12. THE BINOMIAL DISTRIBUTION

Eg: The top line on county ballots is supposed to be assigned by random drawing to either the Republican or Democratic candidate. The clerk of the county is supposed to make this random drawing. In Essex County, New Jersey, the county clerk (who incidentally is a Democrat) assigned the top line to the Democrats 40 out of 41 times. What is the probability that the Democrats would do so well in a truly random lottery?

("Ruling in Jersey upholds idea of equal odds for all", New York Times, Aug 13, 1985).

Eg: Dave interviews with 3 firms. The probability that he will get an offer from a given firm is 0.7. The decision of each firm is independent of the others. What is the probability that Dave will get exactly two offers? At least one offer? No offers?

Eg: Suppose you play five rounds of the “Let’s Make a Deal” game, switching each time. What is the probability that you lose all five rounds? How does the answer change if you stick with your original choice each time?

Eg: In a large batch of computer chips, 0.5% are defective. You randomly select 100 for inspection. How many bad chips will you get? How likely are you to find at least one bad chip?

Eg: If an airline deliberately overbooks a flight, selling 400 seats on a flight having a capacity of 350, and if the probability of a given passenger not showing up for the flight is .15, how many passengers will show up? What is the probability that some passengers will be denied boarding on the flight?

Eg: What is the probability that an analyst will correctly forecast the direction of the Dow for at least 7 of the next 10 trading days, assuming that the Dow is a random walk, and therefore changes in the Dow are actually not forecastable?

• In each of these examples, we have an experiment consisting of a fixed number, n, of "trials," or repetitions.

There are two possible outcomes at each trial, which may be denoted by "success" and "failure".

The results of different trials are independent.

The probability (p) of success on a given trial remains constant for all trials.

We are interested in the distribution of the random variable

X = Total number of successes.

The possible values for X are $x = 0, 1, \cdots, n$.

Under the above conditions, X is said to have a **binomial distribution**, and the probability of x successes in n trials is

$$p(x) = \binom{n}{x} p^x q^{n-x}$$

$x = 0, 1, \cdots, n$,
where $q = 1 - p$.

• Why is this formula correct?
Consider a sequence such as FSFFSFSS, where S and F denote "Success" and "Failure". This particular sequence had $n = 8$ trials, and $x = 4$ successes.

There are $\binom{n}{x}$ sequences of n trials containing exactly x successes. (This is the number of ways of selecting x of the trials to be successes.)

The probability of any particular such sequence is $p^x q^{n-x}$

Therefore, we get the above formula for $p(x)$.

Eg: Toss a coin 3 times (n = 3, p = 1/2).

x	Sequences With x Heads	Number of Sequences	$\binom{n}{x}$	Prob{x Heads}	$\binom{n}{x} p^x q^{n-x}$
0	TTT	1	1	1/8	1/8
1	TTH THT HTT	3	3	3/8	3/8
2	THH HTH HHT	3	3	3/8	3/8
3	HHH	1	1	1/8	1/8

Eg: In the Dave example, we have n = 3, p = .7, so that

$$p(0) = (.3)^3 = .027, \quad p(1) = 3(.7)^1(.3)^2 = .189,$$
$$p(2) = 3(.7)^2(.3)^1 = .441, \quad p(3) = (.7)^3 = .343 \ .$$

There is a 44.1% chance that Dave gets two offers.

He gets at least one offer with probability $p(1)+p(2)+p(3)=.973$.

There is just a 2.7% chance that he will get no offers.

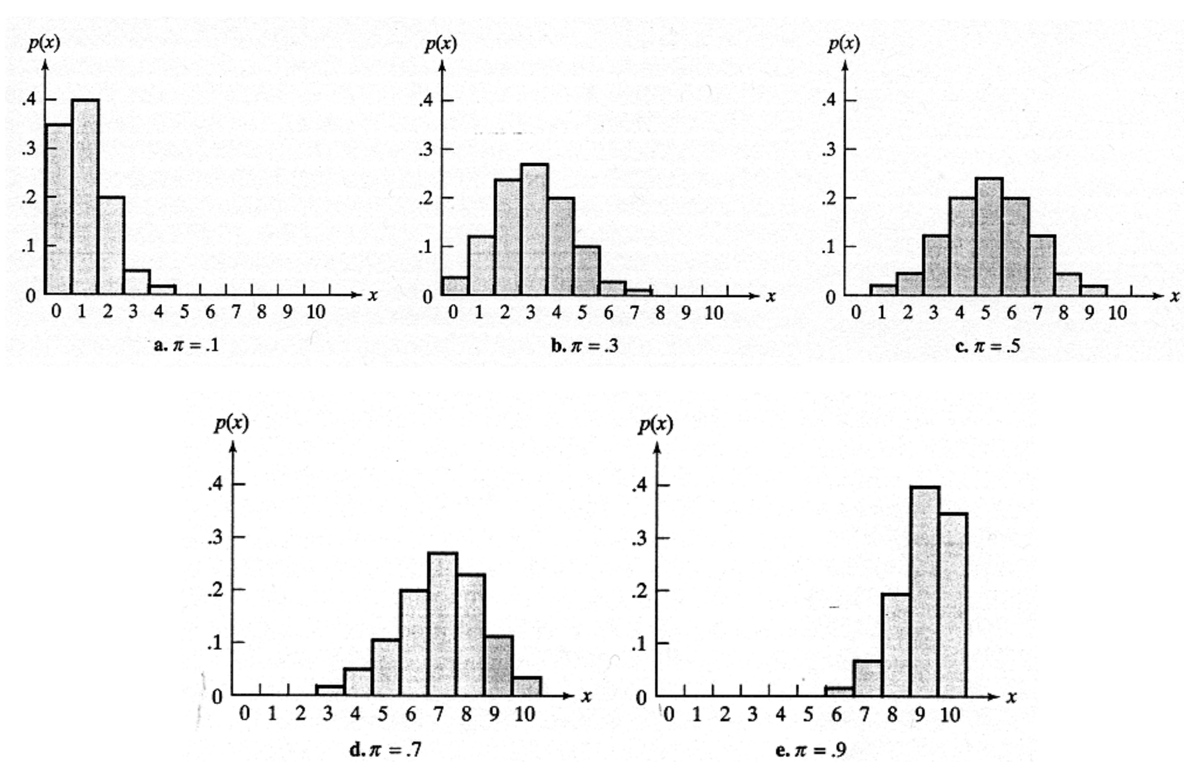
Eg: In tossing a coin 10 times, you are much more likely to get 5 heads than 0 heads. The reason is that there are many more ways to get 5 heads than 0 heads (252, compared to 1).

