BERNOULLI TRIALS AND THE BINOMIAL DISTRIBUTION

1. Independent Bernoulli trials

The archetypical example of a Bernoulli trial is the tossing of a coin, fair or not. The associated random variable is

$$X = \begin{cases} 
1, & \text{if the coin came up heads,} \\
0, & \text{otherwise.}
\end{cases}$$

Consequently, $X = 1$ with probability $p$, if that is the probability that the coin will come up heads, etc. And surely, we know what independent coin tosses are.

A Bernoulli trial is an experiment with probability $p$ of success and probability $1-p$ of failure. We define a Bernoulli random $X$ by

$$X = \begin{cases} 
1, & \text{if the experiment was successful,} \\
0, & \text{otherwise,}
\end{cases}$$

and consequently,

$$\mathbb{P}[X = 1] = p, \quad \mathbb{P}[X = 0] = 1 - p.$$ 

We write concisely

$$X \sim \text{Bernoulli}(p).$$

It is a nice exercise to verify that $\mathbb{E}[X^m] = p$, $m = 1, 2, 3, \cdots$, so that

$$\text{If } X \sim \text{Bernoulli}(p) \text{ then } \mathbb{E}[X] = p, \quad \mathbb{V}[X] = p(1-p).$$

2. Binomial random variables

The archetypical example of a Binomial random variable is the number of heads obtained in $n$ independent tosses of the same coin.

The number $N$ of successes in $n$ independent, identical Bernoulli trials with probability $p$ of success is a Binomial$(n, p)$ random variable,

$$N \sim \text{Binomial}(n, p).$$

Obviously (it takes work actually),

$$\mathbb{P}[N = k] = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \cdots, n.$$
Obviously (here, it is indeed obvious), the only possible outcomes for a Binomial\((n, p)\) random variable are 0, 1, \cdots, \(n\). Depending on one’s temperament, one can prove this by noting that Newton’s binomial formula,

\[
(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k},
\]

implies that

\[
\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1.
\]

Since the probabilities add up to 1, we must have listed all possible outcomes. (Or, all other outcomes have probability 0.)

How shall we compute the mean and the variance of a Binomial random variable. It seems clear that if

\[ N \sim \text{Binomial}(n, p), \]

then \(\mathbb{E}[N] = np\). This very much fits into the frequentist view of probability. But for the variance we are in trouble. The following is the authorized way of going about it.

\textbf{Binomial \equiv sum of iid Bernoulli-s}

Let \(X_1, X_2, \cdots, X_n\) be independent Bernoulli\((p)\) random variables. (So, we are performing \(n\) independent Bernoulli trials.) Then,

\[ M = \sum_{i=1}^{n} X_i \]

counts the number of successes in the \(n\) trials, and so

\[ M \sim \text{Binomial}(n, p). \]

But, since the mean of a sum is the sum of the means

\[ \mathbb{E}[M] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np, \]

but we knew that already.

For the variance, since \(M\) is a sum of independent random variables, its variance is the sum of the variances, so

\[ \text{Var}[M] = \sum_{i=1}^{n} \text{Var}[X_i] = np(1-p). \]

Note that the variance is largest when \(p = \frac{1}{2}\), i.e., when betting on the outcome of a toss is least favorable. When \(p\) is close to 0 or 1, the outcome becomes more certain, and there is less statistical variation in \(M\).
To summarize, if $N \sim \text{Binomial}(n, p)$, then
\[ \mathbb{E}[N] = np, \quad \text{Var}[N] = np(1-p). \]

COMMENT. Statisticians and probabilists are extremely happy with expressions like
\[ N = \sum_{i=1}^{n} X_i, \]
with independent, identically distributed random variables. In our case, they are Bernoulli($p$). In other applications, they will be different (e.g., the heights of students, or annual income of households, etc.) but that hardly matters.