Chapter 4 Probability distributions

Probabilities

Important terms

- **Parameter:** A quantity describes the population, usually unknown.
- **Statistic:** Any quantity describes the sample, is computed from a sample, is used to estimate population parameters.

- We can make conclusions about an unknown population parameter using a representative sample
  - Example: In clinical trials, a new drug is tested and based on the sample evidence, it is concluded the entire target population will benefit.

Statistical model

- DATA = MODEL + ERROR

- DATA are observed, MODEL captures the structural features of interest, ERROR defines the randomness associated with the process.

- Example the population mean model: $x_i = \mu + \varepsilon_i$, where x_i is the observed value, μ is the (unknown) true population mean, and ε_i represents the (unknown) random deviation of x_i from μ .

4.2-4.7 Probabilities objectives

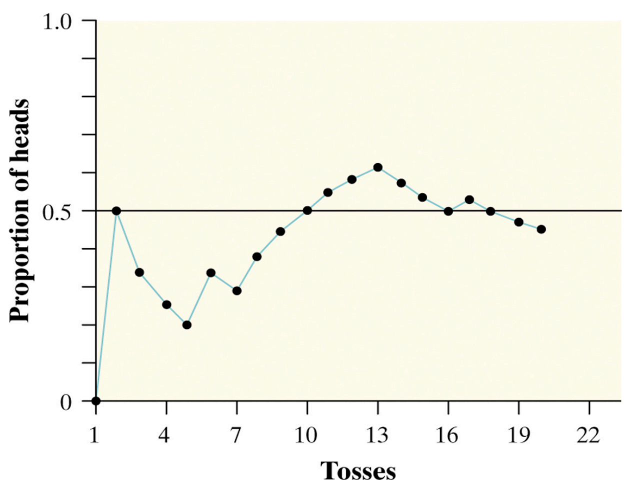
- The idea of probability
- Probability models
- Probability rules
- Finite probability models
- Continuous probability models
- Random variables
- Mean and variance of random variables

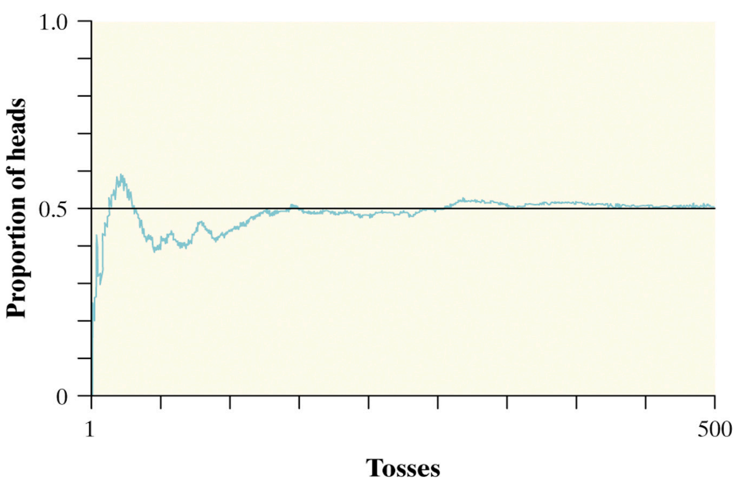
The idea of probability

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

RANDOMNESS AND PROBABILITY

- 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 random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions.





