be negatively related even if A, B are independent. This is known as *Berkson's Paradox.*

To make it more quantitative, consider a sample of 1000 balls. We know that 100 of them are red, 50 of them are shiny and 5 of them are red and shiny. The probability of being shiny is $1/20$ and also the probability of a red ball being shiny is $5/100 = 1/20$. Hence being red and being shiny are independent.

Say, we pick only the balls which are red and/or shiny. Then the probability that a ball is shiny is close to $1/3$ but a red ball being shiny remains at $1/20$. This will show that the a red ball is mostly not shiny.

Exercise 2. Say A be the event that the ball is shiny and B being the event that the ball is red. Prove that $Pr\left(\frac{A}{A \cup B}\right) > Pr(A)$. Convince yourself that this is the reason why events A and B seem to be negatively related.

Exercise 3. Convince yourself that the same thing happened in the above example (survey in a hospital).

1.1 Independence for family of events

The criteria for independence becomes more involved if there are more than two events. Consider a family of events, $\{A_i\}_{i \in I}$, each being indexed by i in some index set I.

The simplest notion of independence is called *pairwise independence*, implying that any two pair of events in the family are independent.

$$P(A_i \cap A_j) = P(A_i)P(A_j) \ \ \forall i \neq j \in I.$$

It can be strengthened to *k-wise independence*, it holds iff

$$P(\cap_{j \in J} A_j) = \mathit{\Pi}_{j \in J} P(A_j) \ \ \forall J \subseteq I, |J| \leq k.$$

Exercise 4. Show that pairwise independence is same as 2-wise independence.

The strongest notion is of *mutual independence*, which says that the events are $|I|$-wise independent.

$$P(\cap_{j \in J} A_j) = \mathit{\Pi}_{j \in J} P(A_j) \ \ \forall J \subseteq I.$$

Exercise 5. Convince yourself that k-wise independence implies l-wise independence if $l \leq k$.

It might seem that all these notions are equivalent. Consider the following example which shows that these definitions are different.

We toss an unbiased coin twice. Define A to be the event that first toss is same as the second toss. Define B, C to be events that first (second) toss is head respectively. You will show in the assignment that these three events are pairwise independent but not mutually (3-wise) independent.

2 Independent random variables

We can similarly define independence of random variables. Remember, a random variable X is a mapping from sample space to real numbers, $X : \Omega \to \mathbb{R}$. The conditional distribution of X given an event B was defined to be,

$$P_{X|B}(x) = P(X = x|B).$$

Using these definitions, two random variables X and Y are independent iff,

$$P(X = x \cap Y = y) = P(X = x)P(Y = y) \ \ \forall x, y.$$

Here x, y are in the range of X, Y respectively.

Notice the difference between independence of events and random variables. For two random variables to be independent, all concerned events defined by $X = s$ kind of events should be independent. Intuitively, two random variables are independent iff the value of one does not give any indication about the value of other random variable.

Again, like events, independence can be generalized to family of random variables $\{X_i\}_{i \in I}$. We start with *pairwise independence*, implying that any two pair of random variables in the family are independent.

$$P((X_i = x_i) \cap (X_j = x_j)) = P(X_i = x_i)P(X_j = x_j) \ \ \forall i, j \in I.$$

Here, x_i, x_j are all possible elements from the range of X_i and X_j respectively.

It can be strengthened to *k-wise independence*, it holds iff

$$P(\cap_{j \in J}(X_j = x_j)) = \Pi_{j \in J} P(X_j = x_j) \ \ \forall J \subseteq I, |J| \leq k.$$

The last extension is to *mutual independence* and is very similar, which says that the events are $|I|$-wise independent.

$$P(\cap_{j \in J}(X_j = x_j)) = \Pi_{j \in J} P(X_j = x_j) \ \ \forall J \subseteq I.$$

Exercise 6. Convince yourself that k-wise independence implies l-wise independence for random variables if $l \leq k$.

Exercise 7. Is pairwise independence same as mutual independence, construct a counterexample.

You might be confused about all these definitions and their use. We will see an application in the next section.

2.1 Hash functions

We will see an application of independence in a concept called *hashing*. It is taken from Victor Shoup's book on Computational Number Theory.

Hashing is the practice of mapping a set bigger set S to a potentially smaller set T using random keys R. The objective is that an element $s \in S$ goes to an element in T uniformly at random and the number of collisions are small.

One application could be to store lots of elements, when it is know that all elements belong to some big set S. To reduce the size of the storage, we can store the hashed image (in T) instead of the actual element. This will require much less storage size.

Exercise 8. Read about hashing from the web.

Let us define it formally.

