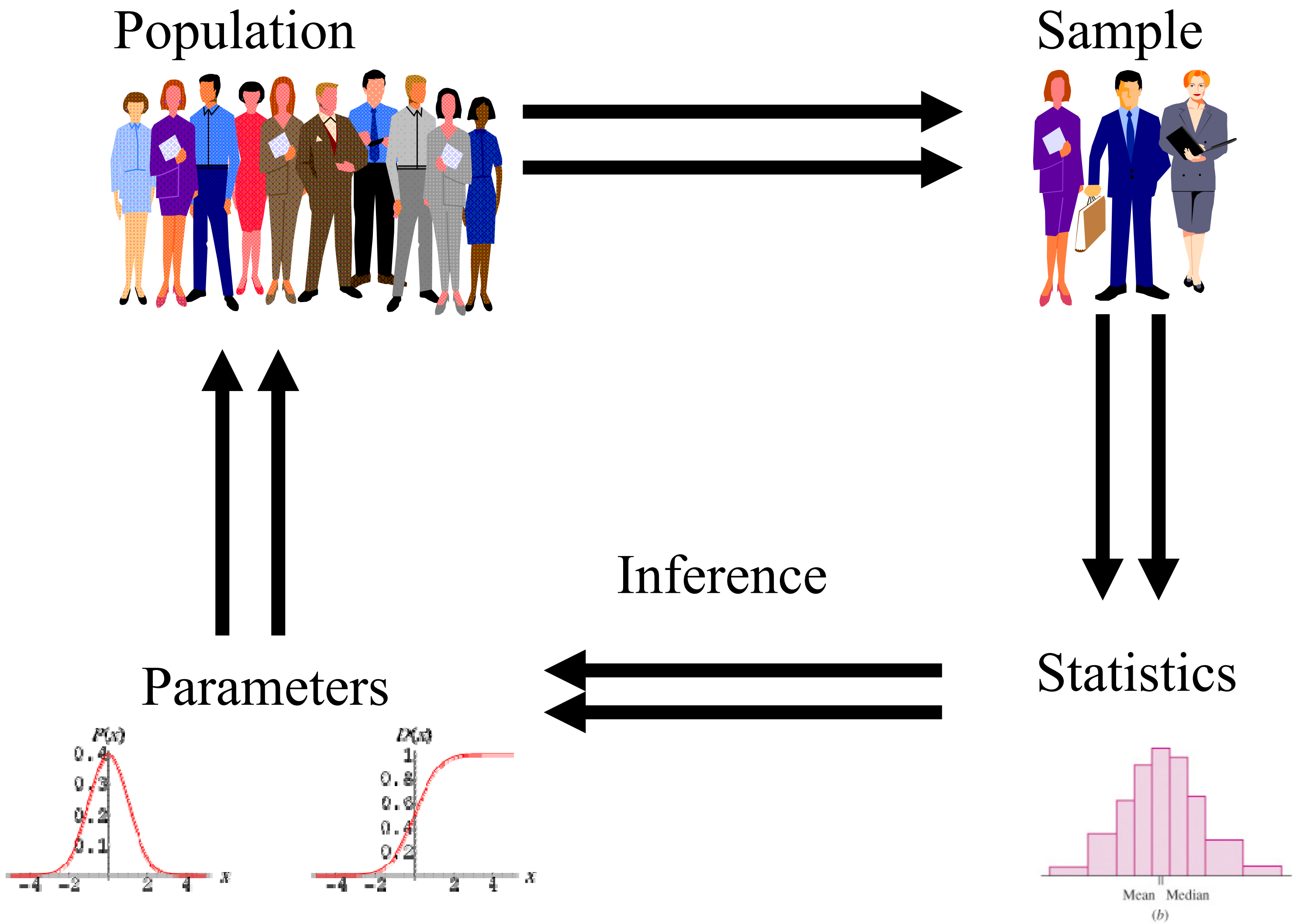
Quick Review of Probability

Berlin Chen
Department of Computer Science & Information Engineering
National Taiwan Normal University

References:
1. W. Navidi. *Statistics for Engineering and Scientists*. Chapter 2 & Teaching Material
2. D. P. Bertsekas, J. N. Tsitsiklis. *Introduction to Probability*.

Sample Statistics and Population Parameters



Basic Ideas

- Definition: An experiment is a process that results in an outcome that cannot be predicted in advance with certainty
  - Examples:
    - Rolling a die
    - Tossing a coin
    - Weighing the contents of a box of cereal
- Definition: The set of all possible outcomes of an experiment is called the sample space for the experiment
  - Examples:
    - For rolling a fair die, the sample space is {1, 2, 3, 4, 5, 6}
    - For a coin toss, the sample space is {heads, tails}
    - For weighing a cereal box, the sample space is $(0, \infty)$, a more reasonable sample space is (12, 20) for a 16 oz. box (with an infinite number of outcomes)

More Terminology

Definition: A subset of a sample space is called an **event**

- The empty set Ø is an event
- The entire sample space is also an event

- A given event is said to have occurred if the outcome of the experiment is one of the outcomes in the event. For example, if a die comes up 2, the events {2, 4, 6} and {1, 2, 3} have both occurred, along with every other event that contains the outcome "2"

Combining Events

- The union of two events A and B, denoted $A \cup B$, is the set of outcomes that belong either to A, to B, or to both
  - In words, $A \cup B$ means "A or B". So the event "A or B" occurs whenever either A or B (or both) occurs

- Example: Let $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$

  Then $A \cup B = \{1, 2, 3, 4\}$

Intersections

- The intersection of two events A and B, denoted by $A \cap B$, is the set of outcomes that belong to A and to B
  - In words, $A \cap B$ means “A and B”. Thus the event “A and B” occurs whenever both A and B occur
- Example: Let $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$

  Then $A \cap B = \{2, 3\}$

Complements

- The complement of an event A, denoted A^c, is the set of outcomes that do not belong to A
  - In words, A^c means "not A". Thus the event "not A" occurs whenever A does not occur

- Example: Consider rolling a fair sided die.