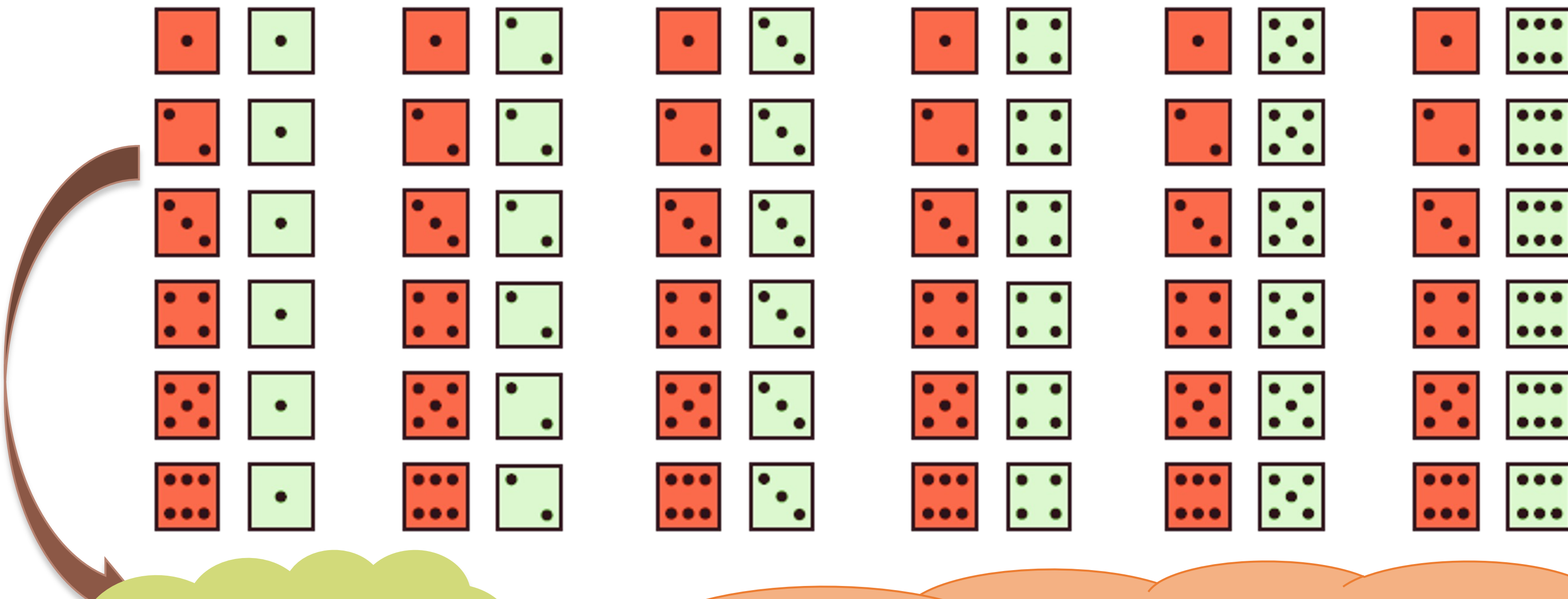
Probability models

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

- The sample space S of a random phenomenon 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 mathematical description of a random phenomenon consisting of two parts: a sample space S and a way of assigning probabilities to events.

Probability models

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



**Sample space
36 outcomes**

**Since the dice are fair, each outcome is equally likely.
Each outcome has probability 1/36.**

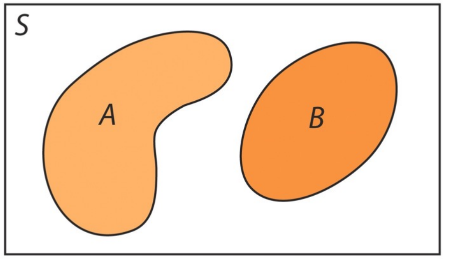
Set theory

- S = Sample space. Contains all possible outcomes.
- ∅= Empty set. The null or impossible event.

Set Operators:

- ∪= union. All the outcomes from combining of two events
$$P(A \cup B) = P(A\ or\ B)$$
- ∩= intersection. The outcomes common to two events.
$$P(A \cap B) = P(A\ and\ B)$$
- $\bar{E}$ = complement (of event E). The event of not getting E
$$P(\bar{E}) = 1 - P(E)$$
- Disjoint Events: A and B are called disjoint if A ∩ B = ∅ ; (no outcomes are common to A and B)

Venn Diagram



Probability axioms

The probability axioms in formal language:

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

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

Rule 3. Two events A and B are disjoint (mutually exclusive)if they have no outcomes in common and, thus, can never occur together. If A and B are disjoint,

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

Or $P(A_1 \cup A_2 \cup \cdots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k)$

This is the addition rule for disjoint events.

Rule 4. For any event A, $P(\bar{A}) = P(A \text{ does not occur}) = 1 - P(A)$

Probability axioms (example)

Example: Favorite vehicle colors. Our preferences for vehicle colors can be related to our personality, our moods, or particular objects. Here is a probability model for color preferences:

Color	**White**	**Black**	**Silver**	**Gray**
Probability	0.24	0.19	0.16	0.15
Color	Red	Blue	Brown	Other
Probability	0.10	0.07	0.05	0.04

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

Each probability is between 0 and 1 and
0.24 + 0.19 + 0.16 + 0.15 + 0.10 + 0.07 + 0.05 + 0.04 = 1

(b)Find the probability that a person's favorite vehicle color is black or silver.

