1 Binomial numbers and related distributions

1.1 Binomial numbers

Texans (used to) play the lottery by selecting six different numbers between 1 and 50, which cost \$1 for each combination. Twice a week, on Wednesday and Saturday at 10 PM, six ping-pong balls are released without replacement from a rotating plastic ball containing 50 ping-pong balls numbered 1 through 50. The winner of the jackpot (which occasionally accumulates to 60 or more million dollars!) is the one who has all six drawn numbers correct, where the order in which the numbers are drawn does not matter. What are the odds of winning if you play one set of six numbers only?

In order to answer this question, suppose first that the order of the numbers does matter. Then the number of ordered sets of 6 out of 50 numbers is: 50 possibilities for the first drawn number, times 49 possibilities for the second drawn number, times 48 possibilities for the third drawn number, times 47 possibilities for the fourth drawn number, times 46 possibilities for the fifth drawn number, times 45 possibilities for the sixth drawn number:

$$\prod_{j=0}^{5}(50-j)=\frac{\prod_{k=1}^{50}k}{\prod_{k=1}^{50-6}k}=\frac{50!}{(50-6)!}.$$

The notation $n!$ (read: n factorial) stands for the product of the natural numbers 1 through n:

$$n!=1\times 2\times ....\times(n-1)\times n \text{ if } n>0,\ 0!=1.$$

The reason for defining $0!=1$ will be explained below.

Since a set of six given numbers can be permutated in $6!$ ways, we need to correct the above number for the $6!$ replications of each unordered set of six given numbers. Therefore, the number of sets of six unordered numbers out of 50 is:

$$\binom{50}{6}=\frac{50!}{6!\,(50-6)!}=15,890,700.$$

Thus, the probability of winning the Texas lottery if you play only one combination of six numbers is 1 out of $15,890,700$.

Note: In the Spring of 2000, the Texas lottery has changed the rules: Now one has to choose 6 different numbers between 1 and 54, which reduces the probability of winning to 1 one out of $25,827,165$. The reason for this change is to boost the jackpot, because the higher the jackpot, the more people play.

The setup of the Pennsylvania lottery is different: One has to choose 5 out of 59 numbers and then a power ball (numbered from 1 to 39). The probability of winning in this case is 1 out of

$$\binom{59}{5} \times 39 = \frac{59!}{5!\,(59-5)!} \times 39 = 5,006,386 \times 39 = 195,249,054$$

which is clearly a rip-off compared with the Texas lotto!

In general, the number of ways we can draw a set of k unordered objects out of a set of n objects without replacement is:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

These numbers, read as: n choose k, also appear as coefficients in the binomial expansion

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

The reason for defining $0! = 1$ is now that the first and last coefficients in this binomial expansion are always equal to 1:

$$\binom{n}{0} = \binom{n}{n} = \frac{n!}{0!n!} = \frac{1}{0!} = 1.$$

1.2 The binomial distribution

Consider a bowl containing r red balls and $N-r$ white balls, where $0 < r < N$. Draw randomly n balls from this bowl *with replacement,* i.e., shake the bowl thoroughly, draw blindfolded a ball, take the blindfold off, observe the color of the ball you have drawn, *put the ball back* in the bowl (and the blindfold on!), and repeat this procedure n times.

The number of ways you can draw an *ordered* sequence of k red balls and $n-k$ white balls in this way is: $r^k\,(N-r)^{n-k}$, and the number of ways you can draw an ordered sequence of n balls (of any color) is N^n. Thus, the probability that you draw a sequence of k red balls and $n-k$ white balls *in a particular order* is: $r^k\,(N-r)^{n-k}/N^n = (p)^k\,(1-p)^{n-k}$, where $p = r/N$.

But the number of *ordered* sequences of k red balls and $n-k$ white balls is: $\binom{n}{k}$. Therefore, if X is the number of red balls you have drawn, then

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,1,...,n.$$

This distribution is called the Binomial (n,p) distribution.

The expectation of X is:

$$E[X]=n.p$$

1.3 The hypergeometric distribution

Consider again a bowl containing r red balls and $N-r$ white balls, where $0<r<N$. Draw randomly n balls ($n\leq r$) from this bowl *without replacement,* i.e., shake the bowl thoroughly, draw blindfolded a ball, *don't put the ball back* in the bowl, and repeat this procedure n times.