  Let A be the event: "rolling a six" = {6}.

  Then A^c = "not rolling a six" = {1, 2, 3, 4, 5}

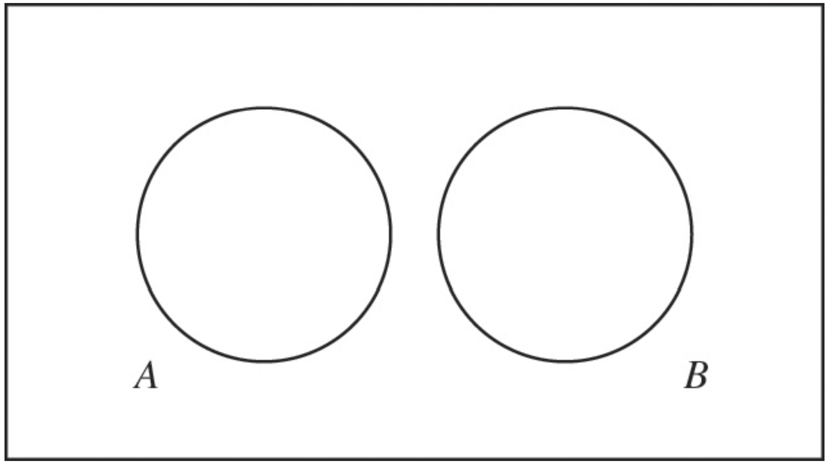
Mutually Exclusive Events

- Definition: The events A and B are said to be mutually exclusive if they have no outcomes in common
  - More generally, a collection of events $A_1, A_2, ..., A_n$ is said to be mutually exclusive if no two of them have any outcomes in common

- Sometimes mutually exclusive events are referred to as disjoint events

Example

- When you flip a coin, you cannot have the coin come up heads and tails
  - The following Venn diagram illustrates mutually exclusive events



Probabilities

- Definition: Each event in the sample space has a probability of occurring. Intuitively, the probability is a quantitative measure of how likely the event is to occur

- Given any experiment and any event A:
  - The expression $P(A)$ denotes the probability that the event A occurs
  - $P(A)$ is the proportion of times that the event A would occur in the long run, if the experiment were to be repeated over and over again

Axioms of Probability

1. Let S be a sample space. Then $P(S) = 1$

2. For any event A, $0 \leq P(A) \leq 1$

3. If A and B are mutually exclusive events, then
   $P(A \cup B) = P(A) + P(B)$

   More generally, if $A_1, A_2, ...$ are mutually exclusive events, then $P(A_1 \cup A_2 \cup ....) = P(A_1) + P(A_2) + ...$

A Few Useful Things

- For any event *A*, $P(A^c) = 1 - P(A)$
- Let Ø denote the empty set. Then $P(Ø) = 0$
- If *A* is an event, and $A = \{E_1, E_2, ..., E_n\}$ (and $E_1, E_2, ..., E_n$ are mutually exclusive), then

  $P(A) = P(E1) + P(E2) + \ldots + P(E_n)$.

- Addition Rule (for when *A* and *B* are not mutually exclusive):

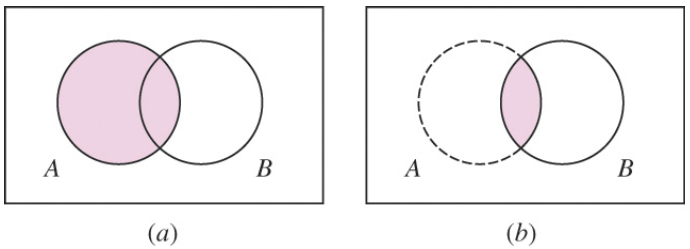
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Conditional Probability and Independence

- Definition: A probability that is based on a part of the sample space is called a conditional probability
  - E.g., calculate the probability of an event given that the outcomes from a certain part of the sample space occur

Let *A* and *B* be events with $P(B) \neq 0$. The conditional probability of *A* given *B* is

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



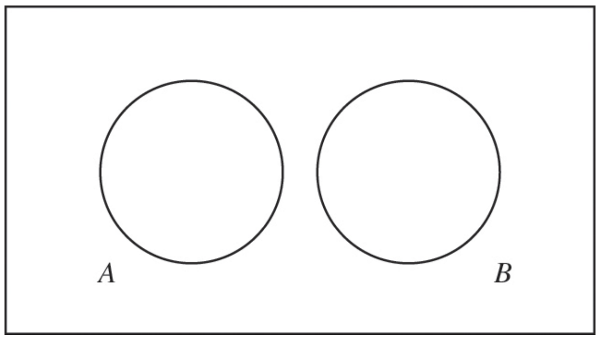
Venn diagram

More Definitions

- Definition: Two events A and B are independent if the probability of each event remains the same whether or not the other occurs

- If $P(B) \neq 0$ and $P(B) \neq 0$, then A and B are independent if $P(B|A) = P(B)$ or, equivalently, $P(A|B) = P(A)$

- If either $P(A) = 0$ or $P(B) = 0$, then A and B are independent



Are A and B independent (?)