P(Black or Silver) = P(Black) + P(Silver)
= 0.19 + 0.16 = 0.35

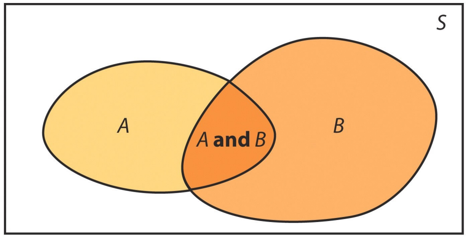
(c) Find the probability that the favorite color is not blue.

P(Not Blue) = 1 – P(Blue) = 1 – 0.07 = 0.93

More probability rules

- General addition rule:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



- Conditional Probability The chance of event A occurring given that event B occurs, denoted by P(A|B), is called the conditional probability of A given B. Provided P(B) > 0

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Independence Events A and B are said to be statistically independent if and only if:

$$P(A \cap B) = P(A)P(B)$$

Example: Addition Rule

- Consider a study on hypertension. Suppose that 10% of patients take drug A, 20% take drug B, and 5% take both A and B.

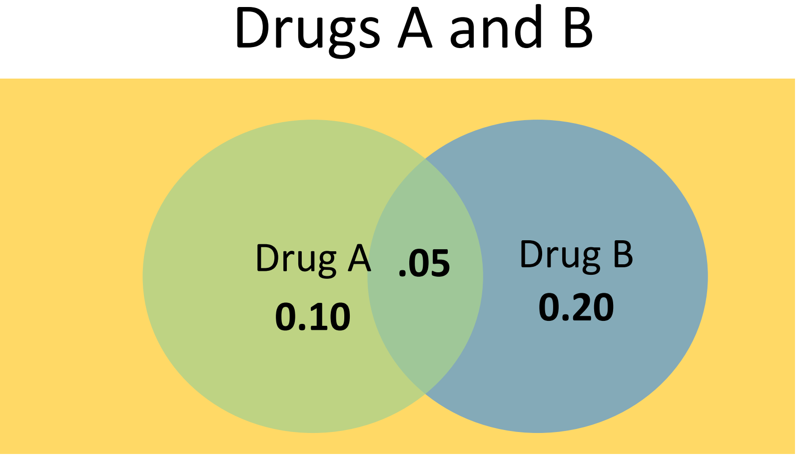
  $P(A) = 0.10$, $P(B) = 0.20$, $P(A \text{ and } B) = 0.05$

  - Are taking drugs A and B mutually exclusive?

    $P(A \cap B) = P(A \text{ and } B) = 0.05 \neq 0$, so ***no***

  - What is the probability of taking drug A or B?

    $$P(A \cup B) = P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$
    $$= 0.10 + 0.20 - 0.05$$
    $$= \mathbf{0.25}$$



Example: Independence and Conditional Probability

- In a study of smoking (S) and coffee drinking (C), 5% of the people smoked but did not drink coffee, 45% drank coffee but did not smoke, 35% drank coffee and smoked, and 15% neither drank coffee nor smoked.
- Is smoking independent of drinking coffee?
  - Does P(S) = P(S|C)?
  - **No:** 0.4 ≠ 0.35/0.8 = 0.4375
  - P(S and NC) = 0.05
    P(NS and C) = 0.45
    P(S and C) = 0.35
    P(NS and NC) = 0.15

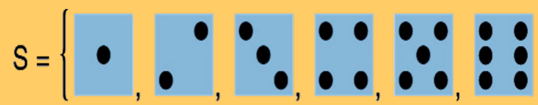
	Coffee drinker	Not a coffee drinker	Total
Smoker	0.35	0.05	0.40
Not a smoker	0.45	0.15	0.60
Total	0.80	0.20	1.00

- **Unconditional probability of smoking**
  - P(S) = P(S and C) + P(S and NC) = 0.35 + 0.05 = **0.40**
- **Conditional probability of smoking**
  - P(S|C) = P(S and C)/P(C)
  - P(C) = P(S and C) + P(NS and C) = 0.35 + 0.45 = 0.80
  - P(S|C) = P(S and C)/P(C) = 0.35/0.80 = **0.4375**
- Since P(S) ≠ P(S|C), smoking and coffee drinking are **not** independent

Probabilities: finite number of outcomes

Finite sample spaces deal with **discrete data**—data that can only take on a limited number of values. These values are often integers or whole numbers.

