3.8 Independence of random variables

Unit 3: Discrete Random Variables

Adapted from Blitzstein-Hwang Chapter 3.

Just as we had the notion of independence of events, we can define independence of random variables. Intuitively, if two r.v.s X and Y are independent, then knowing the value of X gives no information about the value of Y, and vice versa.

Definition 3.8.1 (Independence of two r.v.s).

Random variables X and Y are said to be *independent* if

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y),$$

for all $x, y \in \mathbb{R}$. In the discrete case, this is equivalent to the condition

$$P(X = x, Y = y) = P(X = x)P(Y = y),$$

for all x, y with x in the support of X and y in the support of Y.

The definition for more than two r.v.s is analogous.

Example 3.8.2.

In a roll of two fair dice, if X is the number on the first die and Y is the number on the second die, then $X + Y$ is not independent of $X - Y$. To see why, note that

$$0 = P(X + Y = 12, X - Y = 1) \neq P(X + Y = 12)P(X - Y = 1) = \frac{1}{36} \cdot \frac{5}{36}.$$

Since we have found a pair of values (s, d) for which

$$P(X + Y = s, X - Y = d) \neq P(X + Y = s)P(X - Y = d),$$

$X+Y$ and $X-Y$ are dependent. This also makes sense intuitively: knowing the sum of the dice is 12 tells us their difference must be 0, so the r.v.s provide information about each other. If X and Y are independent then it is also true, for example, that X^2 is independent of Y^3, since if X^2 provided information about Y^3 then X would give information about Y (using X^2 and Y^3 as intermediaries). In general, if X and Y are independent then any function of X is independent of any function of Y.

Definition 3.8.3 (i.i.d.).

We will often work with random variables that are independent and have the same distribution. We call such r.v.s *independent and identically distributed*, or *i.i.d.* for short.

By taking a sum of i.i.d. Bernoulli r.v.s, we can write down the story of the Binomial distribution in an algebraic form.

Theorem 3.8.4.

If $X \sim \text{Bin}(n, p)$, viewed as the number of successes in n independent Bernoulli trials with success probability p, then we can write $X = X_1 + \cdots + X_n$ where the X_i are i.i.d. $\text{Bern}(p)$.

Proof

Let $X_i = 1$ if the ith trial was a success, and 0 if the ith trial was a failure. It's as though we have a person assigned to each trial, and we ask each person to raise their hand if their trial was a success. If we count the number of raised hands (which is the same as adding up the X_i), we get the total number of successes in the n trials, which is X. An important fact about the Binomial distribution is that the sum of independent Binomial r.v.s with the same success probability is another Binomial r.v. The previous theorem gives an easy proof, though we include two others for comparison.

Theorem 3.8.5.

If $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$, and X is independent of Y, then $X + Y \sim \text{Bin}(n + m, p)$.

Proof

We present three proofs, since each illustrates a useful technique.

1. LOTP: We can directly find the PMF of $X+Y$ by conditioning on X (or Y, whichever we prefer) and using the law of total probability:

$$\begin{aligned}
P(X+Y=k) &= \sum_{j=0}^{k} P(X+Y=k|X=j)P(X=j) \\
&= \sum_{j=0}^{k} P(Y=k-j)P(X=j) \\
&= \sum_{j=0}^{k} \binom{m}{k-j} p^{k-j} q^{m-k+j} \binom{n}{j} p^j q^{n-j} \\
&= p^k q^{n+m-k} \sum_{j=0}^{k} \binom{m}{k-j}\binom{n}{j} \\
&= \binom{n+m}{k} p^k q^{n+m-k}.
\end{aligned}$$

$$P(X+Y=k|X=j) = P(Y=k-j|X=j) = P(Y=k-j),$$

and in the last line, we used the fact that

$$\sum_{j=0}^{k} \binom{m}{k-j}\binom{n}{j} = \binom{n+m}{k}$$

by Vandermonde's identity. The resulting expression is the $\text{Bin}(n+m,p)$ PMF, so $X+Y \sim \text{Bin}(n+m,p)$.

2. Representation: A much simpler proof is to represent both X and Y as the sum of i.i.d. $\text{Bern}(p)$ r.v.s: $X = X_1 + \cdots + X_n$ and $Y = Y_1 + \cdots + Y_m$, where the X_i and Y_j are all i.i.d. $\text{Bern}(p)$. Then $X+Y$ is the sum of $n+m$ i.i.d. $\text{Bern}(p)$ r.v.s, so its distribution, by the previous theorem, is $\text{Bin}(n+m,p)$.

3. Story: By the Binomial story, X is the number of successes in n independent trials and Y is the number of successes in m additional independent trials, all with the same success probability, so $X+Y$ is the total number of successes in the $n+m$ trials, which is the story of the $\text{Bin}(n+m,p)$ distribution. Of course, if we have a definition for independence of r.v.s, we should have an analogous definition for conditional independence of r.v.s.

Definition 3.8.6 (Conditional independence of r.v.s).

Random variables X and Y are *conditionally independent* given an r.v. Z if for all $x, y \in \mathbb{R}$ and all z in the support of Z,

$$P(X \le x, Y \le y|Z=z) = P(X \le x|Z=z)P(Y \le y|Z=z).$$

For discrete r.v.s, an equivalent definition is to require

$$P(X=x, Y=y|Z=z) = P(X=x|Z=z)P(Y=y|Z=z).$$

This is the definition of independence, except that we condition on $Z = z$ everywhere, and require the equality to hold for all z in the support of Z.

> **Definition 3.8.7 (Conditional PMF).**
>
> For any discrete r.v.s X and Z, the function $P(X = x|Z = z)$, when considered as a function of x for fixed z, is called the *conditional PMF of X given $Z = z$*.

Independence of r.v.s does not imply conditional independence, nor vice versa. First let us show why independence does not imply conditional independence.

Example 3.8.8 (Matching pennies).

Consider the simple game called *matching pennies*. Each of two players, A and B, has a fair penny. They flip their pennies independently. If the pennies match, A wins; otherwise, B wins. Let X be 1 if A's penny lands Heads and -1 otherwise, and define Y similarly for B. Let $Z = XY$, which is 1 if A wins and -1 if B wins. Then X and Y are unconditionally independent, but given $Z = 1$, we know that $X = Y$ (the pennies match). So X and Y are conditionally dependent given Z.
Next let's see why conditional independence does not imply independence.

Example 3.8.9 (Mystery opponent).

Suppose that you are going to play two games of tennis against one of two identical twins. Against one of the twins, you are evenly matched, and against the other you have a $3/4$ chance of winning. Suppose that you can't tell which twin you are playing against until after the two games. Let Z be the indicator of playing against the twin with whom you're evenly matched, and let X and Y be the indicators of victory in the first and second games, respectively. Conditional on $Z = 1$, X and Y are i.i.d. $\textrm{Bern}(1/2)$, and conditional on $Z = 0$, X and Y are i.i.d. $\textrm{Bern}(3/4)$. So X and Y are conditionally independent given Z. Unconditionally, X and Y are dependent because observing $X = 1$ makes it more likely that we are playing the twin who is worse. Past games give us information which helps us infer who our opponent is, which in turn helps us predict future games!