The Multiplication (Chain) Rule

- If A and B are two events and $P(B) \neq 0$, then $P(A \cap B) = P(B)P(A|B)$

- If A and B are two events and $P(A) \neq 0$, then $P(A \cap B) = P(A)P(B|A)$

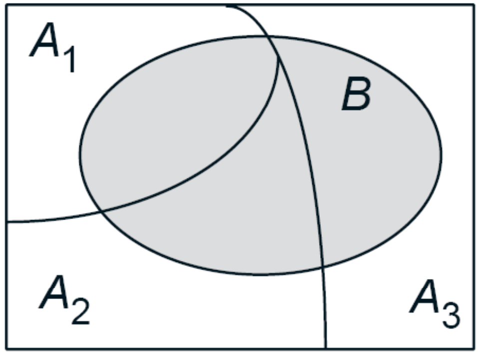
- If $P(A) \neq 0$, and $P(B) \neq 0$, then both of the above hold

- If A and B are two independent events, then $P(A \cap B) = P(A)P(B)$

- This result can be extended to more than two events

Law of Total Probability

- If $A_1, \ldots, A_n$ are mutually exclusive and exhaustive events, and B is any event, then



$$P(B) = P(A_1 \cap B) + \ldots + P(A_n \cap B)$$

  - Exhaustive events:
    - The union of the events cover the sample space

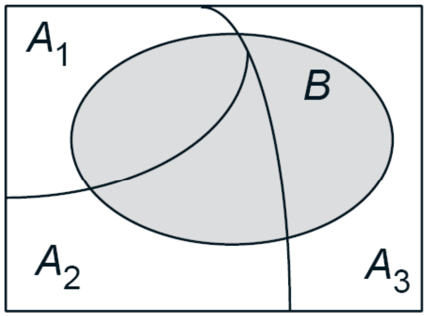
$$S = A_1 \cup A_2 \ldots \cup A_n$$

- Or equivalently, if $P(A_i) \neq 0$ for each A_i,

$$P(B) = P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n)$$

Example

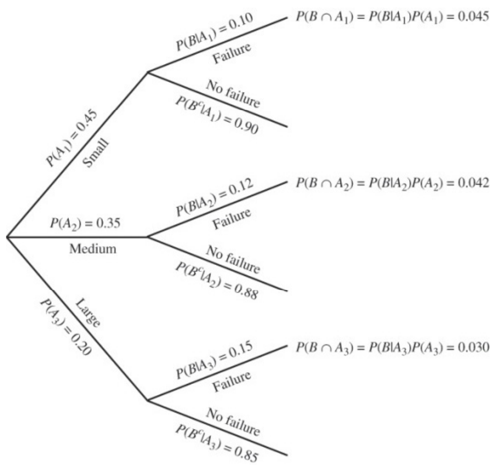
- Customers who purchase a certain make of car can order an engine in any of three sizes. Of all the cars sold, 45% have the smallest engine, 35% have a medium-sized engine, and 20% have the largest. Of cars with smallest engines, 10% fail an emissions test within two years of purchase, while 12% of those with the medium size and 15% of those with the largest engine fail. What is the probability that a randomly chosen car will fail an emissions test within two years?



Solution

- Let B denote the event that a car fails an emissions test within two years. Let A_1 denote the event that a car has a small engine, A_2 the event that a car has a medium size engine, and A_3 the event that a car has a large engine. Then $P(A_1) = 0.45$, $P(A_2) = 0.35$, and $P(A_3) = 0.20$. Also, $P(B|A_1) = 0.10$, $P(B|A_2) = 0.12$, and $P(B|A_3) = 0.15$. By the law of total probability,

$$
\begin{aligned}
P(B) &= P(B|A_1)\,P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)\,P(A_3) \\
&= 0.10(0.45) + 0.12(0.35) + 0.15(0.20) = 0.117
\end{aligned}
$$



Bayes' Rule

- Let $A_1,\ldots, A_n$ be mutually exclusive and exhaustive events, with $P(A_i) \neq 0$ for each A_i. Let B be any event with $P(B) \neq 0$. Then

$$P(A_k \mid B) = \frac{P(A_k \cap B)}{P(B)}$$

$$= \frac{P(B \mid A_k) P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i) P(A_i)}$$

Example

- The proportion of people in a given community who have a certain disease (*D*) is 0.005. A test is available to diagnose the disease. If a person has the disease, the probability that the test will produce a positive signal (+) is 0.99. If a person does not have the disease, the probability that the test will produce a positive signal is 0.01. If a person tests positive, what is the probability that the person actually has the disease?

Solution

- Let *D* represent the event that a person actually has the disease
- Let + represent the event that the test gives a positive signal
- We wish to find $P(D|+)$
- We know $P(D) = 0.005$, $P(+|D) = 0.99$, and $P(+|D^C) = 0.01$

- Using Bayes' rule

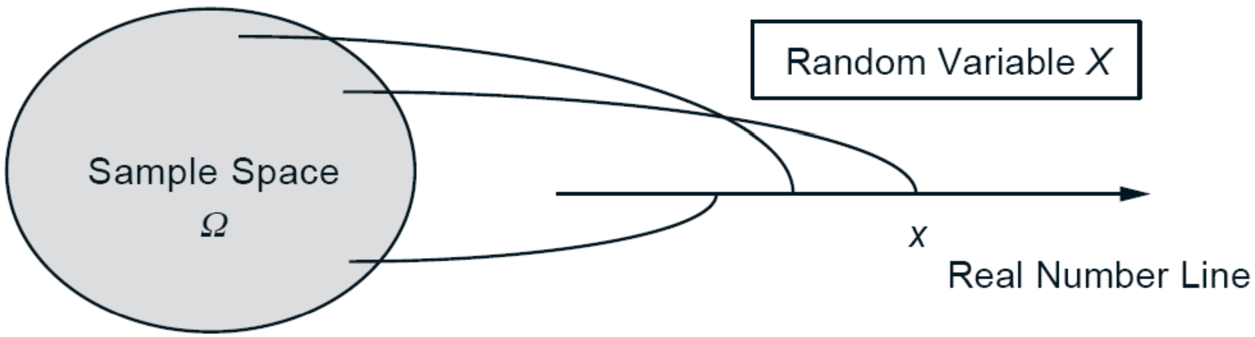
$$P(D \mid +) = \frac{P(+ \mid D)P(D)}{P(+ \mid D)P(D) + P(+ \mid D^C)P(D^C)}$$

$$= \frac{0.99(0.005)}{0.99(0.005) + 0.01(0.995)} = 0.332.$$

Random Variables

- Definition: A random variable assigns a numerical value to each outcome in a sample space
  - We can say a random variable is a real-valued function of the experimental outcome