Throwing a die:



$S = \{1, 2, 3, 4, 5, 6\}$

The individual outcomes of a random phenomenon are always disjoint. ➔ The probability of any event is the sum of the probabilities of the outcomes making up the event (addition rule).

M&M candies

If you draw an M&M candy at random from a bag, the candy will have one of six colors. The probability of drawing each color depends on the proportions manufactured, as described here:

Color	Brown	Red	Yellow	Green	Orange	Blue
Probability	0.3	0.2	0.2	0.1	0.1	?

What is the probability that an M&M chosen at random is blue?

S = {brown, red, yellow, green, orange, blue}

$P(S) = P(\text{brown}) + P(\text{red}) + P(\text{yellow}) + P(\text{green}) + P(\text{orange}) + P(\text{blue}) = 1$

$P(\text{blue}) = 1 - [P(\text{brown}) + P(\text{red}) + P(\text{yellow}) + P(\text{green}) + P(\text{orange})]$

$= 1 - [0.3 + 0.2 + 0.2 + 0.1 + 0.1] = 0.1$

What is the probability that a random M&M is either red, yellow, or orange?

$P(\text{red or yellow or orange}) = P(\text{red}) + P(\text{yellow}) + P(\text{orange})$

$= 0.2 + 0.2 + 0.1 = 0.5$

Probabilities

We can assign probabilities either:

- **empirically** → from our knowledge of numerous similar past events
    - Mendel discovered the probabilities of inheritance of a given trait from experiments on peas without knowing about genes or DNA.

- **or theoretically** → from our understanding of the phenomenon and symmetries in the problem
    - A 6-sided fair die: each side has the same chance of turning up

If a random phenomenon has *k* equally likely possible outcomes, then each individual outcome has probability 1/*k*.

And, for any event A:

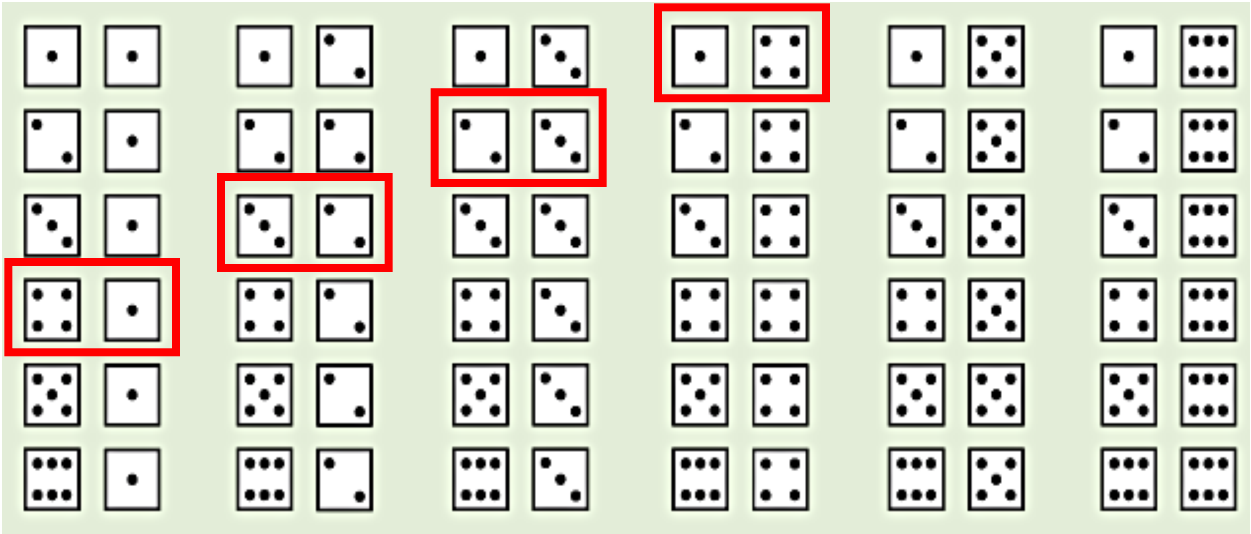
$$P(\mathrm{A}) = \frac{\text{count of outcomes in A}}{\text{count of outcomes in S}}$$

Dice

You toss two dice. What is the probability of the outcomes summing to 5?

