The Binomial Distribution

James H. Steiger

November 10, 2003

1. Topics for this Module

2. The Binomial Process

Perhaps the simplest situation we encounter in probability theory is the case where only two things can happen. If one of the two things has probability p, the alternative thing must have probability $q = 1 - p$. Suppose we call the two outcomes "Success" and "Failure" code the outcomes 1 and 0 with the random variable T. Then T has the following probability distribution

Table 1: Probability Distribution for a Binary Random Variable

t	$P_T(t)$
1	p
0	$1 - p$

It is easy to establish (C.P.) that the random variable T has an expected value of p and a variance of $p(1 - p) = pq$.

Topics for this Module
The Binomial Process
The Binomial . . .
The Binomial . . .
Modelling the . . .
The Normal . . .
Home Page
Print
Title Page
Page 3 of 16
Go Back
Full Screen
Close
Quit

Suppose we have a sequence of N independent trials where, on each trial, we have an outcome on the random variable T. Then such a process has the following characteristics:

- There are N *independent* trials
- Only two things can happen. One, arbitrarily labelled "Success," has probability p.The other, arbitrarily labeled "Failure," has probability $q = 1 - p$.
- Probabilities remain constant from trial to trial.

Such a process is a good approximation to many "real world" processes. For example, consider the following:

- A woman has 5 children. They are either boys or girls.
- A factory produces 100 radios. They are either have a manufacturing defect or they do not.
- A basketball player attempts 385 free throws in a season. On each attempt, he either succeeds or fails.

If you consider the above 3 situations carefully, you can advance reasons why, in one respect or another, they fail to match the ideal description of the binomial process. Typically, either the assumption of independence, or the assumption of constant probability (or both) will be violated. For example, a woman's probability of having either a boy or a girl may change as she gets older. Perhaps, if the first 3 children are all boys or girls, she may take steps to alter the balance of probabilities for subsequent children. Manufacturing defects may occur because of bad "batches" of components that render production "trials" non-independent. Basketball players may try harder after missing a free throw, thereby introducing dependencies into the sequence of free throw attempts.

So we recognize at the outset that the binomial process as described above is an idealistic set of specifications that is seldom met in practice. Nonetheless, the binomial process often provides an excellent approximation to natural processes, and allows us to make excellent predictions in many situations.

3. The Binomial Random Variable

Often, we are not interested in the precise sequence of outcomes in a binomial process, but rather in the *number of successes* in the N trial sequence. For example, a basketball player's free throw percentage is based on the number of successes in N free throws, not on the precise sequence that produces the number of successes. Consequently, we often code a sequence of binomial trials with the binomial random variable, defined as follows:B

Definition 3.1 ***(The Binomial Random Variable)*** *Given N trials of a binomial process with probability of success p, the binomial random variable X is the number of success that occurs in the N trial sequence.*

4. The Binomial Distribution

The binomial distribution is a family of distributions with two parameters — N, the number of trials, and p, the probability of success. We refer to the binomial random variable with general notation $B\left(N,p\right)$. For example, $B(10,1/2)$ refers to a 10 trial binomial process with probability of success equal to $1/2$.

4.1. Computing the Binomial pdf

The $B\left(N,p\right)$ distribution has the following probability distribution function (or *pdf*):

$$P_X(r;N,p) = \binom{N}{r} p^r \left(1-p\right)^{N-r} \tag{1}$$

In Equation 1, the notation $P_X(r;N,p)$ means the probability that exactly r successes occur in a sequence of N trials when the probability of success is p.

Example 4.1 ***(An unfair coin)*** *Consider an unfair coin, with* $p = \Pr(\text{Head}) = 2/3$. *What is the probability that, in 4 tosses of this coin, exactly 3 heads occur?*

Solution 4.1 *Using Equation 1, we find*

$$P_X(3;4;2/3) = \binom{4}{3}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^1 = 4\left(\frac{8}{27}\right)\left(\frac{1}{3}\right) = \frac{32}{81}$$

To see where Equation 1 came from, we need to look at the above problem in slightly more detail. First, consider *just one* sequence that yields exactly 3 heads, e.g.,

$$HHHT$$

Because the 4 trials are assumed to be independent in a binomial process, we have

$$\Pr(HHHT) = \Pr(H)\Pr(H)\Pr(H)\Pr(T) = \left(\frac{2}{3}\right)\left(\frac{2}{3}\right)\left(\frac{2}{3}\right)\left(\frac{1}{3}\right) = \left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^1 = \frac{8}{81}$$

