

Probability and the Binomial Distribution

Introduction to Probability

A probability is a numerical quantity that expresses the likelihood of an event. The probability of an event E is written as $P(E)$. The probability $P(E)$ is always a number between 0 and 1, inclusive.

Example: coin tossing

Consider the familiar chance operation of tossing a coin, and define the event E : Heads. Each time the coin is tossed, either it falls heads or it does not. If the coin is equally likely to fall heads or tails, then

$$P(E) = \frac{1}{2} = 0.5.$$

Such an ideal coin is called a "fair" coin. If the coin is not fair (perhaps because it is slightly bent), then $P(E)$ will be some value other than 0.5, for instance, $P(E) = 0.6$.

Example: sampling Fruitflies

A large population of the fruitfly *Drosophila melanogaster* is maintained in a lab. In the population, 30% of the individuals are black because of a mutation, while 70% of the individuals have the normal gray body color. Suppose one fly is chosen at random from the population. Then the probability that a black fly is chosen is 0.3. More formally, define E : Sampled fly is black. Then $P(E) = 0.3$.

The preceding example illustrates the basic relationship between probability and random sampling: *The probability that a randomly chosen individual has a certain characteristic is equal to the proportion of population members with the characteristic.*

Frequency interpretation of probability

- The probability of an event E is meaningful only in relation to a chance operation that can in principle be repeated indefinitely often.
- Each time the chance operation is repeated, the event E either occurs or does not occur.
- The probability $P(E)$ is interpreted as the relative frequency of occurrence of E in an indefinitely long series of repetitions of the chance operation.

$$\frac{\# \text{ of times } E \text{ occurs}}{\# \text{ of times chance operation is repeated}} \rightarrow P(E),$$

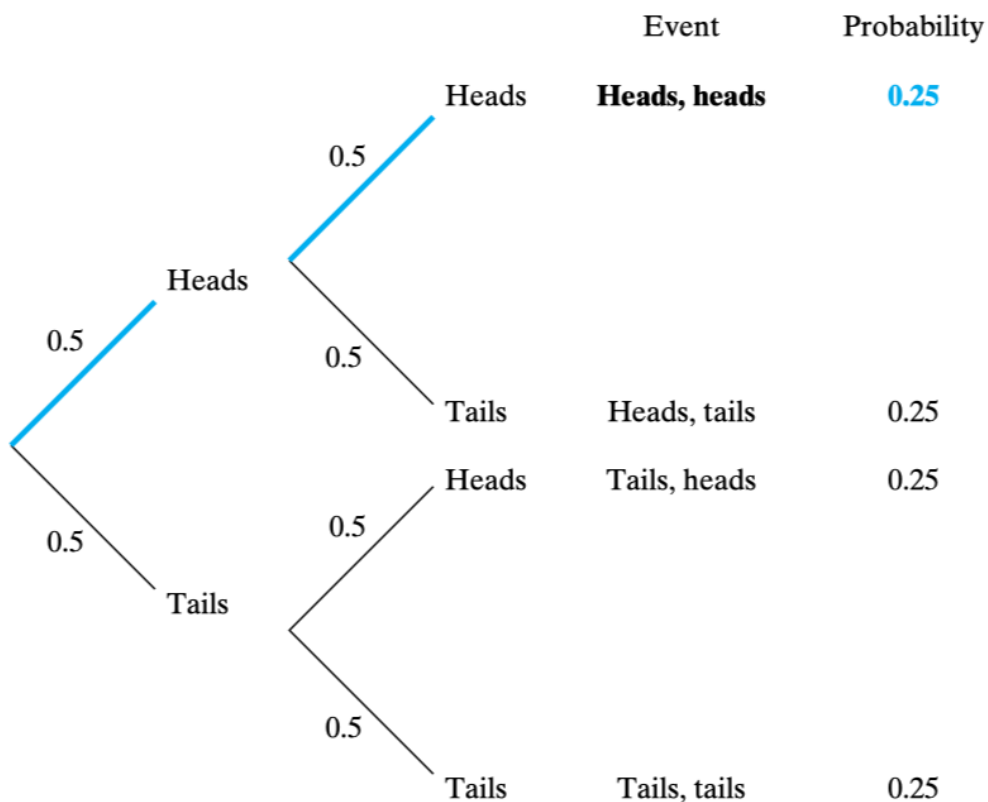
where the arrow indicates that if the chance operation is repeated an unlimited number of times, the two sides of the expression will be approximately equal.