- Definition: A random variable is discrete if its possible values form a discrete set



Example

- The number of flaws in a 1-inch length of copper wire manufactured by a certain process varies from wire to wire. Overall, 48% of the wires produced have no flaws, 39% have one flaw, 12% have two flaws, and 1% have three flaws. Let X be the number of flaws in a randomly selected piece of wire

- Then,
  - $P(X = 0) = 0.48$, $P(X = 1) = 0.39$, $P(X = 2) = 0.12$, and $P(X = 3) = 0.01$
  - The list of possible values 0, 1, 2, and 3, along with the probabilities of each, provide a complete description of the population from which X was drawn

Probability Mass Function

- The description of the possible values of X and the probabilities of each has a name:
  - The probability mass function

- Definition: The **probability mass function** (denoted as pmf) of a discrete random variable X is the function $p(x) = P(X = x)$. The probability mass function is sometimes called the **probability distribution**

Cumulative Distribution Function

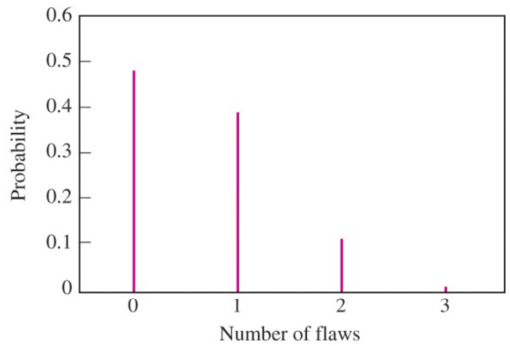
- The probability mass function specifies the probability that a random variable is equal to a given value

- A function called the **cumulative distribution function** (cdf) specifies the probability that a random variable is less than or equal to a given value

- The cumulative distribution function of the random variable X is the function $F(x) = P(X \leq x)$

Example

- Recall the example of the number of flaws in a randomly chosen piece of wire. The following is the pdf:
  - $P(X = 0) = 0.48$, $P(X = 1) = 0.39$, $P(X = 2) = 0.12$, and $P(X = 3) = 0.01$
- For any value x, we compute $F(x)$ by summing the probabilities of all the possible values of x that are less than or equal to x
  - $F(0) = P(X \leq 0) = 0.48$
  - $F(1) = P(X \leq 1) = 0.48 + 0.39 = 0.87$
  - $F(2) = P(X \leq 2) = 0.48 + 0.39 + 0.12 = 0.99$
  - $F(3) = P(X \leq 3) = 0.48 + 0.39 + 0.12 + 0.01 = 1$



More on Discrete Random Variables

- Let X be a discrete random variable. Then
  - The probability mass function (cmf) of X is the function $p(x) = P(X = x)$
  - The cumulative distribution function (cdf) of X is the function $F(x) = P(X \leq x)$

    $$F(x) = \sum_{t \leq x} p(t) = \sum_{t \leq x} P(X = t)$$

  - $\sum_{x} p(x) = \sum_{x} P(X = x) = 1$, where the sum is over all the possible values of X

Mean and Variance for Discrete Random Variables

- The mean (or expected value) of X is given by

$$\mu_X = \sum_x x P(X = x) \quad \text{, also denoted as } \mathbf{E}[X]$$

  where the sum is over all possible values of X

- The variance of X is given by

$$\sigma_X^2 = \sum_x (x - \mu_X)^2 P(X = x) \quad \text{, also denoted as } \mathbf{E}\left[(X - \mu_X)^2\right]$$

$$= \sum_x x^2 P(X = x) - \mu_X^2 . \quad \text{, also denoted as } \mathbf{E}\left[X^2\right] - (\mathbf{E}[X])^2$$

- The standard deviation is the square root of the variance

- Mean, variance, standard deviation provide summary information for a random variable (probability distribution)

The Probability Histogram

- When the possible values of a discrete random variable are evenly spaced, the probability mass function can be represented by a histogram, with rectangles centered at the possible values of the random variable

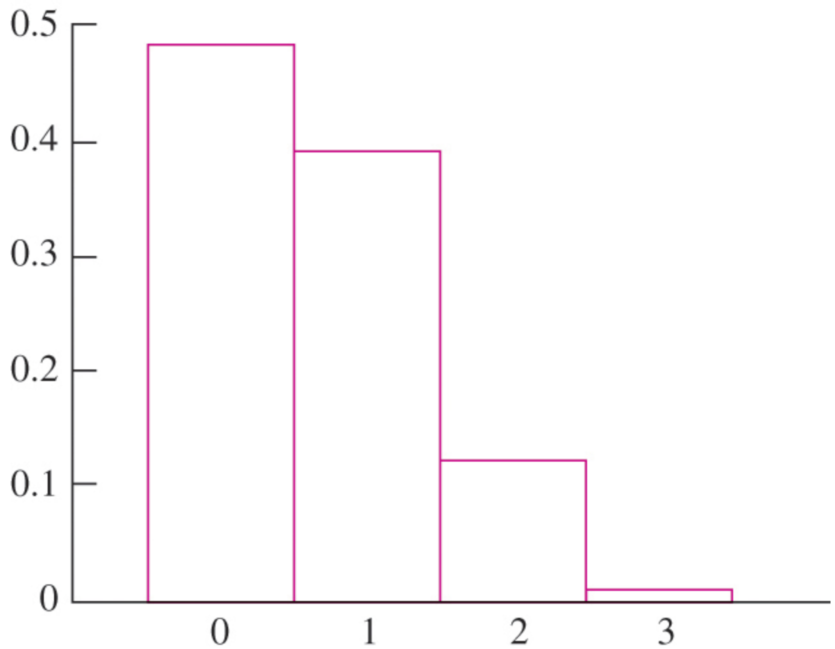
- The area of the rectangle centered at a value x is equal to $P(X = x)$