The number of ways you can draw k red balls and $n-k$ white balls in this way is $\binom{r}{k}\binom{N-r}{n-k}$, and the number of ways you can draw n balls (of any color) is $\binom{N}{n}$. Clearly, if $n>r$ and $k>r$, or $k>n$, there is no way to draw k red balls and $n-k$ white balls. Therefore, if X is the number of red balls you have drawn, then

$$\begin{aligned} P(X &= k)=\frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}\text{ for }k=0,1,...,\min(n,r), \\ P(X &= k)=0\text{ for }k>\min(n,r). \end{aligned}$$

This distribution is called the Hypergeometric (n,r,N) distribution.

If n is very small relative to N, this distribution is approximately equal to the Binomial (n,p) distribution, with $p=r/N$.

The expectation of X is

$$E[X]=\frac{n.r}{N}.$$

1.4 The negative binomial distribution

Consider a sequence of independent repetitions of a random experiment with constant probability p of success. Let the random variable X be the total number of failures in this sequence before the m-th success, where $m \geq 1$. Thus, $X + m$ is equal to the number of trials necessary to produce exactly m successes. The probability $P(X = k)$, $k = 0, 1, 2, ...$, is the product of the probability of obtaining exactly $m - 1$ successes in the first $k + m - 1$ trials, which is equal to the (Binomial) probability

$$\binom{k+m-1}{m-1} p^{m-1} (1-p)^{k+m-1-(m-1)},$$

and the probability p of a success on the $(k + m)$-th trial:

$$P(X = k) = \binom{k+m-1}{m-1} p^{m} (1-p)^{k}, \ k = 0, 1, 2, ...$$

This distribution is called the Negative Binomial (m, p) distribution.

The expectation of X is:

$$E[X] = m \left(p^{-1} - 1 \right).$$

1.5 The Poisson distribution

Let X_n be Binomial (n, p_n) distributed:

$$P(X_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k}, \ k = 0, 1, ..., n,$$

and suppose that for $n = 1, 2, ...$, $p_n \downarrow 0$ as $n \to \infty$, such that for $n > c$, $np_n = c$, where $c > 0$ is a constant. Then for fixed k, and $n > c$,

$$P(X_n = k) = (1 - \frac{c}{n})^{-k} \times \frac{n!}{n^k (n-k)!} \times (1 - \frac{c}{n})^n \times \frac{c^k}{k!}. \tag{1}$$

The first factor in (1) converges to 1:

$$\lim_{n\to\infty} (1 - \frac{c}{n})^{-k} = (1 - \lim_{n\to\infty} \frac{c}{n})^{-k} = 1.$$

The second factor in (1) equals 1 if $k = 0$, and converges to 1 if $k \geq 1$:

$$\lim_{n\to\infty} \frac{n!}{n^k (n-k)!} = \lim_{n\to\infty} \frac{n(n-1) \times ... \times (n-k+1)}{n^k} = \lim_{n\to\infty} \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right) = 1.$$

The third factor in (1) converges to $\exp(-c)$, because

$$\begin{aligned}\lim_{n\to\infty} \ln(1-\frac{c}{n})^n &= -c\lim_{n\to\infty} \frac{\ln(1-\frac{c}{n})-\ln 1}{-c/n} = -c\lim_{\delta\to 0} \frac{\ln(1+\delta)-\ln 1}{\delta} \\ &= -c \times \left.\frac{d\ln(x)}{dx}\right|_{x=1} = -c \times \left.\frac{1}{x}\right|_{x=1} = -c.\end{aligned}$$

Thus, for fixed $k = 0, 1, 2, ...$, $\lim_{n\to\infty} P(X_n = k) = P(X = k)$, where

$$P(X = k) = \exp(-c)\frac{c^k}{k!}.$$

This distribution is called the Poisson (c) distribution. Since it is the limit of a Binomial (n, p) distribution with $p = c/n$ for $n > c$, the Poisson distribution is often used to model the distribution of *rare* events.

The expectation of X is:

$$E[X] = c$$