Recall the coin tossing example. Suppose that a fair coin is tossed and the number of heads is recorded. One expects that

$$\frac{\# \text{ of heads}}{\# \text{ of tosses}} \rightarrow 0.5.$$

How about the probability of the event that both tosses are heads when this coin is tossed twice?

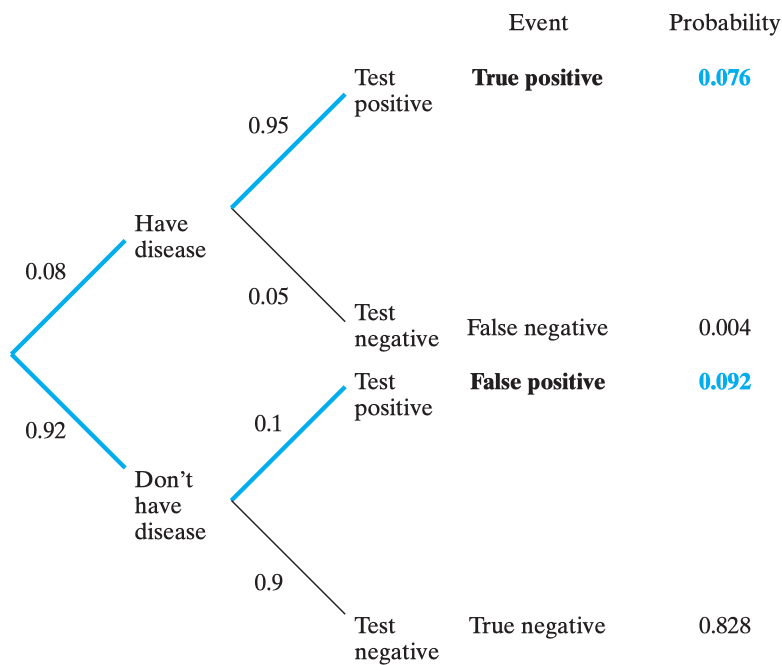
Probability tree



- What is the probability of the event that the first toss is head and the second toss is tail?
- What is the probability of the event that the first toss is tail?

Example: medical testing

Suppose a medical test is conducted on someone to try to determine whether or not the person has a particular disease.



		Predicted condition	
		Positive (PP)	Negative (PN)
Actual condition	Total population = P + N		
	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

- Sensitivity:

$$\frac{TP}{TP + FN}$$

- Specificity:

$$\frac{TN}{TN + FP}$$

Probability Rules

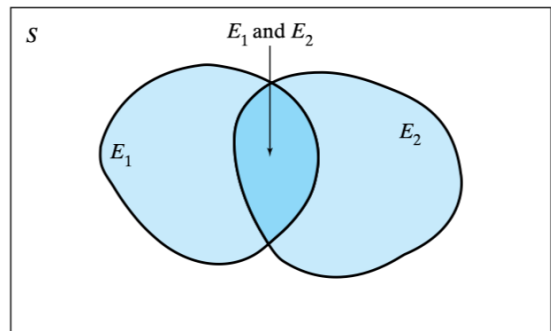
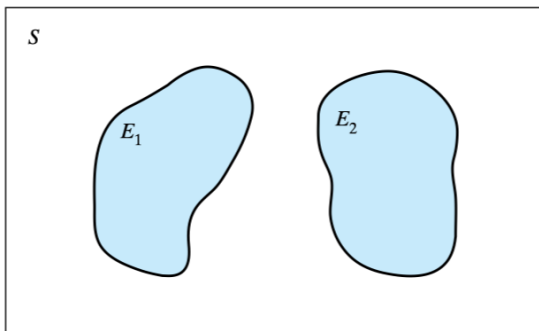
- Rule (1) The probability of an event E is always between 0 and 1. That is, $0 \leq P(E) \leq 1$.
- Rule (2) The sum of the probabilities of all possible events equals 1. That is, if the set of all possible disjoint events is E_1, E_2, \dots, E_k , then $\sum_{i=1}^k P(E_i) = 1$.
- Rule (3) The probability that an event E does not happen, denoted by E^C , is one minus the probability that the event happens. That is, $P(E^C) = 1 - P(E)$. (We refer to E^C

as the complement of E .)