- Such a histogram is called a probability histogram, because the areas represent probabilities

Example

- The following is a probability histogram for the example with number of flaws in a randomly chosen piece of wire
  - $P(X = 0) = 0.48$, $P(X = 1) = 0.39$, $P(X = 2) = 0.12$, and $P(X = 3) = 0.01$

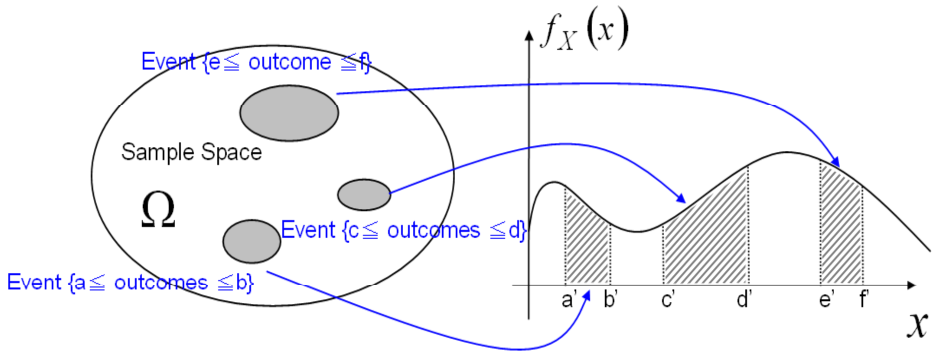
- Figure 2.8



Continuous Random Variables

- A random variable is continuous if its probabilities are given by areas under a curve
- The curve is called a probability density function (pdf) for the random variable. Sometimes the pdf is called the probability distribution
- Let *X* be a continuous random variable with probability density function *f*(*x*). Then

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$



Computing Probabilities

- Let *X* be a continuous random variable with probability density function *f*(*x*). Let *a* and *b* be any two numbers, with *a* < *b*. Then

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = \int_a^b f(x)dx.$$

- In addition,

$$P(X \leq a) = P(X < a) = \int_{-\infty}^{a} f(x)dx$$

$$P(X \geq a) = P(X > a) = \int_a^{\infty} f(x)dx.$$

More on Continuous Random Variables

- Let X be a continuous random variable with probability density function $f(x)$. The cumulative distribution function (cdf) of X is the function

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

- The mean of X is given by

$$\mu_X = \int_{-\infty}^{\infty} xf(x)dx. \qquad \text{, also denoted as } \mathbf{E}[X]$$

- The variance of X is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x-\mu_X)^2 f(x)dx \qquad \text{, also denoted as } \mathbf{E}\left[(X-\mu_X)^2\right]$$

$$= \int_{-\infty}^{\infty} x^2 f(x)dx - \mu_X^2. \qquad \text{, also denoted as } \mathbf{E}\left[X^2\right] - (\mathbf{E}[X])^2$$

Median and Percentiles

- Let X be a continuous random variable with probability mass function $f(x)$ and cumulative distribution function $F(x)$
  - The median of X is the point x_m that solves the equation

$$F(x_m) = P(X \le x_m) = \int_{-\infty}^{x_m} f(x)dx = 0.5.$$

  - If p is any number between 0 and 100, the pth percentile is the point x_p that solves the equation

$$F(x_p) = P(X \le x_p) = \int_{-\infty}^{x_p} f(x)dx = p/100.$$

  - The median is the 50th percentile

Linear Functions of Random Variables

- If X is a random variable, and a and b are constants, then

$$\mu_{aX+b} = a\mu_X + b$$

$$\sigma^2_{aX+b} = a^2\sigma^2_X$$

$$\sigma_{aX+b} = |a|\sigma_X$$

More Linear Functions

- If X and Y are random variables, and a and b are constants, then

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

- More generally, if $X_1, \ldots, X_n$ are random variables and $c_1, \ldots, c_n$ are constants, then the mean of the linear combination $c_1X_1, \ldots, c_nX_n$ is given by

$$\mu_{c_1X_1+c_2X_2+\ldots+c_nX_n} = c_1\mu_{X_1} + c_2\mu_{X_2} + \ldots + c_n\mu_{X_n}.$$

Two Independent Random Variables

- If X and Y are **independent** random variables, and S and T are sets of numbers, then

$$P(X \in S \text{ and } Y \in T) = P(X \in S)P(Y \in T).$$

- More generally, if X_1, …, X_n are independent random variables, and S_1, …, S_n are sets, then

$$P(X_1 \in S_1, X_2 \in S_2, \ldots, X_n \in S_n) = P(X_1 \in S_1)P(X_2 \in S_2)\ldots P(X_n \in S_n).$$

Variance Properties

- If X_1, …, X_n are *independent* random variables, then the variance of the sum X_1+ …+ X_n is given by

$$\sigma^2_{X_1+X_2+...+X_n} = \sigma^2_{X_1} + \sigma^2_{X_2} + ... + \sigma^2_{X_n}.$$

- If X_1, …, X_n are *independent* random variables and c_1, …, c_n are constants, then the variance of the linear combination $c_1 X_1$+ …+ $c_n X_n$ is given by

$$\sigma^2_{c_1X_1+c_2X_2+...+c_nX_n} = c_1^2\sigma^2_{X_1} + c_2^2\sigma^2_{X_2} + ... + c_n^2\sigma^2_{X_n}.$$

More Variance Properties

- If *X* and *Y* are *independent* random variables with variances σ_X^2 and σ_Y^2, then the variance of the sum *X* + *Y* is

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

The variance of the difference *X* – *Y* is

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

Independence and Simple Random Samples

- Definition: If $X_1, \ldots, X_n$ is a **simple random sample**, then $X_1, \ldots, X_n$ may be treated as independent random variables, all from the same population

  – Phrased another way, $X_1, \ldots, X_n$ are independent, and identically distributed (i.i.d.)