Given three sets R, S, T, a *hashing* from S to T is a collection of functions $\Phi_r : S \to T$, for each $r \in R$. We could think of elements of R as keys which allow us to know the mapping from S to T. We can think of $\Phi_R(s)$ for an $s \in S$ as a random variable where r is picked uniformly. Again, we would want $\Phi_R(s)$ to be distributed uniformly in T. Ideally, knowledge of $\Phi_R(s)$ should not give any information about $\Phi_R(s')$ for any $s' \neq s$.

To make these requirements concrete, a family of hash functions (hashing) is called *pairwise independent* iff

- Random variables $\{\Phi_R(s)\}_{s \in S}$ are pairwise independent.
- $\Phi_R(s)$ is uniformly distributed over T for every s.

You will show in the assignment that a hashing is pairwise independent iff

$$P(\Phi_R(s) = t \cap \Phi_R(s') = t') = \frac{1}{|T|^2} \ \ \forall s, s' \in S(s \neq s'); t, t' \in T.$$

As expected, pairwise independent hashing implies that no s favors any particular area of T, and given knowledge of one hash (value of $\Phi_r(s)$) we have no clue about any other $\Phi_r(s')$ (assuming, we don't know r). There are many ways known to construct such hash functions. We will now take a look at an application of such possible constructions in cryptography.

Assume that two parties, let us call them Raj and Simran, want to communicate a message. Though, Simran wants to make sure that the message is actually from Raj and not transmitted by Amrish. Hashing provides a solution to this age old problem, though with two assumptions.

- Raj and Simran share a secret key.
- Simran is ready to accept a small probability of error.

The solution is quite natural given a family of hash functions.

Exercise 9. Why don't you think about a possible protocol for this problem using hashing.

So, as hinted above, assume that Raj and Simran know of a hashing scheme $R, S, T, \Phi_R(s)$. The set S will become the possible set of messages and R is the set of random keys. Raj and Simran will decide upon the key in advance, some $r \in R$. At the time of transmission, Raj will send $(s, \Phi_r(s))$ as a message to Simran.

We can model it as random variables $(X, Y = \Phi_R(X))$, where X is the random message and $\Phi_R(X)$ is the corresponding authentication tag. The message X is uniformly distributed in S and $r \in R$ is picked uniformly.

After receiving the message, Simran checks if the authentication tag Y is indeed the hash of X under the secret key R. Amrish (our adversary) fools Simran if he can find another pair (X', Y'), such that, $\Phi_r(X') = Y'$.

Ideally, ever after seeing the original message (X, Y), Amrish should not have any knowledge of key r. So, the probability that he can find X', Y' such that $\Phi_r(X') = Y'$ should be really small. We will explicitly calculate the probability of this event (say E).

$$\begin{aligned}
P(E) &= \sum_{s \in S} \sum_{t \in T} P(X = s \cap Y = t \cap E) \\
&= \sum_{s \in S} \sum_{t \in T} P((X = s) \cap (\Phi_R(s) = t) \cap (\Phi_R(X') = Y') \cap (X' \neq s)) \\
&= \sum_{s \in S} \sum_{t \in T} P(X = s) P((\Phi_R(s) = t) \cap (\Phi_R(X') = Y') \cap (X' \neq s)) \\
&\leq \sum_{s \in S} \sum_{t \in T} P(X = s) \frac{1}{|T|^2} \\
&= \frac{1}{|T|}
\end{aligned}$$

Here, third equality follows from the fact that the message and the key are picked independently.

Exercise 10. Why did the fourth inequality follow?

So, the probability of Amrish being successful is inversely proportional to the size of the image T. By picking a big enough T, Simran can ensure her failure probability to be small.

3 Assignment

Exercise 11. Given that events A, B are independent, prove that A^c, B^c are also independent. Hint: first prove that A, B^c are independent.

Exercise 12. Let X and Y be two independent random variables. Prove that,

$$E[XY] = E[X]E[Y].$$

Exercise 13. We toss an unbiased coin twice. Define A to be the event that first toss is same as the second toss. Define B, C to be events that first (second) toss is head respectively. Show that these three events are pairwise independent but not mutually (3-wise) independent.

Exercise 14. Show that a hashing is pairwise independent iff

$$P(\mathit{\Phi}_R(s) = t \cap \mathit{\Phi}_R(s') = t') = \frac{1}{|T|^2} \quad \forall s, s' \in S(s \neq s'); t, t' \in T.$$

References

1. D. Stirzaker. Elementary Probability. *Cambridge University Press*, 2003.
2. D. Kahneman. Thinking, Fast and Slow. *Farrar, Straus and Giroux*, 2011.