Example: blood types

In the United States, 44% of the population has type O blood, 42% has type A, 10% has type B, and 4% has type AB. Consider choosing someone at random and determining the person's blood type. The probability of a given blood type will correspond to the population percentage.

- The probability that the person will have type O blood = $P(O) = 0.44$.
- $P(O) + P(A) + P(B) + P(AB) = 0.44 + 0.42 + 0.10 + 0.04 = 1$.
- The probability that the person will not have type O blood = $P(O^C) = 1 - 0.44 = 0.56$. This could also be found by adding the probabilities of the other blood types:
 $P(O^C) = P(A) + P(B) + P(AB) = 0.42 + 0.10 + 0.04 = 0.56$.
- We say that two events are **disjoint (mutually exclusive)** if they cannot occur simultaneously. Recall the blood types example.
- The **union** (\cup) of two events is the event that one or the other occurs or both occur.
- The **intersection** (\cap) of two events is the event that they both occur.



Addition rules

- Rule (4) If two events E_1 and E_2 are disjoint, then

$$P(E_1 \cup E_2) = P(E_1) + P(E_2).$$

- Rule (5) For any two events E_1 and E_2 ,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2).$$

Example: hair color and eye color

The following table shows the relationship between hair color and eye color for a group of 1,770 German men.

Table 3.3.1 Hair color and eye color					
		Hair color			Total
		Brown	Black	Red	
Eye color	Brown	400	300	20	720
	Blue	800	200	50	1,050
	Total	1,200	500	70	1,770

Calculate the probabilities of the following events:

- black hair
 - blue eyes
 - black hair or red hair: rule (4)
 - black hair or blue eyes: rule (5)
-
- Two events are said to be **independent** if knowing that one of them occurred does not change the probability of the other one occurring.
 - For example, if a coin is tossed twice, the outcome of the second toss is independent of the outcome of the first toss, since knowing whether the first toss resulted in heads or in tails does not change the probability of getting heads on the second toss.
 - Events that are not independent are said to be **dependent**.
 - When events are dependent, we need to consider the conditional probability of one event, given that the other event has happened. We use the notation $P(E_2|E_1)$ to represent the probability of E_2 happening, given that E_1 happened.

Conditional probability

The conditional probability of E_2 , given E_1 , is

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)}$$

provided that $P(E_1) > 0$.

Consider the hair color and eye color example, what is the probability of the man having blue eyes given that he has black hair?

Multiplication rules

- Rule (6) If two events E_1 and E_2 are independent, then

$$P(E_1 \cap E_2) = P(E_1) \times P(E_2).$$

- Rule (7) For any two events E_1 and E_2 ,

$$P(E_1 \cap E_2) = P(E_1) \times P(E_2|E_1).$$

- Recall the coin tossing example, use rule (6) to calculate the probability of getting heads on both tosses.
- Consider the hair color and eye color example, what is the probability of the man having red hair and brown eyes?

Rules of total probability

- Rule (8) For any two events E_1 and E_2 ,

$$P(E_1) = P(E_2) \times P(E_1|E_2) + P(E_2^C) \times P(E_1|E_2^C).$$

Consider the hair color and eye color example and two events E_1 : red hair, E_2 : brown eyes. Verify rule (8).