This sequence has probability $8/81$, but it is not the only sequence that yields exactly 3 heads. For example, consider the following sequence, which also yields 3 heads

$$HTHH$$

What is the probability of this sequence? It is

$$\begin{aligned}\Pr(HTHH) &= \Pr(H)\Pr(T)\Pr(H)\Pr(H) = \left(\frac{2}{3}\right)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)\left(\frac{2}{3}\right) \\ &= \left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^1 = \frac{8}{81}\end{aligned}$$

A little thought should convince you that each sequence yielding exactly 3 heads must have this same probability. Consequently, computing the total probability involves finding out exactly how many different sequences yield 3 heads, and multiplying that number by $8/81$. More generally, we can say that

$$P_X(r; N, p) = (\text{Number of sequences yielding } r \text{ successes}) \times p^r (1-p)^{N-r}$$

The number of different sequences that yield r successes in N trials must be $\binom{N}{r}$, because that is the number of ways one can select r trials (out of N) on which the successes occur. In this example, we have 4 distinct sequences that yield exactly 3 heads. So the total probability is $4(8/81) = 32/81$. Thus we end up with Equation 1.

4.2. Computing the Binomial cdf

We begin with a definition.

Definition 4.1 ***(cdf)*** *The cumulative distribution function, or cdf, for the random variable X, evaluated at x is the probability that X takes on a value less than or equal to x, i.e.*

$$F_X(x) = \Pr(X \le x)$$

Many interesting problems involve the cumulative binomial distribution, which is computed as

$$F_X(r; N, p) = \sum_{i=0}^{r} P_X(i; N, p) = \sum_{i=0}^{r} \binom{N}{i} p^i (1-p)^{N-i} \tag{2}$$

Example 4.2 ***(The Cumulative Binomial)*** *Suppose you play 5 rounds of a gambling game where the odds are in your favor, i.e., your probability of winning a round is 5/9. If you bet the same amount of money in each of the 5 rounds, what is the probability that you will lose money, that is, win 2 or fewer rounds?*

Solution 4.2 *The answer is*

$$\begin{aligned}
F_X(2;5,5/9) &= \sum_{i=0}^{2} \binom{5}{i} \left(\frac{5}{9}\right)^{i} \left(\frac{4}{9}\right)^{5-i} \\
&= \binom{5}{0} \left(\frac{5}{9}\right)^{0} \left(\frac{4}{9}\right)^{5} + \binom{5}{1} \left(\frac{5}{9}\right)^{1} \left(\frac{4}{9}\right)^{4} + \binom{5}{2} \left(\frac{5}{9}\right)^{2} \left(\frac{4}{9}\right)^{3} \\
&= \frac{(1)(1)(4^5) + (5)(5)(4^4) + (10)(5^2)(4^3)}{9^5} \\
&= \frac{1024 + 6400 + 16000}{59049} = \frac{7808}{19\,683} = 0.396\,687
\end{aligned}$$

Note! Even though the odds are 5 *to* 4 *in your favor, you will lose money about 40% of the time if you play only 5 rounds of the game. But suppose you play 21 rounds of the game. What is the probability you will lose money? It is much lower, i.e.*

$$F_X(10;21,5/9) = \sum_{i=0}^{10} \binom{21}{i} \left(\frac{5}{9}\right)^{i} \left(\frac{4}{9}\right)^{21-i} = \frac{33\,112\,451\,722\,283\,843\,584}{109\,418\,989\,131\,512\,359\,209}$$

which is only $0.302\,621$*. The message is clear, if you play enough rounds, the odds of losing money will become arbitrarily small. Las Vegas was founded on this principle.*

5. Modelling the Public Opinion Poll

The binomial process is used frequently as a model for public opinion polls. Suppose that a public opinion poll is taken by a politician who is interested in how many people want her to run for president. Suppose she samples 100 people completely at random from a very large population of potential voters. Imagine that exactly 50% of the entire population wants her to run (but she does not know this). What is the probability that the public opinion poll will provide her with an answer that is accurate to within ±10%?

