2.5. Bernoulli Trials and the Binomial Distribution

1. *Definition.* A Bernoulli trial is an experiment with two, and only two, possible outcomes.

2. *Example.* The toss of a coin is a Bernoulli trial.

3. *Definition.* Let X be a random variable, and let p be a constant such that $0 \leq p \leq 1$. If

$$P(X = 1) = p, \quad P(X = 0) = 1 - p,$$

then we say X has the Bernoulli(p) distribution.

4. Let $X \sim$ Bernoulli(p), our convention is to consider the event $X = 1$ as a success and the event $X = 0$ as a failure. In this context, p is usually called the success rate or success probability, $1 - p$ is often called the failure rate.

5. *Example.* Think of a multiple-choice question with three choices only one of which is correct. Suppose that we decide to make a random guess. Define

$$X = \begin{cases} 1, & \text{we get the right answer,} \\ 0, & \text{we don't get the right answer.} \end{cases}$$

Then, X is a Bernoulli random variable. And, X has pmf

x	0	1
$f(x)$	$\frac{2}{3}$	$\frac{1}{3}$

Here, the event $X = 1$ is a success, $\frac{1}{3}$ is the success rate; the event $X = 0$ is a failure, $\frac{2}{3}$ is the failure rate.

6. *Definition.* Suppose that $X \sim \text{Bernoulli}(p)$. Suppose that $X_1, X_2, \cdots, X_n$ are a sample from the population X.

   Let $Y = X_1 + X_2 + \cdots + X_n$. Then the range of Y is $\{0, 1, 2, 3, \cdots, n\}$, and we can show that (will be proved later) pmf of Y is

$$f(y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, 1, 2, \cdots, n.$$

   We say Y has the $\text{Binomial}(n, p)$ distribution.

7. *Example.* An algebra test has five multiple choice questions. Each question has three choices, of which only one is correct. Suppose a certain student just randomly guesses on each of the five questions. Let Y be the number of questions this student will answer correctly. Find the distribution of Y.

— *Solution.* For each $i = 1, 2, 3, 4, 5$, define the random variable X_i as:

$$X_i = \begin{cases} 1, & \text{the } i\text{-th question is answered correctly,} \\ 0, & \text{the } i\text{-th question is not answered correctly.} \end{cases}$$

Then, each X_i is a Bernoulli random variable. And, $X_1, X_2, \cdots, X_5$ have a common pmf

x	0	1
$f(x)$	$\frac{2}{3}$	$\frac{1}{3}$

It is obvious that $X_1, X_2, \cdots, X_5$ are independent. So, by definition, $X_1, X_2, \cdots, X_5$ form a sample of size five from the population $f(x)$. Here, the population is the Bernoulli($\frac{1}{3}$) distribution. It is also clear that

$$Y = X_1 + X_2 + \cdots + X_5.$$

Y has possible values $0, 1, 2, 3, 4, 5$. We will now find, for each $i = 0, 1, 2, 3, 4, 5$,

$$P(Y = i).$$

In fact, we can show that, for each $i = 0, 1, 2, 3, 4, 5$,

$$P(Y = i) = \binom{5}{i} \left(\frac{1}{3}\right)^i \left(\frac{2}{3}\right)^{5-i}.$$

The proof is combinatorial in nature, and will be provided later. The key point is, the event $Y = i$ can occur in $\binom{5}{i}$ ways, and the probability of each of these ways is $\left(\frac{1}{3}\right)^i \left(\frac{2}{3}\right)^{5-i}$.

(Example continues on the next page.)

Let consider the $P(Y = 3)$ in detail. $Y = 3$ means three questions are answered correctly, and these three questions can be any three of the five questions. So $Y = 3$ is a not a single sample point but a subset of the sample space. In fact, the event $Y = 3$ can occur in $\binom{5}{3} = 10$ ways, and the probability of each of these ten ways is $\left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^2$ — two failures plus three successes. The ten ways in which the event $Y = 3$ can occur are listed in the table on the next page, and each of the ten ways corresponds to a 3-element subset of $\{1, 2, 3, 4, 5\}$.

Since the set $\{1, 2, 3, 4, 5\}$ has $\binom{5}{3} = 10$ 3-element subsets, so the event $Y = 3$ can occur in $\binom{5}{3} = 10$ ways. It follows that

$$P(Y = 3) = \binom{5}{3} \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^2 .$$

X_1	X_2	X_3	X_4	X_5	corresponding set
1	1	1	0	0	$\{1, 2, 3\}$ $\leftarrow$
1	1	0	1	0	$\{1, 2, 4\}$
1	1	0	0	1	$\{1, 2, 5\}$
1	0	1	1	0	$\{1, 3, 4\}$
1	0	1	0	1	$\{1, 3, 5\}$
1	0	0	1	1	$\{1, 4, 5\}$
0	1	1	1	0	$\{2, 3, 4\}$
0	1	1	0	1	$\{2, 3, 5\}$
0	1	0	1	1	$\{2, 4, 5\}$
0	0	1	1	1	$\{3, 4, 5\}$