Density Curves

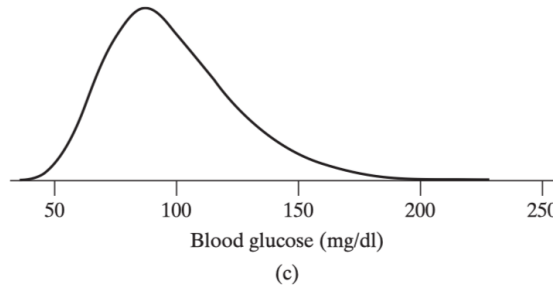
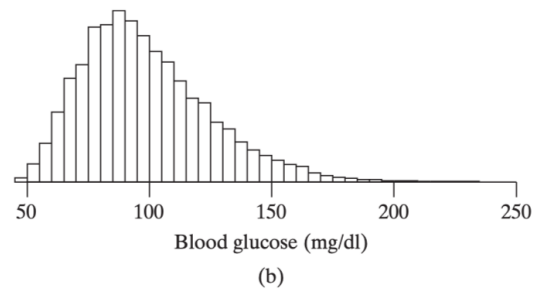
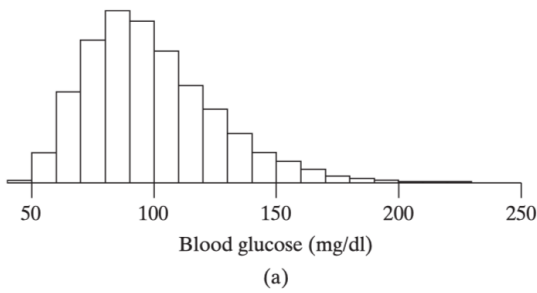
The examples presented in the previous section dealt with probabilities for discrete variables. In this section we will consider probability when the variable is continuous

Relative frequency histograms and density curves

- A relative frequency histogram is a histogram in which we indicate the proportion (i.e., the relative frequency) of observations in each category, rather than the count of observations in the category.
- We can think of the relative frequency histogram as an approximation of the underlying true population distribution from which the data came.
- We may visualize the density curve as an idealization of a relative frequency histogram with very narrow classes.

Example: blood glucose

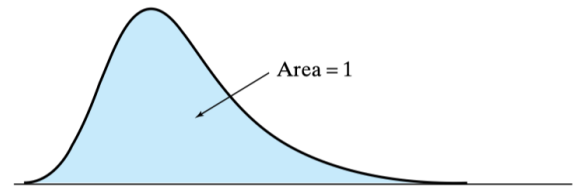
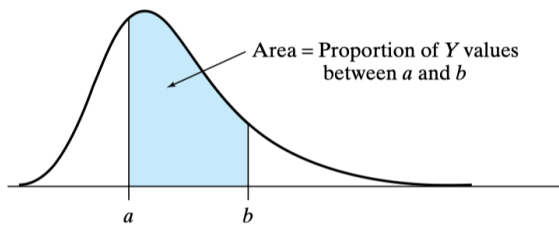
A glucose tolerance test can be useful in diagnosing diabetes. The blood level of glucose is measured one hour after the subject has drunk 50 mg of glucose dissolved in water. The distribution is represented by histograms with class widths equal to (a) 10 and (b) 5, and by (c) a smooth curve.



A smooth curve representing a frequency distribution is called a **density curve**.

For any two numbers a and b ,

Area under density curve between a and b = Proportion of Y values between a and b

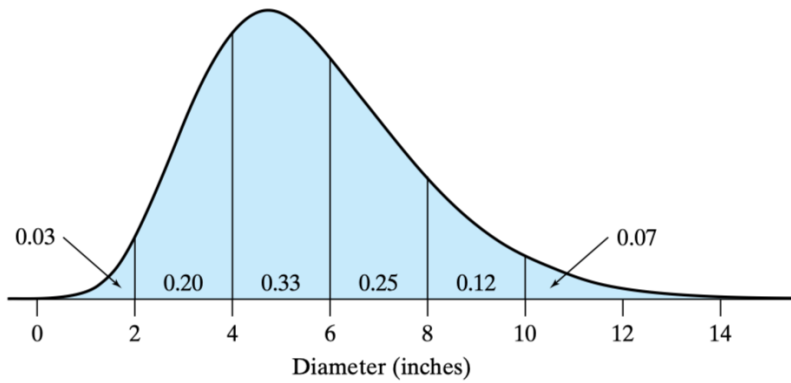


- If a variable has a continuous distribution, then we find probabilities by using the density curve for the variable.
- A probability for a continuous variable equals the area under the density curve for the variable between two points.

