

2. The Binomial Distribution

Basic Theory

Suppose that our [random experiment](#) is to perform a sequence of [Bernoulli trials](#) :

$$X = (X_1, X_2, \dots)$$

Recall that X is a sequence of independent indicator random variables with common probability of success p , the basic parameter of the process. In statistical terms, the first n trials (X_1, X_2, \dots, X_n) form a [random sample](#) of size n from the Bernoulli distribution. In this section we will study the [random variable](#) that gives the number of successes in the first n trials and the random variable that gives the proportion of successes in the first n trials. The underlying distribution, the binomial distribution, is one of the most important in probability theory.

- ❑ 1. Show that the number of successes in the first n trials is the random variable

$$Y_n = \sum_{i=1}^n X_i$$

The Density Function

- ❑ 2. Denote the index set for the first n trials by $N = \{1, 2, \dots, n\}$. Use the assumptions of Bernoulli trials to show that for $K \subseteq N$ with $\#(K) = k$,

$$\mathbb{P}(X_i = 1 \text{ for all } i \in K \text{ and } X_i = 0 \text{ for all } i \in N \setminus K) = p^k (1 - p)^{n-k}$$

Recall that the number of subsets of size k from a set of size n is the [binomial coefficient](#)

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

- ❑ 3. Use Exercise 2 and basic properties of probability to show that the probability density function of Y_n is given by

$$\mathbb{P}(Y_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k \in \{0, 1, \dots, n\}$$

The distribution with this probability density function is known as the **binomial distribution** with parameters n and p .

- ❑ 4. In the [binomial coin experiment](#), vary n and p with the scrollbars, and note the shape and location of the probability density function. For selected values of the parameters, run the simulation 1000 times,

updating every 10 runs. Note the apparent convergence of the relative frequency function to the density function.

5. Use the binomial theorem to show that the binomial probability density function really is a probability density function.

6. Show that

- $\mathbb{P}(Y_n = k) > \mathbb{P}(Y_n = k - 1)$ if and only if $k < (n + 1)p$
- $\mathbb{P}(Y_n = k) = \mathbb{P}(Y_n = k - 1)$ if and only if $k = (n + 1)p$, an integer between 1 and n .

Thus, the density function at first increases and then decreases, reaching its maximum value at $\lfloor (n + 1)p \rfloor$. This integer is a *mode* of the distribution. In the case that $m = (n + 1)p$ is an integer between 1 and n , there are two consecutive modes, at $m - 1$ and m . In any event, the shape of the binomial distribution is **unimodal**.

7. Suppose that U is a random variable having the binomial distribution with parameters n and p . Show that $n - U$ has the binomial distribution with parameters n and $1 - p$.

- Give a probabilistic proof, based on Bernoulli trials
- Give an analytic proof, based on probability density functions

Moments

We will compute the **mean** and **variance** of the binomial distribution several different ways. The method using indicator variables is the best.

8. Show that $\mathbb{E}(Y_n) = np$ in two ways:

- using the probability density function.
- using the representation as a sum of indicator variables.

This result makes intuitive sense, since p should be approximately the proportion of successes in a large number of trials. We will discuss the point further in the subsection on the **proportion of successes**.

9. Show that $\text{var}(Y_n) = np(1 - p)$ in two ways:

- using the probability density function
- using the representation as a sum of independent indicator variables.

10. Sketch the graph of the variance as a function of p . Note in particular that the variance takes its maximum value $\frac{1}{4}$ when $p = \frac{1}{2}$ and its minimum value 0 when $p = 0$ or $p = 1$.

11. In the **binomial coin experiment**, vary n and p with the scrollbars and note the location and size of the

mean/standard deviation bar. For selected values of the parameters, run the simulation 1000 times, updating every 10 runs. Note the apparent convergence of the sample mean and standard deviation to the distribution mean and standard deviation.

12. Show that the **probability generating function** is $\mathbb{E}(t^{Y_n}) = ((1-p) + pt)^n$ for $t \in \mathbb{R}$ in two ways:

- using the probability density function
- using the representation as a sum of independent indicator variables.