Note: $\binom{n}{x}$ = (10)(9)(8)(7)(6) / [(5)(4)(3)(2)(1)] = 252.

• Each pair of values (n, p) determines a distinct binomial distribution.

The shape of a binomial distribution:

$p < 0.5$	skewed to the right
$p = 0.5$	symmetric
$p > 0.5$	skewed to the left



• Table II, Appendix B gives the cumulative probabilities for the binomial distribution,

$$\Pr(X \le k) = \sum_{x=0}^{k} p(x).$$

$n = 10$

	p												
k	.01	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95	.99
0	9044	.5987	.3487	.1074	.0282	.0060	.0010	.0001	.0000	.0000	.0000	.0000	.0000
1	.9957	.9139	.7361	.3758	.1493	.0464	.0107	.0017	.0001	.0000	.0000	.0000	.0000
2	.9999	.9885	.9298	.6778	.3828	.1673	.0547	.0123	.0016	.0001	.0000	.0000	.0000
3	1.0000	.9990	.9872	.8791	.6496	.3823	.1719	.0548	.0106	.0009	.0000	.0000	.0000
4	1.0000	.9999	.9984	.9672	.8497	.6331	.3770	.1662	.0473	.0064	.0001	.0000	.0000
5	1.0000	1.0000	.9999	.9936	.9527	.8338	.6230	.3669	.1503	.0328	.0016	.0001	.0000
6	1.0000	1.0000	1.0000	.9991	.9894	.9452	.8281	.6177	.3504	.1209	.0128	.0010	.0000
7	1.0000	1.0000	1.0000	.9999	.9984	.9877	.9453	.8327	.6172	.3222	.0702	.0115	.0001
8	1.0000	1.0000	1.0000	1.0000	.9999	.9983	.9893	.9536	.8507	.6242	.2639	.0861	.0043
9	10000	1.0000	1.0000	1.0000	1.0000	.9999	.9990	.9940	.9718	.8926	.6513	.4013	.0956

(continued)

Eg: Compute the probability of getting at least 7 correct forecasts out of 10 for the direction of the Dow, assuming that the Dow is a random walk.

[Forecasting Lab Results]

Eg: A basketball foul shooter has been averaging 80% from the line. Assuming his skill stays the same, what is the probability that at least 10 of his next 15 foul shots will be good?

$n = 15$						p							
k	.01	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95	.99
0	.8601	.4633	.2059	.0352	.0047	.0005	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1	.9904	.8290	.5490	.1671	.0353	.0052	.0005	.0000	.0000	.0000	.0000	.0000	.0000
2	.9996	.9638	.8159	.3980	.1268	.0271	.0037	.0003	.0000	.0000	.0000	.0000	.0000
3	1.0000	.9945	.9444	.6482	.2969	.0905	.0176	.0019	.0001	.0000	.0000	.0000	.0000
4	1.0000	.9994	.9873	.8358	.5155	.2173	.0592	.0093	.0007	.0000	.0000	.0000	.0000
5	1.0000	.9999	.9978	.9389	.7216	.4032	.1509	.0338	.0037	.0001	.0000	.0000	.0000
6	1.0000	1.0000	.9997	.9819	.8689	.6098	.3036	.0950	.0152	.0008	.0000	.0000	.0000
7	1.0000	1.0000	1.0000	.9958	.9500	.7869	.5000	.2131	.0500	.0042	.0000	.0000	.0000
8	1.0000	1.0000	1.0000	.9992	.9848	.9050	.6964	.3902	.1311	.0181	.0003	.0000	.0000
9	1.0000	1.0000	1.0000	.9999	.9963	.9662	.8491	.5968	.2784	.0611	.0022	.0001	.0000
10	1.0000	1.0000	1.0000	1.0000	.9993	.9907	.9408	.7827	.4845	.1642	.0127	.0006	.0000
11	1.0000	1.0000	1.0000	1.0000	.9999	.9981	.9824	.9095	.7031	.3518	.0556	.0055	.0000
12	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9963	.9729	.8732	.6020	.1841	.0362	.0004
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9995	.9948	.9647	.8329	.4510	.1710	.0096
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9995	.9953	.9648	.7941	.5367	.1399