Example: tree diameters

The diameter of a tree trunk is an important variable in forestry. The density curve shown below represents the distribution of diameters (measured at breast height) in a population of 30-year-old Douglas fir trees; areas under the curve are shown, as well. Consider the diameter, in inches, of a randomly chosen tree. Then, for example,
 $P(4 < \text{diameter} < 6) = 0.33$.

- $P(4 \leq \text{diameter} \leq 6)$?
- How about the probability that a randomly chosen tree has a diameter greater than 8 inches?



Random Variables

A **random variable** is simply a variable that takes on numerical values that depend on the outcome of a chance operation.

Example: dice

Consider the chance operation of tossing a die. Let the random variable Y represent the number of spots showing. The possible values of Y are $Y = 1, 2, 3, 4, 5,$ or 6 . We do not know the value of Y until we have tossed the die. If the die is perfectly balanced so that each of the six faces is equally likely, then

$$P(Y = i) = \frac{1}{6},$$

for $i = 1, 2, 3, 4, 5, 6$.

Example: family size

Suppose a family is chosen at random from a certain population, and let the random variable Y denote the number of children in the chosen family. The possible values of Y are $0, 1, 2, 3, \dots$. The probability that Y has a particular value is equal to the percentage of families with that many children. For instance, if 23% of the families have 2 children, then

$$P(Y = 2) = 0.23.$$

Example: heights of men

Let the random variable Y denote the height of a man chosen at random from a certain population. If we know the distribution of heights in the population, then we can specify the probability that Y falls in a certain range. For instance, if 46% of the men are between 65.2 and 70.4 inches tall, then

$$P(65.2 < Y < 70.4) = 0.46.$$

- Each of the variables in dice and family size examples is a **discrete random variable**, because in each case we can list the possible values that the variable can take on.
- In contrast, the variable in heights of men example, height, is a **continuous random variable**: Height, at least in theory, can take on any of an infinite number of values in an interval.
- We use density curves (**probability density functions**) to model the distributions of continuous random variables, such as blood glucose level or tree diameter.
- For discrete random variables, the counterpart of probability density functions (pdf) are **probability mass functions** (pmf), which gives the probability that a discrete random variable is exactly equal to some value. For example, in dice example the pmf is $P(Y = i) = 1/6$ for $i = 1, 2, 3, 4, 5, 6$.

Mean and variation of a random variable

For the case of a discrete random variable, we can calculate the population mean and standard deviation if we know the probability mass function for the random variable.

- The mean (expected value) of a discrete random variable Y is defined as

$$\mu_Y = E(Y) = \sum y_i P(Y = y_i),$$

where the y_i are the values that the variable takes on and the sum is taken over all possible values.

- The variance of a discrete random variable Y is defined as

$$\sigma_Y^2 = \text{Var}(Y) = \sum (y_i - \mu_Y)^2 P(Y = y_i),$$

where the y_i are the values that the variable takes on and the sum is taken over all possible values.

Calculate the mean (3.5), variance(35/12), and stand deviation of the discrete random variable defined in the dice example.

Rules for means of random variables

- Rule (1) If X and Y are two random variables, then

$$\mu_{X+Y} = \mu_X + \mu_Y, \mu_{X-Y} = \mu_X - \mu_Y.$$

- Rule (2) If Y is a random variable and a and b constants, then

$$\mu_{a+bY} = a + b\mu_Y.$$

Rules for Variances of Random Variables

- Rule (3) If Y is a random variable and a and b constants, then

$$\sigma_{a+bY}^2 = b^2 \sigma_Y^2.$$

- Rule (4) If X and Y are two **independent** random variables, then

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2, \quad \sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2.$$

If we add two random variables that are not independent of one another, then the variance of the sum depends on the degree of dependence between the variables. To take an extreme case, suppose that one of the random variables is the negative of the other.

