Probability

Marc H. Mehlman
marcmehlman@yahoo.com

University of New Haven

"The theory of probabilities is at bottom nothing but common sense reduced to calculus. – Laplace, Théorie analytique des probabilités, 1820

"Baseball is 90 percent mental. The other half is physical." – Yogi Berra

Table of Contents

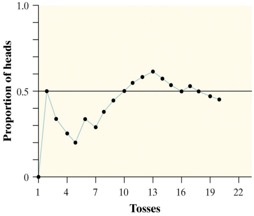
Probability Models

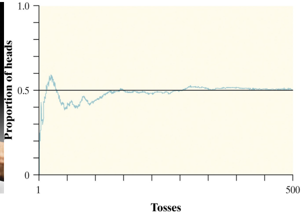
Probability Models

The Language of Probability

Chance behavior is unpredictable in the short run, but has a regular and predictable pattern in the long run.

> We call a phenomenon **random** if individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions.
>
> The **probability** of any outcome of a chance process is the proportion of times the outcome would occur in a very long series of repetitions.





4

Probability Models

Descriptions of chance behavior contain two parts: a list of possible outcomes and a probability for each outcome.

The **sample space *S*** of a chance process is the set of all possible outcomes.

An **event** is an outcome or a set of outcomes of a random phenomenon. That is, an event is a subset of the sample space.

A **probability model** is a description of some chance process that consists of two parts: a sample space S and a probability for each outcome.

7

Definition (Empirical Probability)

A series of trails are **independent** if and only if the outcome of one trail does not effect the outcome of any other trail. Consider the proportion of times an event occurs in a series of independent trails. That proportion approaches the **empirical probability** of an event occurring as the number of independent trails increase.

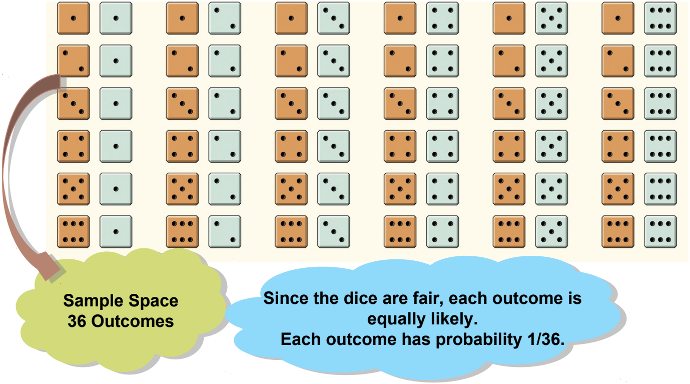
Definition (Equally Likely Outcome Probability)

Given a probability model, if there are only a finite number of outcomes and each outcome is equally likely, the probability of any event A is

$$P(A) \stackrel{\text{def}}{=} \frac{\#\text{ outcomes in } A}{\#\text{ possible outcomes in } S}.$$

Probability Models

Example: Give a probability model for the chance process of rolling two fair, six-sided dice—one that's red and one that's green.



8

Probability Rules

1. Any probability is a number between 0 and 1.
2. All possible outcomes together must have probability 1.
3. If two events have no outcomes in common, the probability that one or the other occurs is the sum of their individual probabilities.
4. The probability that an event does not occur is 1 minus the probability that the event does occur.

Rule 1. The probability $P(A)$ of any event A satisfies $\mathbf{0 \leq P(A) \leq 1}$.

Rule 2. If S is the sample space in a probability model, then $\mathbf{P(S) = 1}$.

Rule 3. If A and B are **disjoint**, $\mathbf{P(A \text{ or } B) = P(A) + P(B)}$.

This is the **addition rule for disjoint events.**

Rule 4: The **complement** of any event A is the event that A does not occur, written A^c. $\mathbf{P(A^c) = 1 - P(A)}$.

9

Probability Rules

Distance-learning courses are rapidly gaining popularity among college students. Randomly select an undergraduate student who is taking distance-learning courses for credit and record the student's age. Here is the probability model:

Age group (yr):	18 to 23	24 to 29	30 to 39	40 or over
Probability:	0.57	0.17	0.14	0.12

(a) Show that this is a legitimate probability model.

Each probability is between 0 and 1 and
0.57 + 0.17 + 0.14 + 0.12 = 1

(b) Find the probability that the chosen student is not in the traditional college age group (18 to 23 years).

***P*(not 18 to 23 years) = 1 – *P*(18 to 23 years)**
= 1 – 0.57 = 0.43

10

Finite Probability Models

One way to assign probabilities to events is to assign a probability to every individual outcome, then add these probabilities to find the probability of any event. This idea works well when there are only a finite (fixed and limited) number of outcomes.