Eg: Logic analyzers come off the assembly line with a 3% defective rate. You must ship 17 of these analyzers tomorrow. How many analyzers should you schedule for production today in order to be reasonably sure that 17 or more of the scheduled machines will work?

Sol: If you schedule n machines, then X = #Working Machines has a binomial distribution with $p = 0.97$.

• If you schedule 17 machines (no margin for error), you might think that the high (97%) rate would help you, but in fact the probability that all 17 machines will work is just 0.596. So by leaving no margin for error, you are taking a 40.4% chance that you will fail to ship the entire order in working condition.

• If you schedule 18 machines, then

P(At least 17 machines work) is 0.900.

• If you schedule 19 machines, then

P(At least 17 machines work) is 0.982.

• Conclusion: You'd better schedule at least 19 machines to get 17 good ones.

Note: Since Table 2 doesn't cover these values of *n*, I computed the binomial probabilities by hand. Try it yourself!

Eg: For a "Best 4 of 7" series between two equally matched teams, the duration of the series has the following distribution:

Duration of Series	Probability
4	0.125
5	0.25
6	0.3125
7	0.3125

Let's show that Prob{Series Lasts 6 Games} = 0.3125.
(The proofs for the other cases are similar; you can try them yourself).

We'll calculate the probability that the series is won on the 6^{th} game by Team A, and then double this probability.

We have Prob{Series is won on 6^{th} game by Team A} =

Prob{Team A wins 3 of first 5 games}×Prob{Team A wins 6^{th} game}

$$= \frac{\binom{5}{3}}{32} \times \frac{1}{2} = \frac{10}{64} = 0.15625$$

where we used the fact that the number of wins for Team A in the first five games has a binomial distribution with $n = 5$, $p = 1/2$.

Therefore, Prob{Series lasts 6 games} = 2(0.15625) = 0.3125.

By the way, the distribution given above does not seem to fit the data for the Stanley Cup Finals, a Best 4 out of 7 championship series for the National Hockey League. See article in Chance, Vol 11, No. 1, by Morrison and Schmittlein.

Eg: Suppose we play the New York State 6/51 Lotto twice every week for 10 years, a total of 1040 tickets. The number of jackpot-winning tickets has a binomial distribution with $n = 1040$, $p = 1/18{,}009{,}460$. The probability of getting at least 1 jackpot-winning ticket is $1 - p(0) = 1 - (1 - p)^{1040} = .0000577$, about 1 in 20,000.

For Second Prize, (matching 5 of the 6 numbers plus bonus) we use $p = 1/3{,}001{,}577$ and the probability of winning at least once in 10 years becomes $1 - (1 - p)^{1040} = .00035$, about 1 in 3,000.

Mean and Variance of Binomial Random Variables

If X is binomial (n, p) then

$$E(X) = \mu = np,$$

$$Var(X) = \sigma^2 = npq.$$

Eg: For $n = 17$, $p = 0.97$, we get $\mu = 16.49$, $\sigma = 0.703$. The relatively large value of σ helps to "explain" why Prob(17 successes) is so low (0.596), despite the large value of p.

Eg: The manufacturer of peanut M&Ms claims that only one bag out of 10 will contain any defectives. We inspect 5 bags, and 3 contain defectives. Do we reject the manufacturer's claim? How many defective bags should be expected? With what variation?

[Taste Test Lab]

Eg: Susan Lucci, a longtime star on the ABC soap "All My Children" failed to win an Emmy Award 28 times, finally winning on the 29th nomination. Was she unfairly passed over for the award?

Let's assume that there were four nominees for each award, that she was as good as the other nominees (all of whom were equally good), and that the Emmy outcomes are independent, the chances of failing 28 times, followed by a success are

$$(3/4)^{28} \cdot (1/4) = 7.93 \cdot 10^{-5}, \text{ i.e, about 8 in 100,000.}$$

Even if we assume she was only half as good as the other nominees, the probability increases to

$$(6/7)^{28} \cdot (1/7) = 0.0019, \text{ about 2 in 1,000.}$$

It looks like she's not been given a fair shake!

TVG: After so many disappointments, did you allow yourself to believe that you still could win?
SL: No. Except for a little corner of me that was still hopeful, I couldn't really believe I would ever win.

TVG: Your husband once said that he was beginning to suspect there was a conspiracy of some sort when you lost year after year. Did you ever think you were being blackballed?
SL: I couldn't believe that, no. My peers in the industry nominated me.

TV GUIDE JUNE 19, 1999