13. Use the probability generating function to compute the mean and variance.

14. Use the identity $j \binom{n}{j} = n \binom{n-1}{j-1}$ for $n \in \mathbb{N}_+$ and $j \in \mathbb{N}_+$ to show that

$$\mathbb{E}(Y_n^k) = n p \mathbb{E}((Y_{n-1} + 1)^{k-1}); \quad n \in \mathbb{N}_+, k \in \mathbb{N}_+$$

15. Use the recursion result in the previous exercise to give yet one more derivation of the mean and variance.

The Proportion of Successes

The **proportion of successes** in the first n trials is the random variable

$$M_n = \frac{Y_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

In statistical terms, M_n is the **sample mean** of the random sample (X_1, X_2, \dots, X_n) . The proportion of successes M_n is typically used to estimate the probability of success p when this probability is unknown. It is basic to the very notion of probability, that if the number of trials is large, then M_n should be close to p . The mathematical formulation of this idea is a special case of the **law of large numbers**.

It is easy to express the probability density function of the proportion of successes M_n in terms of the probability density function of the number of successes Y_n . First, note that M_n takes the values $\frac{k}{n}$ where $k \in \{0, 1, \dots, n\}$.

16. Show that

$$\mathbb{P}\left(M_n = \frac{k}{n}\right) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \dots, n\}$$

17. In the **binomial coin experiment**, select the proportion of heads. Vary n and p with the scroll bars and note the shape of the probability density function. For selected values of the parameters, run the experiment 1000 times, updating every 10 runs. Watch the apparent convergence of the relative frequency function to the probability density function.

18. Show that $\mathbb{E}(M_n) = p$. In statistical terms, this means that M_n is an **unbiased estimator** of p .

19. Show that $\text{var}(M_n) = \frac{p(1-p)}{n}$. In particular, $\text{var}(M_n) \leq \frac{1}{4n}$ for any $p \in [0, 1]$.

Note that $\text{var}(M_n) \rightarrow 0$ as $n \rightarrow \infty$ (and actually, the convergence is uniform in $p \in [0, 1]$). Thus, the estimate improves as n increases; in statistical terms, this is known as **consistency**.

20. Use the result in the last exercise and [Chebyshev's inequality](#) to show that $\mathbb{P}(|M_n - p| \geq \delta) \rightarrow 0$ as $n \rightarrow \infty$ for every $\delta > 0$. In fact, the convergence is uniform in $p \in [0, 1]$.

The result in the last exercise is a special case of the [weak law of large numbers](#) and means that $M_n \rightarrow p$ as $n \rightarrow \infty$ in probability. The [strong law of large numbers](#) states that the convergence actually holds with probability 1. See [Estimation in the Bernoulli Model](#) in the chapter on [Set Estimation](#) for a different approach to the problem of estimating p .

21. In the [binomial coin experiment](#), select the proportion of heads. Vary n and p and note the size and location of the mean/standard deviation bar. For selected values of the parameters, run the experiment 10000 times, updating every 10 runs, and note the apparent convergence of the empirical moments to the distribution moments.

Sums of Independent Binomial Variables

Several important properties of the random process $Y = (Y_1, Y_2, \dots)$ stem from the fact that it is a [partial sum process](#) corresponding to the sequence X of independent, identically distributed indicator variables.

22. Use the representation in terms of indicator variables to prove the following properties.

- If m and n are positive integers with $m \leq n$ then $Y_n - Y_m$ has the same distribution as Y_{n-m} , namely binomial with parameters $n - m$ and p . Thus, the process Y has **stationary increments**.
- If $n_1 \leq n_2 \leq n_3 \leq \dots$ then $(Y_{n_1}, Y_{n_2} - Y_{n_1}, Y_{n_3} - Y_{n_2}, \dots)$ is a sequence of independent random variables. Thus, the process Y has **independent increments**.

Actually, any partial sum process corresponding to an independent, identically distributed sequence will have stationary, independent increments.

23. Show that if U and V are independent random variables for an experiment, and that U has the binomial distribution with parameters m and p , and V has the binomial distribution with parameters n and p , then $U + V$ has the binomial distribution with parameters $m + n$ and p .

- Give a probabilistic proof using the previous exercise.
- Give an analytic proof using probability density functions.
- Give an analytic proof using probability generating functions.

24. Use the stationary, independent increments properties to find the joint probability density functions

of the Y process: if $n_1 < n_2 < \dots < n_k$ and $j_1 \leq j_2 \leq \dots \leq j_k$ then

$$\mathbb{P}(Y_{n_1} = j_1, Y_{n_2} = j_2, \dots, Y_{n_k} = j_k) = \binom{n_1}{j_1} \binom{n_2 - n_1}{j_2 - j_1} \dots \binom{n_k - n_{k-1}}{j_k - j_{k-1}} p^{j_k} (1-p)^{n_k - j_k}$$

where as always, we use the standard conventions for binomial coefficients: $\binom{a}{b} = 0$ if $b < 0$ or $b > a$.