Properties of $\bar{X}$ (1/4)

- If $X_1, \ldots, X_n$ is a simple random sample from a population with mean μ and variance σ^2, then the sample mean $\bar{X}$ is a random variable with

mean of sample mean $$\mu_{\bar{X}} = \mu$$

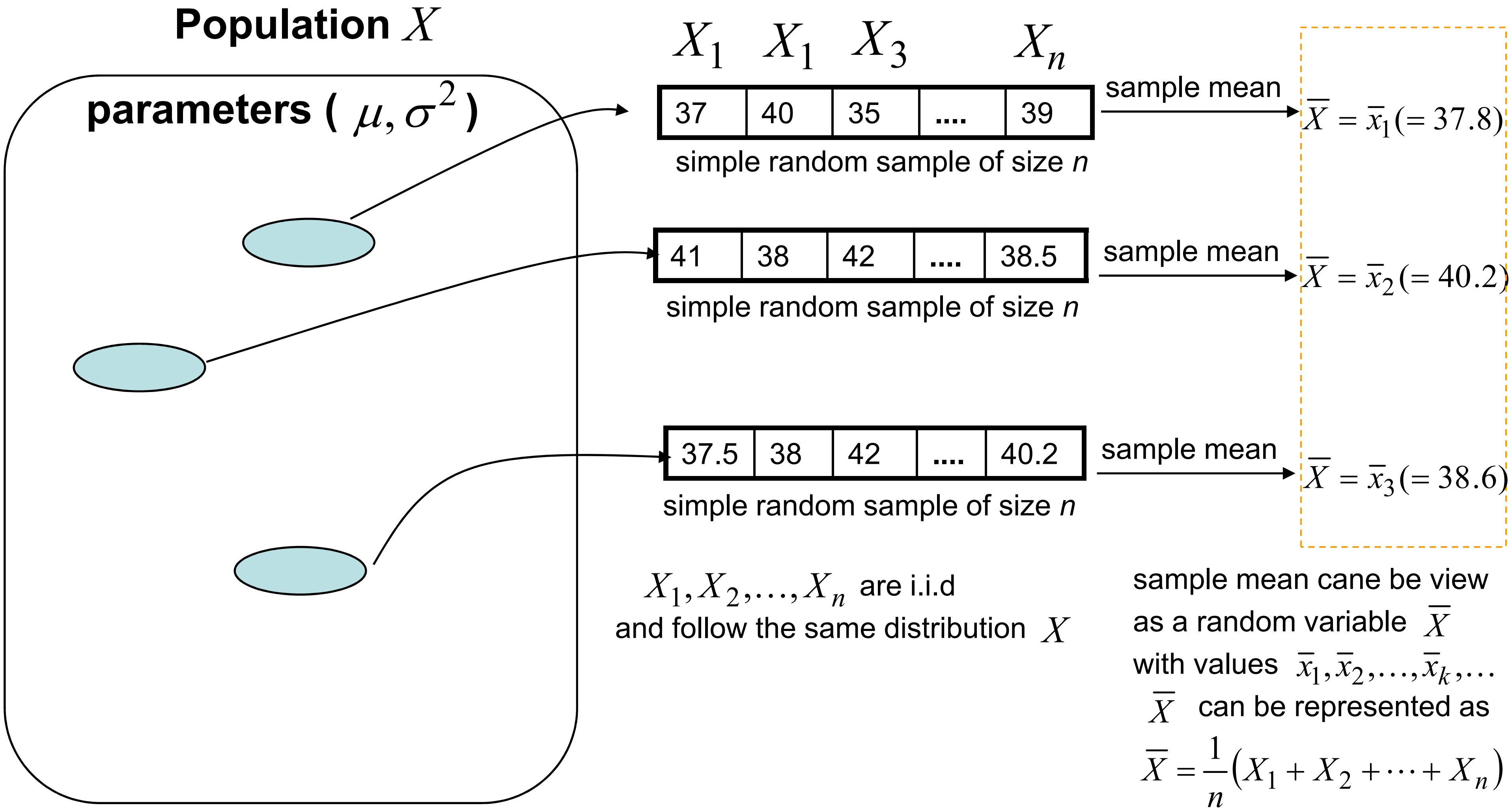
$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

variance of sample mean $$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}.$$

The standard deviation of $\bar{X}$ is

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

Properties of $\bar{X}$ (2/4)



$X_1, X_2, \ldots, X_n$ are i.i.d
and follow the same distribution X

sample mean cane be view as a random variable $\bar{X}$ with values $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k, \ldots$

$\bar{X}$ can be represented as

$$\bar{X} = \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right)$$

Properties of $\bar{X}$ (3/4)

$$\mu_{\bar{X}} = \mathbf{E}\left[\bar{X}\right]$$

$$= \mu_{\frac{1}{n}(X_1+X_2+\cdots+X_n)}$$

$$= \frac{1}{n}\mu_{X_1} + \frac{1}{n}\mu_{X_2} + \cdots + \frac{1}{n}\mu_{X_n}$$

$$= \frac{1}{n}\mu + \frac{1}{n}\mu + \cdots + \frac{1}{n}\mu$$

$$= \mu$$

$X_1, X_2, \ldots, X_n$ are i.i.d and follow the same distribution X with mean μ

$$\sigma^2_{\bar{X}} = \mathbf{E}\left[\left(\bar{X} - \mu_{\bar{X}}\right)^2\right]$$

$$= \sigma^2_{\frac{1}{n}(X_1+X_2+\cdots+X_n)}$$

$$= \frac{1}{n^2}\sigma^2_{X_1} + \frac{1}{n^2}\sigma^2_{X2} + \cdots + \frac{1}{n^2}\sigma^2_{X_n}$$

$$= \frac{1}{n^2}\sigma^2 + \frac{1}{n^2}\sigma^2 + \cdots + \frac{1}{n^2}\sigma^2$$

