Lecture 18. Adding Random Variables

Suppose X and Y are discrete random variables on $\{0, 1, 2, 3, \ldots\}$. Suppose f and g are the corresponding pmfs:

$$P(X = x) = f(x), \qquad P(Y = y) = g(y).$$

We assume there is no influence of one variable upon the other.

Problem. What is $P(X + Y = n)$?

Illustration. Suppose X is the result obtained by rolling one die, and Y is the result obtained by rolling another. Let $n = 8$. We must consider *all ways* that the sum can be 8, and we see that there are the following possibilities: $(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)$, where the ordered pairs represent a possible value of X first and a possible value of Y second. Each of these outcomes has probability $1/36$, and if we add these probabilities together, we get the probability of the event.

Solution. Just as in the case of the dice, we need to consider all ways that X and Y may sum to n. If (a, b) is one such way, then $P(X = a \& Y = b) = P(X = a) \cdot P(Y = b) = f(a) \cdot g(b)$, since we have assumed there is no interaction between the variables. We want to sum the probabilities of all the outcomes in this event, so our result is:

$$P(X + Y = n) = \sum_{x+y=n} f(x) \cdot g(y), \quad x, y \in \{0, 1, 2, 3, \ldots\}.$$

Elaboration. The formula above is reminiscent of the formula for multiplying polynomials. Suppose $p(t) = a_0 + a_1 t + a_2 t^2 + \cdots$ and $q(t) = b_0 + b_1 t + b_2 t^2 + \cdots$. If we use c_i to denote the i^{th} coefficient of the product, i.e.,

$$p(t) \cdot q(t) = (a_0 + a_1 t + \cdots)(b_0 + b_1 t + \cdots) = c_0 + c_1 t + c_2 t^2 + \cdots,$$

then

$$c_n = \sum_{i+j=n} a_i \cdot b_j, \quad i, j \in \{1, 2, 3, \ldots\}.$$

Note that p and q could just as well have been power series. The same formula for the coefficients of the product apply.

We can associate with the variables X and Y the power series

$$p_X(t) = f(0) + f(1)t + f(2)t^2 + \cdots$$

$$p_Y(t) = g(0) + g(1)t + g(2)t^2 + \cdots.$$

It follows that

$$\text{if } W = X + Y, \text{ then } p_W = p_X \cdot p_Y$$

Let us illustrate the technique with the dice problem. In this case,

$$p_X(t) = p_Y(t) = \frac{1}{6}\big(t + t^2 + t^3 + t^4 + t^5 + t^6\big).$$

(The constant term vanishes because the probability that $X = 0$ is zero.) Thus,

$$\begin{aligned} p_{X+Y}(t) &= \frac{1}{36}\big(t + t^2 + t^3 + t^4 + t^5 + t^6\big)^2 \\ &= \tfrac{1}{36}t^2 + \tfrac{2}{36}t^3 + \tfrac{3}{36}t^4 + \tfrac{4}{36}t^5 + \tfrac{5}{36}t^6 + \tfrac{6}{36}t^7 + \tfrac{5}{36}t^8 + \tfrac{4}{36}t^9 + \tfrac{3}{36}t^{10} + \tfrac{2}{36}t^{11} + \tfrac{1}{36}t^{12} \end{aligned}$$

Problem. We say X *is Bernoulli with parameter* p if X is discrete and

$$P(X = 0) = 1 - p, \quad P(X = 1) = p, \quad \text{and } \; P(X = x) = 0 \; \text{ for all } x \neq 0, 1.$$

(See the textbook, pages 188-9.) If X is Bernoulli with parameter p, then

$$p_X(t) = (1 - p) + p\,t.$$

Accordingly, if X and Y are both Bernoulli with parameter p, then

$$p_{X+Y}(t) = ((1 - p) + p\,t)^2 = (1 - p)^2 + 2(1 - p)\,p\,t + p^2\,t^2.$$

From this, we can read off the pmf of $X + Y$ by looking at the coefficients of t:

$$\begin{aligned} P(X + Y = 0) &= \text{constant term} = (1 - p)^2 \\ P(X + Y = 1) &= \text{coefficient of } t = 2\,(1 - p)\,p \\ P(X + Y = 2) &= \text{coefficient of } t^2 = p^2 \end{aligned}$$

Thus, $X + Y$ has a binomial distribution with parameters $n = 2$ and p.

Homework. What if $X_1, \ldots, X_m$ are independent, and each is Bernoulli with parameter p. Find the pmf of $W = X_1 + X_2 + \cdots + X_m$. Is it related to a binomial pmf?