The Binomial Distribution

Independent-Trials Model: A series of n *independent* trials is conducted. Each trial results in success or failure. The probability of success is equal to the same quantity, p , for each trial, regardless of the outcomes of the other trials.

A **binomial random variable** is a random variable that satisfies the following four conditions, abbreviated as BlnS:

- *Binary outcomes:* There are two possible outcomes for each trial (success and failure).
- *Independent trials:* The outcomes of the trials are independent of each other.
- *n is fixed:* The number of trials, n , is fixed in advance.
- *Same value of p :* The probability of a success on a single trial is the same for all trials.

Example: coin tossing

Suppose we have a fair coin and we flip it 10 times. We are interested in the number of times we get heads (success) out of the 10 flips.

The binomial distribution comes into play because each flip of the coin is *independent*, and there are only *two possible outcomes* (heads or tails) with a *fixed probability of success* (0.5 for a fair coin).

In general, if the random variable X follows the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$, we write $X \sim B(n, p)$. The probability of getting exactly j successes in n independent Bernoulli trials is given by the probability mass function:

$$P(X = j) = \binom{n}{j} p^j (1 - p)^{n-j}$$

where:

- n is the number of trials (10 flips)
- j is the number of successes (3 heads)
- p is the probability of success (0.5 for a fair coin)
- the binomial coefficient $\binom{n}{j} = \frac{n!}{j!(n-j)!}$ and $n! = n(n-1) \cdots 1$

Consider the coin tossing example, let X be the number of heads out of the 10 flips. It follows that $X \sim B(10, 0.5)$. Using the above formula, the probability of getting exactly 3 heads out of the 10 flips is

$$P(X = 3) = \binom{10}{3} (0.5)^3 (1 - 0.5)^{10-3} = 0.117.$$

Properties of binomial coefficients

- $0! = 1$

- $$\binom{n}{0} = \binom{n}{n} = 1$$

- Symmetry:

$$\binom{n}{j} = \binom{n}{n-j}$$

Using the complement to calculate probability

$$P(E) = 1 - P(E^C)$$

Sometimes calculating probability of the complement of an event can be easier than calculating the probability of the event itself. We can use the probability of the complement to find the probability of the event by subtracting it from one. This trick is often used when calculating the probability of multiple events.

What is the probability of getting at most 9 heads out of 10 flips?

Mean and variance of a binomial

For a binomial random variable $X \sim B(n, p)$, we have

- the mean is

$$\mu_X = E(X) = np;$$

- the variance is

$$\sigma_X^2 = \text{Var}(X) = np(1 - p).$$

The Bernoulli distribution

- The discrete probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability $1 - p$.

- It can be thought of as a model for the set of possible outcomes of any single experiment that asks a yes-no question. Such questions lead to outcomes that are boolean-valued: a single bit whose value is success/yes/true/one with probability p and failure/no/false/zero with probability q .
- It can be used to represent a (possibly biased) coin toss where 1 and 0 would represent "heads" and "tails", respectively, and p would be the probability of the coin landing on heads (or vice versa where 1 would represent tails and p would be the probability of tails).
- The Bernoulli distribution is a special case of the binomial distribution where a single trial is conducted ($n = 1$).

Deriving the mean and variance of a binomial using Bernoulli distribution

The binomial random variable $X \sim B(n, p)$ is the sum of n identical Bernoulli random variables, each with expected value p and variance $p(1 - p)$. In other words, if X_1, \dots, X_n are identical (and independent) Bernoulli random variables with parameter p , then $X = X_1 + \dots + X_n$.

Example: blood types

In the United States, 85% of the population has Rh positive blood. Suppose we take a random sample of 6 persons and count the number with Rh positive blood. The binomial model can be applied here, since the BInS conditions are met: There is a binary outcome on each trial (Rh positive or Rh negative blood), the trials are independent (due to the random sampling), n is fixed at 6, and the same probability of Rh positive blood applies to each person ($p = 0.85$).

Let Y denote the number of persons, out of 6, with Rh positive blood. Calculate

- $P(Y = 4)$
- $P(Y \geq 2)$
- $E(Y)$ and $\text{Var}(Y)$