This is S:

{(1,1), (1,2), (1,3), ……etc.}



There are 36 possible outcomes in *S*, all equally likely (given fair dice).

Thus, the probability of any one of them is 1/36.

P(the roll of two dice sums to 5) =

P(1,4) + *P*(2,3) + *P*(3,2) + *P*(4,1) = 4 / 36 = 0.111

Finite probability models

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

- 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.

- Finite probability models are sometimes called discrete probability models. Statisticians often refer to finite probability models as discrete.

Continuous probability models

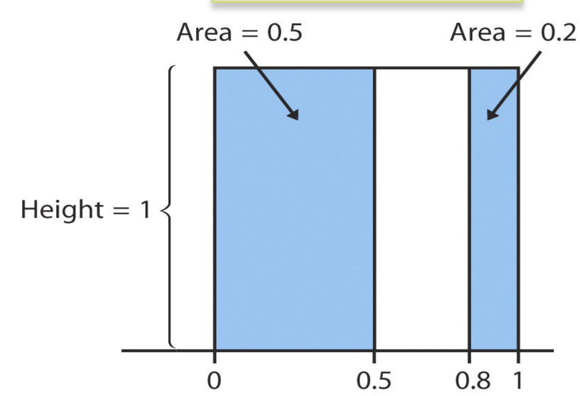
- Suppose we want to choose a number at random between 0 and 1, allowing any number between 0 and 1 as the outcome.
- We cannot assign probabilities to each individual value because there is an infinite continuum of possible values.

- A continuous probability model assigns probability as an area 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



Combinatorics

- Combinatoric: the number of ways of selecting x objects from n objects without considering the order in which they are selected:

$\binom{n}{x} = \frac{n!}{x!(n-x)!}$, $\binom{n}{x}$ is read as "n choose x", k!=1x2x3…(k-1)k, is read as k factorial.

- Example: how many ways to choose 2 individuals from 10 individuals.

$$\binom{10}{2} = \frac{10!}{2!(10-2)!} = \frac{10!}{2!8!} = \frac{10\times 9}{2\times 1} = 45$$

Random Variables

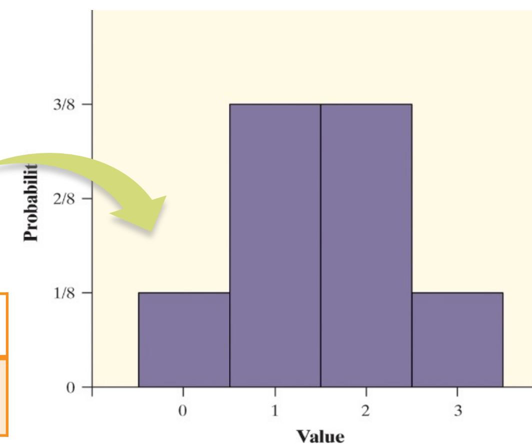
A **random variable** takes numerical values that describe the outcomes of some chance process, often denoted by X, Y, Z etc. eg, X~N(μ, σ)

The **probability distribution** of a random variable gives its possible values and their probabilities.

Example: Consider tossing a fair coin three 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



Discrete Random Variables

- Random variables that have a finite (countable) list of possible outcomes, with probabilities assigned to *each* of these outcomes, are called **discrete**

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:

1. Every probability p_i is a number between 0 and 1.

2. 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.

Discrete random variables

- Example: denote the output of tossing a coin as X, then $X = 1$(Head) or $X = 0$(Tail), the probability distribution function is: $P(X = x) = 0.5, x = 0, 1.$
- Fair dice throw: $P(X = x) = \frac{1}{6}, x = 1,2, \dots, 6$
- $P(X = x)$ can be denoted as $P_X(x)$ or $P(x)$

Properties of $P_X(x)$:

- 1. $P_X(x) \geq 0$ for any value x
- 2. $\Sigma_x P_X(x) = 1$

Discrete Random Variable

Example

A liberal arts college posts the grade distribution for its courses. In a recent semester, students in one section of English 130 received 32% A's, 42% B's, 19% C's, 3% D's, and 4% F's.