Connection to the Hypergeometric Distribution

25. Suppose that $m \leq n$. Show that

$$\mathbb{P}(Y_m = j | Y_n = k) = \frac{\binom{m}{j} \binom{m-n}{k-j}}{\binom{n}{k}}, \quad j \in \{0, 1, \dots, n\}$$

Interestingly, the distribution in the last exercise is independent of p . It is known as the [hypergeometric distribution](#) with parameters n , m and k , and governs the number of type 1 objects in a sample of size k chosen at random and without replacement from a population of n objects of which m are type 1 (and the remainder type 0). Try to interpret this result probabilistically.

The Normal Approximation

26. Open the [binomial timeline experiment](#). For selected values of $p \in (0, 1)$, start with $n = 1$ and successively increase n by 1. For each value of n , Note the shape of the probability density function of the number of successes and the proportion of successes. With $n = 100$, run the experiment 1000 time, updating by 10. For the number of successes and the proportion of successes, note the apparent convergence of the empirical density function to the probability density function.

The characteristic bell shape that you should observe in the previous exercise is an example of the [central limit theorem](#), because the binomial variable can be written as a sum of n independent, identically distributed random variables (the indicator variables).

27. Show that the standard score of Y_n is the same as the standard score of M_n :

$$\frac{Y_n - np}{\sqrt{np(1-p)}} = \frac{M_n - p}{\sqrt{\frac{p(1-p)}{n}}}$$

28. Apply the central limit theorem to show that the distribution of the standard score given in the previous exercise converges to the standard [normal distribution](#) as $n \rightarrow \infty$:

This version of the central limit theorem is known as the [DeMoivre-Laplace theorem](#), and is named after [Abraham DeMoivre](#) and [Simeon Laplace](#). From a practical point of view, this result means that, for large n , the distribution of Y_n is approximately normal, with mean np and standard deviation $\sqrt{np(1-p)}$ and the

distribution of M_n is approximately normal, with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$. Just how large n needs to be for the normal approximation to work well depends on the value of p . The rule of thumb is that we need $np \geq 5$ and $n(1-p) \geq 5$. Finally, when using the normal approximation, we should remember to use the [continuity correction](#), since the binomial is a discrete distribution.

Examples and Applications

29. A student takes a multiple choice test with 20 questions, each with 5 choices (only one of which is correct). Suppose that the student blindly guesses. Let X denote the number of questions that the student answers correctly.

- Find the probability density function of X .
- Find the mean of X .
- Find the variance of X .
- Find the probability that the student answers at least 12 questions correctly (the score that she needs to pass).



30. A certain type of missile has failure probability 0.02. Let Y denote the number of failures in 50 tests.

- Find the probability density function of Y .
- Find the mean of Y .
- Find the variance of Y .
- Find the probability of at least 47 successful tests.



31. Suppose that in a certain district, 40% of the registered voters prefer candidate A. A random sample of 50 registered voters is selected. Let Z denote the number in the sample who prefer A.

- Find the probability density function of Z .
- Find the mean of Z .
- Find the variance of Z .
- Find the probability that Z is less than 19.
- Compute the normal approximation to the probability in (d).



Coins and Dice

Recall that a **standard** die is a six-sided die. A **fair** die is one in which the faces are equally likely. An **ace-six flat** die is a standard die in which faces 1 and 6 have probability $\frac{1}{4}$ each, and faces 2, 3, 4, and 5 have

probability $\frac{1}{8}$.

32. A standard, fair die is tossed 10 times. Let N denote the number of aces.

- Find the probability density function of N
- Find the mean of N .
- Find the variance of N .



33. A coin is tossed 100 times and results in 30 heads. Find the probability density function of the number of heads in the first 20 tosses.



34. An ace-six flat die is rolled 1000 times. Let Z denote the number of times that a score of 1 or 2 occurred.

- Find the probability density function of Z
- Find the mean of Z .
- Find the variance of Z .
- Find the probability that Z is at least 400.
- Find the normal approximation of the probability in (d)



35. In the **binomial coin experiment**, select the proportion of heads. Set $n = 10$ and $p = 0.4$. Run the experiment 100 times, updating after each run. Over all 100 runs, compute the square root of the average of the squares of the errors, when M_n used to estimate p . This number is a measure of the quality of the estimate.