A probability model with a finite sample space is called **finite.**

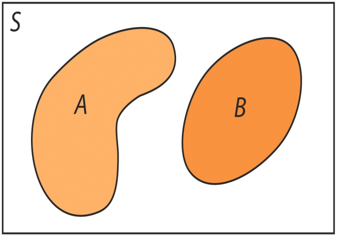
To assign probabilities in a finite model, list the probabilities of all the individual outcomes. These probabilities must be numbers between 0 and 1 that add to exactly 1. The probability of any event is the sum of the probabilities of the outcomes making up the event.

11

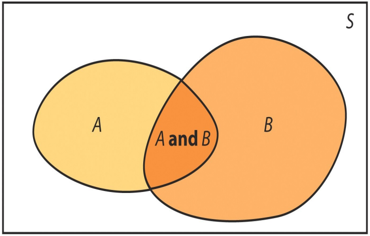
Venn Diagrams

Sometimes it is helpful to draw a picture to display relations among several events. A picture that shows the sample space *S* as a rectangular area and events as areas within *S* is called a **Venn diagram.**

Two disjoint events:



Two events that are not disjoint, and the event {*A* and *B*} consisting of the outcomes they have in common:

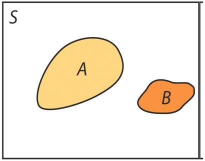


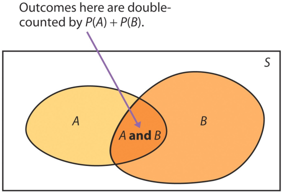
12

The General Addition Rule

Addition Rule for Disjoint Events
If *A*, *B*, and *C* are **disjoint** in the sense that no two have any in common, then:

$$P(A \text{ or } B) = P(A) + P(B)$$





Addition Rule for Unions of Two Events
For any two events *A* and *B*:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

34

Multiplication Rule for Independent Events

If two events A and B do not influence each other, and if knowledge about one does not change the probability of the other, the events are said to be **independent** of each other.

Multiplication Rule for Independent Events

Two events A and B are **independent** if knowing that one occurs does not change the probability that the other occurs. If A and B are independent:

$$P(A \text{ and } B) = P(A) \times P(B)$$

13

"...when you have eliminated the impossible, whatever remains, however improbably, must be the truth." – Sherlock Holmes in the Sign of Four

Conditional Probability

The probability we assign to an event can change if we know that some other event has occurred. This idea is the key to many applications of probability.

When we are trying to find the probability that one event will happen under the *condition* that some other event is already known to have occurred, we are trying to determine a **conditional probability.**

The probability that one event happens given that another event is already known to have happened is called a **conditional probability.**

When $P(A) > 0$, the probability that event B happens *given* that event A has happened is found by:

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

36

The General Multiplication Rule

The definition of conditional probability reminds us that in principle all probabilities, including conditional probabilities, can be found from the assignment of probabilities to events that describe a random phenomenon. The definition of conditional probability then turns into a rule for finding the probability that both of two events occur.

The probability that events A and B both occur can be found using the **general multiplication rule:**

$$P(A \text{ and } B) = P(A) \bullet P(B \mid A)$$

where $P(B \mid A)$ is the conditional probability that event B occurs given that event A has already occurred.

Note: Two events A and B that both have positive probability are **independent** if:

$$P(B|A) = P(B)$$

37

Random Variables

Random Variables

Random Variables

A **probability model** describes the possible outcomes of a chance process and the likelihood that those outcomes will occur.

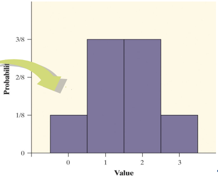
A numerical variable that describes the outcomes of a chance process is called a **random variable.** The probability model for a random variable is its probability distribution.

> A **random variable** takes numerical values that describe the outcomes of some chance process.
> The **probability distribution** of a random variable gives its possible values and their probabilities.

Example: Consider tossing a fair coin 3 times.
Define X = the number of heads obtained

X = 0: TTT
X = 1: HTT THT TTH
X = 2: HHT HTH THH
X = 3: HHH

Value	**0**	**1**	**2**	**3**
Probability	1/8	3/8	3/8	1/8



15

Marc Mehlman

Discrete Random Variable

There are two main types of random variables: *discrete* and *continuous*. If we can find a way to list all possible outcomes for a random variable and assign probabilities to each one, we have a **discrete random variable.**

A **discrete random variable *X*** takes a fixed set of possible values with gaps between. The probability distribution of a discrete random variable X lists the values x_i and their probabilities p_i:

Value:	x_1	x_2	x_3	…
Probability:	p_1	p_2	p_3	…

The probabilities p_i must satisfy two requirements:

- Every probability p_i is a number between 0 and 1.
- The sum of the probabilities is 1.

To find the probability of any event, add the probabilities p_i of the particular values x_i that make up the event.

16

Continuous Random Variable

Discrete random variables commonly arise from situations that involve counting something. Situations that involve measuring something often result in a **continuous random variable.**

> A **continuous random variable** Y takes on all values in an interval of numbers. The probability distribution of Y is described by a **density curve.** The probability of any event is the area under the density curve and above the values of Y that make up the event.