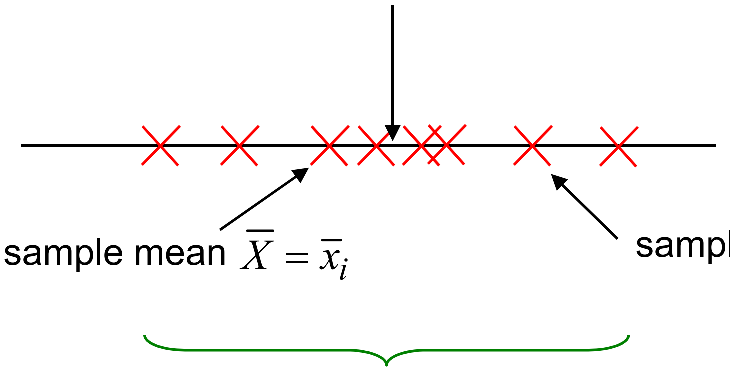
$$= \frac{\sigma^2}{n}$$

$X_1, X_2, \ldots, X_n$ are independent

$X_1, X_2, \ldots, X_n$ are identically distributed (follow the same distribution X with variance σ^2)

Properties of $\bar{X}$ (4/4)

mean of sample mean $\mu_{\bar{X}}$ (equal to population mean μ)



sample mean $\bar{X} = \bar{x}_i$

sample mean $\bar{X} = \bar{x}_j$

The spread of sample mean is determined by the variance of sample mean $\sigma^2_{\bar{X}}$ (equal to $\frac{\sigma^2}{n}$ where σ^2 is the population variance)

Jointly Distributed Random Variables

- If X and Y are jointly discrete random variables:
  - The joint probability mass function of *X* and *Y* is the function

$$p(x, y) = P(X = x \text{ and } Y = y)$$

  - The marginal probability mass functions of *X* and *Y* can be obtained from the joint probability mass function as follows:

$$p_X(x) = P(X = x) = \sum_y p(x, y) \quad p_Y(y) = P(Y = y) = \sum_x p(x, y)$$

    where the sums are taken over all the possible values of *Y* and of *X*, respectively (marginalization)

  - The joint probability mass function has the property that

$$\sum_x \sum_y p(x, y) = 1$$

    where the sum is taken over all the possible values of *X* and *Y*

Jointly Continuous Random Variables

- If X and Y are jointly continuous random variables, with joint probability density function $f(x,y)$, and $a < b$, $c < d$, then

$$P(a \le X \le b \text{ and } c \le Y \le d) = \int_a^b \int_c^d f(x,y)\,dy\,dx.$$

The joint probability density function has the property that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,dy\,dx = 1.$$

Marginals of *X* and *Y*

- If *X* and *Y* are jointly continuous with joint probability density function *f*(*x*,*y*), then the marginal probability density functions of *X* and *Y* are given, respectively, by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,dx.$$

  - Such a process is called “marginalization”

More Than Two Random Variables

- If the random variables X_1, …, X_n are jointly discrete, the joint probability mass function is

$$p(x_1,...,x_n) = P(X_1 = x_1,..., X_n = x_n).$$

- If the random variables X_1, …, X_n are jointly continuous, they have a joint probability density function $f(x_1, x_2, …, x_n)$, where

$$P(a_1 \le X_1 \le b_1,....,a_n \le X_n \le b_n) = \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1,...,x_n)dx_1...dx_n.$$

for any constants $a_1 \le b_1, …, a_n \le b_n$

Means of Functions of Random Variables (1/2)

- If the random variables $X_1, \ldots, X_n$ are jointly discrete, the joint probability mass function is

$$p(x_1,...,x_n) = P(X_1 = x_1,...,X_n = x_n).$$

- If the random variables $X_1, \ldots, X_n$ are jointly continuous, they have a joint probability density function $f(x_1, x_2, \ldots, x_n)$, where

$$P(a_1 \le X_1 \le b_1,....,a_n \le X_n \le b_n) = \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1,...,x_n)dx_1...dx_n.$$

for any constants $a_1 \le b_1, \ldots, a_n \le b_n$.

Means of Functions of Random Variables (2/2)

- Let X be a random variable, and let $h(X)$ be a function of X. Then:
  - If X is a discrete with probability mass function $p(x)$, then mean of $h(X)$ is given by

$$\mu_{h(x)} = \sum_{x} h(x)p(x). \text{ , also denoted as } \mathbf{E}[h(X)]$$

    where the sum is taken over all the possible values of X

  - If X is continuous with probability density function $f(x)$, the mean of $h(x)$ is given by

$$\mu_{h(x)} = \int_{-\infty}^{\infty} h(x)f(x)dx. \text{ , also denoted as } \mathbf{E}[h(X)]$$

Functions of Joint Random Variables

- If *X* and *Y* are jointly distributed random variables, and *h*(*X*,*Y*) is a function of *X* and *Y*, then

  – If *X* and *Y* are jointly discrete with joint probability mass function *p*(*x*,*y*),

$$\mu_{h(X,Y)} = \sum_{x} \sum_{y} h(x,y)\, p(x,y).$$

  where the sum is taken over all possible values of *X* and *Y*

  – If *X* and *Y* are jointly continuous with joint probability mass function *f*(*x*,*y*),

$$\mu_{h(X,Y)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)\, f(x,y)\, dx dy.$$