Here is the distribution of the discrete random variable called "Grade Points," where A = 4, B = 3, etc.:

Value:	0	1	2	3	4
Probability:	0.04	0.03	0.19	0.42	0.32

What is the probability that a randomly selected student got a B or better?

$P(X \geq 3) = P(X = 3) + P(X = 4)$

$= 0.42 + 0.32 = 0.74.$

Continuous Random Variables

- Random variables that can take on any value in an interval, with probabilities given as areas under a density curve, are called **continuous**

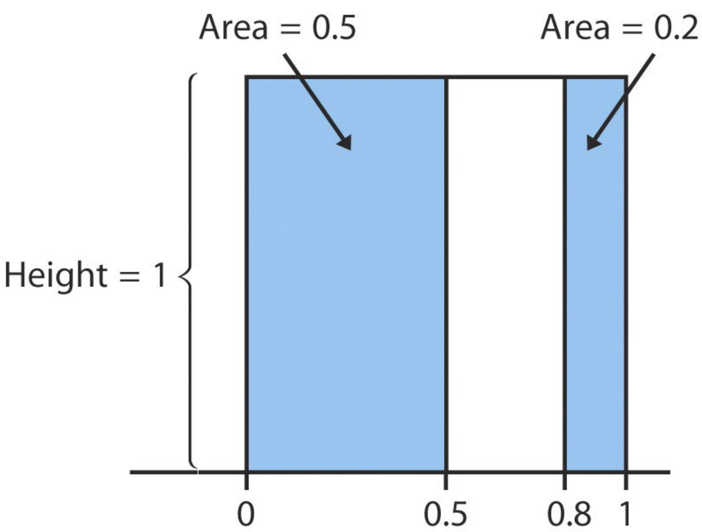
- A continuous random variable X has ***infinitely many* possible values**. All continuous probability models assign probability 0 to every individual outcome. Only ***intervals*** of values have positive probability.

Assigning Probabilities for continuous random variables

Random number generators give output (digits) spread uniformly across the interval from 0 to 1.

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

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



Continuous random variable

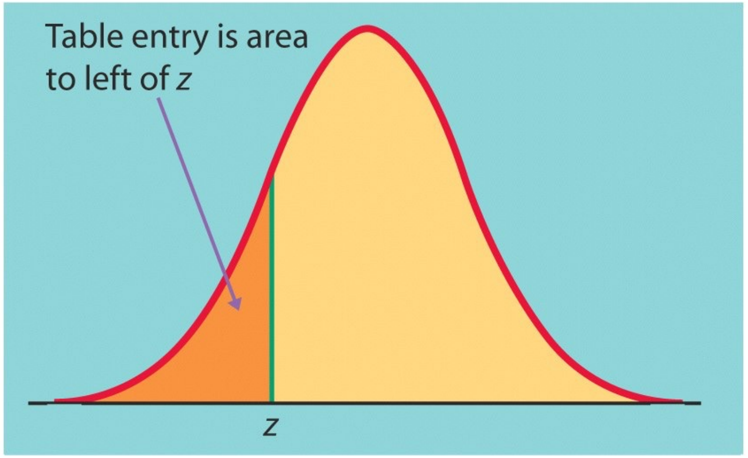
- For continuous random variables, probability measures are only associated with intervals on the real line.
- $P(X = x) = 0, \text{for any } x$.
- Probability Density Function $f_X(x)$ defines the density curve for the random variable X. Areas under this curve correspond to probabilities of the interval events of interest.

Properties of $f_X(x)$:

- 1. $f_X(x) \geq 0$ for all x
- 2. The total area under the curve $f_X(x)$ is always 1
- 3. $\mathrm{P(a < X < b)}$ is the area under the curve $f_X(x)$ between $x = a$ and $x = b$.

Normal Probability Models

- Often the density curve used to assign probabilities to intervals of outcomes is the Normal curve
  - Normal distributions are probability models: probabilities can be assigned to intervals of outcomes using the Standard Normal probabilities in Table A of the text.