36. In the **binomial coin experiment**, select the number of heads Y , and set $p = 0.5$ and $n = 15$. Run the experiment 1000 times with an update frequency of 100. Compute and compare the following:

- $\mathbb{P}(5 \leq Y \leq 10)$
- The relative frequency of the event $\{5 \leq Y \leq 10\}$
- The normal approximation to $\mathbb{P}(5 \leq Y \leq 10)$



37. In the **binomial coin experiment**, select the proportion of heads M and set $n = 30$, $p = 0.6$. Run the experiment 1000 times with an update frequency of 100. Compute and compare each of the following:

- $\mathbb{P}(0.5 \leq M \leq 0.7)$
- The relative frequency of the event $\{0.5 \leq M \leq 0.7\}$
- The normal approximation to $\mathbb{P}(0.5 \leq M \leq 0.7)$



Famous Problems

In 1693, **Samuel Pepys** asked **Isaac Newton** whether it is more likely to get at least one ace in 6 rolls of a die or at least two aces in 12 rolls of a die. This problem is known as a **Pepys' problem**; naturally, Pepys had fair dice in mind.

38. Guess the answer to Pepys' problem based on empirical data. With fair dice and $n = 6$, run the **simulation of the dice experiment** 500 times and compute the relative frequency of at least one ace. Now with $n = 12$, run the simulation 500 times and compute the relative frequency of at least two aces. Compare the results.

39. Solve Pepys' problem using the binomial distribution.



40. Which is more likely: at least one ace with 4 throws of a fair die or at least one double ace in 24 throws of two fair dice? This is known as **DeMere's problem**, named after **Chevalier De Mere**.



Data Analysis Exercises

41. In the **cicada data**, compute the proportion of males in the entire sample, and the proportion of males of each species in the sample.



42. In the **M&M data**, pool the bags to create a large sample of M&Ms. Now compute the sample proportion of red M&Ms.



The Galton Board

The **Galton board** is a triangular array of pegs. The rows are numbered by the natural numbers $\mathbb{N} = \{0, 1, \dots\}$ from top downward. Row n has $n + 1$ pegs numbered from left to right by the integers $\{0, 1, \dots, n\}$. Thus a peg can be uniquely identified by the ordered pair (n, k) where n is the row number and k is the peg number in that row. The Galton board is named after **Francis Galton**.

Now suppose that a ball is dropped from above the top peg $(0, 0)$. Each time the ball hits a peg, it bounces to the right with probability p and to the left with probability $1 - p$, independently from bounce to bounce.

43. Show that the number of the peg that the ball hits in row n has the binomial distribution with parameters n and p .

44. In the **Galton board experiment**, select random variable Y (the number of moves right). Vary the

parameters n and p and note the shape and location of the probability density function and the mean/standard deviation bar. For selected values of the parameters, click single step several times and watch the ball fall through the pegs. Then run the experiment 1000 times, updating every 10 runs and watch the path of the ball. Note the apparent convergence of the relative frequency function and empirical moments to the density function and distribution moments, respectively.

Reliability

Recall the discussion of **structural reliability** given in the last section on **Bernoulli trials**. In particular, we have a system of n similar components that function independently, each with reliability p . Suppose now that the system as a whole functions properly if and only if at least k of the n components are good. Such a system is called, appropriately enough, a **k out of n system**. Note that the **series** and **parallel** systems considered in the previous section are n out of n and 1 out of n systems, respectively.

45. Show that for a k out of n system

- The state of the system is $\mathbf{1}(Y_n \geq k)$ where Y_n is the number of working components.
- The reliability function is $r_{n,k}(p) = \sum_{i=k}^n \binom{n}{i} p^i (1-p)^{n-i}$

46. In the **binomial coin experiment**, set $n = 10$ and $p = 0.9$ and run the simulation 1000 times, updating every 100 runs. Compute the empirical reliability and compare with the true reliability in each of the following cases:

- 10 out of 10 (series) system.
- 1 out of 10 (parallel) system.
- 4 out of 10 system.

47. Consider a system with $n = 4$ components. Sketch the graphs of $r_{4,1}$, $r_{4,2}$, $r_{4,3}$, and $r_{4,4}$, on the same set of axes.

48. An n out of $2n - 1$ system is a **majority rules** system.

- Compute the reliability of a 2 out of 3 system.
- Compute the reliability of a 3 out of 5 system
- For what values of p is a 3 out of 5 system more reliable than a 2 out of 3 system?
- Sketch the graphs of $r_{3,2}$ and $r_{5,3}$, on the same set of axes.



49. In the **binomial coin experiment**, compute the empirical reliability, based on 100 runs, in each of the following cases. Compare your results to the true probabilities.