Discrete Conditional Distributions

- Let X and Y be jointly discrete random variables, with joint probability density function $p(x,y)$, let $p_X(x)$ denote the marginal probability mass function of X and let x be any number for which $p_X(x) > 0$.
  - The conditional probability mass function of Y given $X = x$ is

$$p_{Y|X}(y \mid x) = \frac{p(x, y)}{p(x)}.$$

  - Note that for any particular values of x and y, the value of $p_{Y|X}(y|x)$ is just the conditional probability $P(Y=y|X=x)$

Continuous Conditional Distributions

- Let X and Y be jointly continuous random variables, with joint probability density function $f(x,y)$. Let $f_X(x)$ denote the marginal density function of X and let x be any number for which $f_X(x) > 0$.
  - The conditional distribution function of Y given $X = x$ is

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f(x)}.$$

Conditional Expectation

- Expectation is another term for mean
- A **conditional expectation** is an expectation, or mean, calculated using the conditional probability mass function or conditional probability density function
- The conditional expectation of Y given $X = x$ is denoted by $E(Y|X = x)$ or $\mu_{Y|X}$

Independence (1/2)

- Random variables X_1, …, X_n are independent, provided that:
  - If X_1, …, X_n are jointly discrete, the joint probability mass function is equal to the product of the marginals:

$$p(x_1,...,x_n) = p_{X_1}(x_1)...p_{X_n}(x_n).$$

  - If X_1, …, X_n are jointly continuous, the joint probability density function is equal to the product of the marginals:

$$f(x_1,...,x_n) = f(x_1)...f(x_n).$$

Independence (2/2)

- If X and Y are independent random variables, then:
  - If X and Y are jointly discrete, and x is a value for which $p_X(x) > 0$, then

$$p_{Y|X}(y|x) = p_Y(y)$$

  - If X and Y are jointly continuous, and x is a value for which $f_X(x) > 0$, then

$$f_{Y|X}(y|x) = f_Y(y)$$

Covariance

- Let *X* and *Y* be random variables with means μ_X and μ_Y
  - The covariance of *X* and *Y* is

$$\operatorname{Cov}(X,Y) = \mu_{(X-\mu_X)(Y-\mu_Y)}.$$

  - An alternative formula is

$$\operatorname{Cov}(X,Y) = \mu_{XY} - \mu_X \mu_Y.$$

Correlation

- Let X and Y be jointly distributed random variables with standard deviations σ_X and σ_Y
  - The correlation between X and Y is denoted $\rho_{X,Y}$ and is given by

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

Or, called "correlation coefficient"

- For any two random variables X and Y

$$-1 \le \rho_{X,Y} \le 1.$$

Covariance, Correlation, and Independence

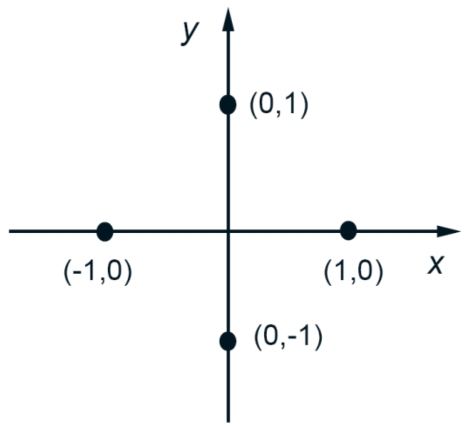
- If Cov(X, Y) = $\rho_{X,Y}$ = 0, then X and Y are said to be uncorrelated

- If X and Y are independent, then X and Y are uncorrelated

- It is mathematically possible for X and Y to be uncorrelated without being independent. This rarely occurs in practice

Example

- The pair of random variables (X, Y) takes the values $(1, 0)$, $(0, 1)$, $(-1, 0)$, and $(0, -1)$, each with probability ¼ Thus, the marginal pmfs of X and Y are symmetric around 0, and $\mathbf{E}[X] = \mathbf{E}[Y] = 0$
- Furthermore, for all possible value pairs (x, y), either x or y is equal to 0, which implies that $XY = 0$ and $\mathbf{E}[XY] = 0$. Therefore, $\mathrm{cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] = 0$, and X and Y are uncorrelated
- However, X and Y are not independent since, for example, a nonzero value of X fixes the value of Y to zero



Variance of a Linear Combination of Random Variables (1/2)

- If $X_1, \ldots, X_n$ are random variables and $c_1, \ldots, c_n$ are constants, then

$$\mu_{c_1 X_1 + \ldots + c_n X_n} = c_1 \mu_{X_1} + \ldots + c_n \mu_{X_n}$$

$$\sigma^2_{c_1 X_1 + \ldots + c_n X_n} = c_1^2 \sigma^2_{X_1} + \ldots + c_n^2 \sigma^2_{X_n} + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} c_i c_j \operatorname{Cov}(X_i, X_j).$$

For the case of two random variables

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2 \cdot \operatorname{Cov}(X, Y)$$

Variance of a Linear Combination of Random Variables (2/2)

- If $X_1, \ldots, X_n$ are *independent* random variables and $c_1, \ldots, c_n$ are constants, then

$$\sigma^2_{c_1 X_1 + \ldots + c_n X_n} = c_1^2 \sigma^2_{X_1} + \ldots + c_n^2 \sigma^2_{X_n}.$$

  – In particular,

$$\sigma^2_{X_1 + \ldots + X_n} = \sigma^2_{X_1} + \ldots + \sigma^2_{X_n}.$$

Summary (1/2)

- Probability and axioms (and rules)
- Counting techniques
- Conditional probability
- Independence
- Random variables: discrete and continuous
- Probability mass functions

Summary (2/2)

- Probability density functions
- Cumulative distribution functions
- Means and variances for random variables
- Linear functions of random variables
- Mean and variance of a sample mean
- Jointly distributed random variables