Cumulative distribution function (cdf)

- The cdf of a random variable X is defined by $F_X(x) = P(X \leq x)$.
- $P(X > x) = 1 - P(X \leq x) = 1 - F_X(x)$.
- For continuous random variable,

$$P(X \leq b) = P(X < b) \text{ since } P(X = b) = 0$$
$$P(a < X < b) = P(a \leq X \leq b) = F_X(b) - F_X(a)$$

- For discrete random variable,

$$P(X \leq b) > P(X < b) \text{ since } P(X = b) \neq 0$$
$$P(a < X < b) \neq P(a \leq X \leq b)$$

The cdf plays an important role in statistics.

1. It provides critical values and p-values for assessing significance.
2. Percentiles are obtained by solving for x_p such that
$$F(x_p) = P(X \leq x_p) = p \text{ with } 0 < p < 1$$
3. The median M is of interest in dose-response studies:
$$F(M) = P(X \leq M) = 0.5$$

Mean of a random variable (Expectation)

The mean *x* bar of a set of observations is their arithmetic average.

The mean ***μ*** of a random variable *X* is a weighted average of the possible values of *X*, reflecting the fact that all outcomes might not be equally likely.

For a discrete random variable *X* with probability distribution →

Value of X	x_1	x_2	x_3	$\cdots$	x_k
Probability	p_1	p_2	p_3	$\cdots$	p_k

the mean μ of *X* is found by multiplying each possible value of *X* by its probability, and then adding the products.

$$E(X) = \mu_X = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k$$
$$= \sum x_i p_i$$

The mean of a random variable *X* is also called **expected value** of *X*, *E*(*X*).

Note: The expected value does not need to be a possible value of *X* or an integer!

It is a long-term average over many repetitions.

Mean of a discrete random variable

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

Value of X	0	1	2	3
Probability	1/8	3/8	3/8	1/8

The mean μ of X is

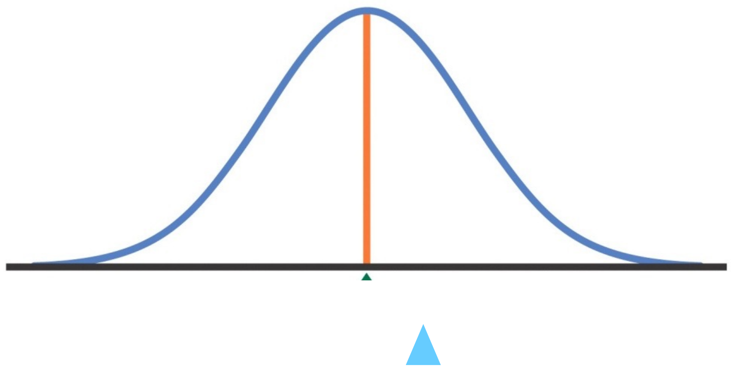
$E(X)=\mu_x = \ldots\ldots.$

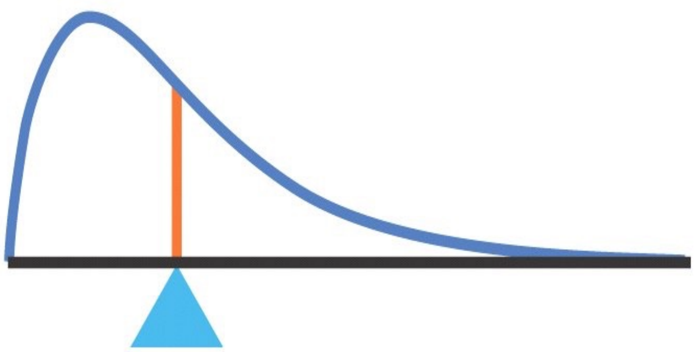
$= 1.5$

Mean of a continuous random variable

The probability distribution of continuous random variables is described by a density curve.

The mean lies at the center of symmetric density curves such as the normal curves.





Exact calculations for the mean of a distribution with a skewed density curve are more complex.

$\mu_X = E(X) = \int x f_X(x) dx,$
where X is a continuous random variable with probability density function $f_X(x)$.

Variance of a random variable

The variance and the standard deviation are the measures of spread that accompany the choice of the mean to measure center.

The **variance σ^2_X** of a random variable is a weighted average of the squared deviations $(X - \mu_X)^2$ of the variable X from its mean μ_X. Each outcome is weighted by its probability in order to take into account outcomes that are not equally likely.