- A 2 out of 3 system with $p = 0.3$
- A 3 out of 5 system with $p = 0.3$

- c. A 2 out of 3 system with $p = 0.8$
- d. A 3 out of 5 system with $p = 0.8$

50. Show that $r_{2n-1,n}\left(\frac{1}{2}\right) = \frac{1}{2}$.

Bernstein Polynomials

The **Weierstrass Approximation Theorem**, named after **Karl Weierstrass**, states that any real-valued function that is continuous on a closed, bounded interval can be uniformly approximated on that interval, to any degree of accuracy, with a polynomial. The theorem is important, since polynomials are simple and basic functions, and a bit surprising, since continuous functions can be quite strange.

In 1911, **Sergi Bernstein** gave an explicit construction of polynomials that uniformly approximate a given continuous function, using Bernoulli trials. Bernstein's result is a beautiful example of the **probabilistic method**, the use of probability theory to obtain results in other areas of mathematics that are seemingly unrelated to probability,

Suppose that f is a real-valued function that is continuous on the interval $[0, 1]$. The **Bernstein polynomial** of degree n for f is defined by

$$b_n(p) = \mathbb{E}_p(f(M_n)), \quad p \in [0, 1]$$

where M_n is the proportion of successes in the first n Bernoulli trials with success parameter p , as defined earlier. Note that we are emphasizing the dependence on p in the expected value operator. The next exercise gives a more explicit representation, and shows that the Bernstein polynomial is, in fact, a polynomial

51. Use basic properties of expected value to show that

$$b_n(p) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} p^k (1-p)^{n-k}, \quad p \in [0, 1]$$

52. In particular, show that

- a. $b_n(0) = f(0)$ and $b_n(1) = f(1)$. Thus, the graph of b_n passes through the endpoints $(0, f(0))$ and $(1, f(1))$
- b. $b_1(p) = f(0) + (f(1) - f(0))p$, $p \in [0, 1]$. Thus, the graph of b_1 is the line connecting the endpoints.
- c. $b_2(p) = f(0) + 2\left(f\left(\frac{1}{2}\right) - f(0)\right)p + \left(f(1) - 2f\left(\frac{1}{2}\right) + f(0)\right)p^2$, $p \in [0, 1]$. Thus, the graph of b_2 is the parabola passing through the endpoints and the point $\left(\frac{1}{2}, \frac{1}{4}f(0) + \frac{1}{2}f\left(\frac{1}{2}\right) + \frac{1}{4}f(1)\right)$.

53. Prove Bernstein's theorem: $b_n \rightarrow f$ as $n \rightarrow \infty$ uniformly on $[0, 1]$. Fill in the details of the major steps below:

- a. Since f is continuous on the closed, bounded interval $[0, 1]$, it is bounded on this interval. Thus, there

exists a constant C such that $|f(p)| \leq C$ for all $p \in [0, 1]$

- b. Since f is continuous on the closed, bounded interval $[0, 1]$, it is uniformly continuous on this interval. Thus, for any $\varepsilon > 0$ there exists $\delta > 0$ such that if $p \in [0, 1]$, $q \in [0, 1]$, and $|p - q| < \delta$ then $|f(p) - f(q)| < \varepsilon$
- c. From basic properties of expected value,
 $|b_n(p) - f(p)| \leq \mathbb{E}_p(|f(M_n) - f(p)|, |M_n - p| < \delta) + \mathbb{E}_p(|f(M_n) - f(p)|, |M_n - p| \geq \delta)$
- d. Using parts (a) and (b), $|b_n(p) - f(p)| \leq \varepsilon + 2C P_p(|M_n - p| \geq \delta)$ for any $p \in [0, 1]$.
- e. By [Exercise 20](#), $P_p(|M_n - p| \geq \delta) \rightarrow 0$ as $n \rightarrow \infty$ uniformly in $p \in [0, 1]$.

54. Compute the Bernstein polynomials of orders 1, 2, and 3 for the function f defined by $f(x) = \cos(\pi x)$, $x \in [0, 1]$. Graph f and the three polynomials on the same set of axes.



55. Use a computer algebra system to compute the Bernstein polynomials of orders 10, 20, and 30 for the function f defined below. Use the CAS to graph the function and the three polynomials on the same axes.

$$f(x) = \begin{cases} 0, & x = 0 \\ x \sin\left(\frac{\pi}{x}\right), & x \in (0, 1] \end{cases}$$

[Virtual Laboratories](#) > [11. Bernoulli Trials](#) > [1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

[Contents](#) | [Applets](#) | [Data Sets](#) | [Biographies](#) | [External Resources](#) | [Key words](#) | [Feedback](#) | [©](#)