Strictly speaking, the politician is sampling without replacement. However, when the population is very large relative to the size of the sample, the dependencies caused by sampling without replacement have a miniscule effect, so the binomial model may be an excellent approximation. For the poll to be accurate within ±10%, between 40 and 60 people must say "Yes" when asked the question. If the true population proportion p is $1/2$, the probability of this occurring is

$$\begin{aligned}
\Pr(40 &\leq X \leq 60) = F_X(60; 100, 1/2) - F_X(39; 100, 1/2) \\
&= \sum_{i=40}^{60} \binom{100}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{100-i} \\
&= \frac{38\,219\,657\,665\,440\,688\,759\,455\,013\,113}{39\,614\,081\,257\,132\,168\,796\,771\,975\,168} = .9648
\end{aligned}$$

With modern computer software, we can perform such calculations routinely. However, when such software is not readily available, one can employ an approximation to the binomial distribution that, in this case, will be quite accurate.

6. The Normal Approximation to the Binomial Distribution

When N is reasonably large and p is not too far from 1/2 (and more generally, when both Np and $N(1-p)$ are greater than or equal to 10), the binomial distribution can be approximated by a normal distribution with mean Np and variance $Np(1-p)$. These results follow from two distinct sources:

1. Linear Combination Theory
2. The Central Limit Theorem

If events on each trial of a binomial process are coded 1 for Success, 0 for failure, the binomial random variable X is simply the sum of the outcomes on the N trials. Recall we showed earlier that a binary random variable coded 0-1 has a mean of p and a variance of $p(1-p)$. So the sum of N independent trials must have a mean of Np and a variance of $Np(1-p)$, so long as the trials are independent. The fact that the binomial distribution is well approximated by the normal distribution follows from the Central Limit Theorem. We state a simplified version of the theorem below.

Theorem 6.1 ***(The Central Limit Theorem for the Sample Mean).*** *Given a sample of N independent observations from a distribution with* *mean μ and finite variance σ^2, the sample mean $\overline{X}_\bullet$ approaches a normal distribution with mean μ and variance σ^2/N as $N \to \infty$. (Note: Strictly speaking the theorem should be stated slightly differently. It should say that as $N \to \infty$, the distribution of $\sqrt{N}\left(\overline{X}_\bullet - \mu\right)$ converges to a normal distribution with mean μ and variance σ^2, because, as $N \to \infty$, σ^2/N must converge to zero. However, since the variance of $\overline{X}_\bullet$ is σ^2/N for any value of N, writers of applied statistics texts are given license to state the result in a way that many students find simpler to comprehend and use.)*

Remark 6.1 *The Central Limit Theorem is a remarkable result. From minimal assumptions, we end up with normality! It is important to remember that the CLT is a result in* asymptotic distribution theory. *Frequently, in practice, it is not possible to derive the exact distribution of a statistic, but it is possible to derive what the distribution converges to as N becomes arbitrarily large. Asymptotic results may or may not represent good approximations to the sampling distribution of a statistic at realistic sample sizes. It all depends on the "rate of convergence," i.e., how fast the distribution converges in shape to the asymptotic form. In the case of the sample mean, the convergence tends to be quite fast. Glass and Hopkins (p. 235–238) examine the speed of convergence of the sample mean under a number of conditions. In practice, the CLT often implies that*

Example 6.1 ***(Revisiting the Public Opinion Poll)*** *In the preceding example, we performed an exact calculation using integer arithmetic.*

We can approximate the calculation using the normal approximation to the binomial. With $N = 100$*, and* $p = 1/2$*, the binomial distribution is approximated by a normal distribution with* $\mu = Np = 100(1/2) = 50$*, and variance* $\sigma^2 = Np(1-p) = 100(1/2)(1/2) = 25$*.So the binomial is approximated by a normal distribution with* $\mu = 50$ *and* $\sigma = 5$*. If one computes the probability that* X *is between* 40 *and* 60 *by simply evaluating this normal curve between those points, one obtains* $.9772 - .0228 = .9544$*. Recall, however, that the normal random variable is continuous, while the binomial is discrete. So, we can obtain a closer approximation by "correcting for continuity,"i.e., computing the area between* 39.5 *and* 60.5*. One obtains* 0.9643*, which is extremely close to the exact probability.*

Exercise 6.1 *Suppose the politician wanted to be within* $\pm 5\%$ *of the correct value, rather than* $\pm 10\%$*. That is, suppose she wished to double the precision of the poll. How could she accomplish this?*