The larger the variance of X, the more scattered the values of X on average. The positive square root of the variance gives the **standard deviation σ** of X.

Variance of a discrete random variable

For a discrete random variable X

with probability distribution →

Value of X	x_1	x_2	x_3	…	x_k
Probability	p_1	p_2	p_3	…	p_k

and mean μ_X, the variance σ^2 of X is found by multiplying each squared deviation of X by its probability and then adding all the products.

$$\sigma_X^2 = E(X-\mu_X)^2 = (x_1-\mu_X)^2 p_1 + (x_2-\mu_X)^2 p_2 + \ldots + (x_k-\mu_X)^2 p_k$$
$$= \sum (x_i-\mu_X)^2 p_i$$

$\sigma_X^2 = \mathrm{E}(\mathrm{X}-\mu_X)^2 = \int (\mathrm{x}-\mu_X)^2 f_X(x)dx$ for continuous random variable

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

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

. $\mu_X = 1.5$

The variance σ^2 of X is

σ^2 = …………

= 0.75

Class exercise

A random variable X has the following distribution:
Find the mean and standard deviation for this random variable.

X	-2	-1	0	1
Probability	0.1	0.2	0.4	0.3

Discrete random variable distributions

- Bernoulli: Let X be the RV associated with a process having only two outcomes, success (1) and failure (0), with P(X = 1)=p, P(X=0)=1-p. Then X is called a Bernoulli random variable and has pdf.

$$P(x) = p^x(1-p)^{1-x}, x = 0, 1. \quad \mu = p, \sigma^2 = p(1-p)$$

- Binomial: If a Bernoulli process is repeated n times, then the random variable associated with the number of successes in n trials is called a Binomial random variable, denoted by X~ Bin(n, p) and has pdf:

$$P(x) = \binom{n}{x} p^x(1-p)^{n-x,} \; x = 0,1, \ldots, \mathrm{n}. \; \mu = np, \sigma^2 = np(1-p)$$

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

Discrete random variable distributions

- Geometric: Let X be the number of Bernoulli trials required to obtain the first success. Then X has a geometric distribution and has pdf.

$$P(x) = p(1-p)^{x-1}, x = 1,2 \dots \quad \mu = \frac{1}{p}, \sigma^2 = \frac{(1-p)}{p^2}$$

- Hypergeometric: Suppose n items are to be drawn from a hat containing N objects, Np of which are labeled success, N(1-p) are labeled fail. When the selection is done without replacement, then the chance of getting a success after each draw will change. The number of successes X from n draws follows a hypergeometric distribution and has pdf.

$$P(X = k) = \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{n}{k}}, \text{ k} = 0,1, \dots, \text{n.} \ \mu = np, \sigma^2 = np(1-p)\frac{N-n}{N-1}$$

example: If 6 people are randomly selected from a group consisting of 12 men and 8 women, then the number of women chosen is a hypergeometric random variable with parameters (N=?, n=?, p=?)

- Poisson: expresses the probability of a given number of events occurring in a fixed interval of time if these events occur with a known constant mean rate λ and independently of the time since the last event. X represents the number of times the event occurs over a unit time period and has pdf:

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0,1,2 \dots, \qquad \mu = \lambda = \sigma^2$$

Suppose the average number of accidents occurring weekly on a particular highway is equal to 1.2. Approximate the probability that there is at least one accident this week.

Continuous random variable distributions

- Uniform: X~U(a; b) The continuous equivalent of a discrete process whose outcomes occur with equal probability is a Uniform process over the interval (a, b). The pdf is:

$$f(x) = \frac{1}{b-a}, a < x < b,$$

- Normal: $X \sim N(\mu, \sigma^2)$. This is also called the Gaussian distribution, Bell curve, symmetric around the center. The pdf is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp \frac{-(x-\mu)^2}{2\sigma^2}, -\infty < x < \infty$$

- Exponential: Failure time processes are often characterized by memory-loss process. That is, knowing that no failure has occurred up to time t does not change the nature of the process. It has no memory of what has happed so far. The pdf is parametrized by the mean-waiting time (to failure) of the process:

$$f(x) = \lambda e^{-\lambda x}, x > 0, \mu = \frac{1}{\lambda}, \sigma^2 = \frac{1}{\lambda^2}$$