The probability model of a discrete random variable X assigns a probability between 0 and 1 to each possible value of X.

A continuous random variable Y has *infinitely many* possible values. All continuous probability models assign probability 0 to every individual outcome. Only *intervals* of values have positive probability.

17

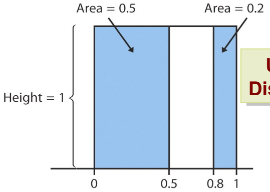
Continuous Probability Models

Suppose we want to choose a number at random between 0 and 1, allowing *any* number between 0 and 1 as the outcome.
We cannot assign probabilities to each individual value because there is an infinite interval of possible values.

A **continuous probability model** assigns probabilities as areas under a density curve. The area under the curve and above any range of values is the probability of an outcome in that range.

Example: Find the probability of getting a random number that is less than or equal to 0.5 OR greater than 0.8.

$$P(X \leq 0.5 \text{ or } X > 0.8)$$
$$= P(X \leq 0.5) + P(X > 0.8)$$
$$= 0.5 + 0.2$$
$$= 0.7$$



Uniform Distribution

18

Definition

The **expectation** (average value) of a random variable, X, is denoted as $\mathrm{E}(X)$.

Definition (Mean and Variance of a Discrete Random Variable)

The **mean** of a discrete random variable, X, is

$$\mu_X \stackrel{\text{def}}{=} \mathrm{E}(X) = \sum_j x_j p_j = x_1 p_1 + x_2 p_2 + \cdots$$

The **variance** of a discrete random variable, X, is

$$\sigma_X^2 \stackrel{\text{def}}{=} \mathrm{E}((X-\mu_X)^2) = \sum_j (x_j - \mu_X)^2 p_j = (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + \cdots$$

The **standard deviation** of a random variable is $\sigma_X \stackrel{\text{def}}{=} \sqrt{\sigma_X^2}$.

Definition

Two random variables are **independent** if knowing the values of one random variable gives no clue to what the values of the other random variable are.

Let X and Y be random variables and a and b be fixed numbers.

Rules for Means:

Rule 1: $\mu_{aX+b} = a\mu_X + b$.

Rule 2: $\mu_{X+Y} = \mu_X + \mu_Y$.

Rules for Variances:

Rule 1: $\sigma^2_{aX+b} = a^2\sigma^2_X$.

Rule 2:

$$\begin{aligned} \sigma^2_{X+Y} &= \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y \\ \sigma^2_{X-Y} &= \sigma^2_X + \sigma^2_Y - 2\rho\sigma_X\sigma_Y. \end{aligned}$$

Rule 3: If X and Y are independent, $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$.

Example

Let X be the number of heads from tossing a coin four times. The coin has a probability of being a head of 0.25 for each toss.

Values	0	1	2	3	4
Probabilities	0.3164	0.4219	0.2109	0.0469	0.0039

Find the mean and variance of X and the mean and variance of $3X + 4$.
Solution:

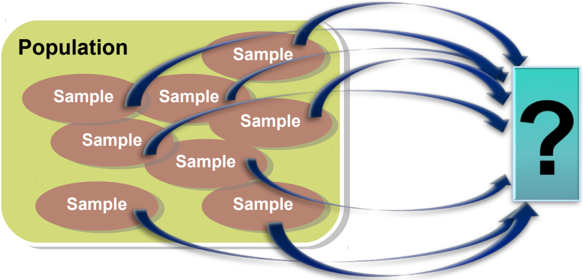
$$\begin{aligned}
\mu_X &= 0 * 0.3164 + 1 * 0.4219 + 2 * 0.2109 + 3 * 0.0469 + 4 * 0.0039 = 1 \\
\sigma_X^2 &= (0-1)^2 * 0.3164 + (1-1)^2 * 0.4219 + (2-1)^2 * 0.2109 \\
&\qquad + (3-1)^2 * 0.0469 + (4-1)^2 * 0.0039 \\
&= 0.75 \\
\mu_{3X+4} &= 3 * 1 + 4 = 7 \\
\sigma_{3X+4}^2 &= 3^2 * 0.75 = 6.75.
\end{aligned}$$

Statistical Estimation

Suppose we would like to estimate an unknown μ. We could select an SRS and base our estimate on the sample mean. However, a different SRS would probably yield a different sample mean.

This basic fact is called **sampling variability:** the value of a statistic varies in repeated random sampling.

To make sense of sampling variability, we ask, "What would happen if we took many samples?"



25

$\bar{X}$ is a random variable that associates with each random sample, the random sample's mean, $\bar{x}$. One might ask "Does $\bar{X}$ tend to be close to being μ?" The answer become more and more "yes" as the sample size increases:

Theorem (Law of Large Numbers)

If we keep on taking larger and larger samples, the statistic $\bar{x}$ becomes closer and closer to the parameter μ.