Introduction to Mathematical Finance

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 500

Subsection 1

Probabilities and Events

Experiments and Outcomes

- Consider an experiment.
- The **sample space** S is the set of all possible outcomes of the experiment.
- If there are m possible outcomes of the experiment, then we will generally number them 1 through m.
- So $S = \{1, 2, \ldots, m\}$.
- When dealing with specific examples, we will usually give more descriptive names to the outcomes.

Example

- Let the experiment consist of flipping a coin.
- Let the outcome be the side that lands face up.
- Thus, the sample space of this experiment is

$$S = \{h, t\},$$

where the outcome is:
  - h if the coin shows heads;
  - t if the coin shows tails.

Example

- Suppose the experiment consists of rolling a pair of dice.
- The outcome is the pair (i, j), where:
  - i is the value that appears on the first die;
  - j is the value that appears on the second die.
- Then the sample space consists of the following 36 outcomes:

$$\begin{array}{l}(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\(6,1),(6,2),(6,3),(6,4),(6,5),(6,6).\end{array}$$

Example

- Suppose the experiment consists of a race of r horses numbered $1, 2, 3, \ldots, r$.
- The outcome is the order of finish of these horses.
- Then the sample space is

$$S = \{\text{all orderings of the numbers } 1, 2, 3, \ldots, r\}.$$

- For instance, if $r = 4$ then the outcome is $(1, 4, 2, 3)$ if:
  - The number 1 horse comes in first;
  - The number 4 horse comes in second;
  - The number 2 horse comes in third;
  - The number 3 horse comes in fourth.

Probability

- Consider once again an experiment with the sample space

$$S = \{1, 2, \ldots, m\}.$$

- We will now suppose that there are numbers $p_1, \ldots, p_m$ with

$$p_i \geq 0,\ i = 1, \ldots, m, \quad \text{and} \quad \sum_{i=1}^{m} p_i = 1,$$

and such that p_i is the **probability** that i is the outcome of the experiment.

Example

- Consider again the experiment of flipping a coin.
- The coin is said to be **fair** or **unbiased** if it is equally likely to land on heads as on tails.
- Thus, for a fair coin we would have that $p_h = p_t = \frac{1}{2}$.
- If the coin were biased and heads were twice as likely to appear as tails, then we would have

$$p_h = \frac{2}{3}, \quad p_t = \frac{1}{3}.$$

Example

- Consider the experiment of rolling a pair of dice.
- If an unbiased pair of dice were rolled, then all possible outcomes would be equally likely.
- So
$$p(i,j) = \frac{1}{36}, \quad 1 \le i \le 6,\ 1 \le j \le 6.$$

Example

- Consider the experiment consisting of a race of three horses.
- Then we are given the six nonnegative numbers that sum to 1:

$$p_{1,2,3},\ p_{1,3,2},\ p_{2,1,3},\ p_{2,3,1},\ p_{3,1,2},\ p_{3,2,1},$$

where $p_{i,j,k}$ represents the probability that:
  - Horse i comes in first;
  - Horse j comes in second;
  - Horse k comes in third.

Events

- Any set of possible outcomes of the experiment is called an **event**.
- That is, an event is a subset of S, the set of all possible outcomes.
- For any event A, we say that A **occurs** whenever the outcome of the experiment is a point in A.
- Let $P(A)$ denote the probability that event A occurs.
- We can determine $P(A)$ by using the equation

$$P(A) = \sum_{i \in A} p_i.$$

- Note that this implies $P(S) = \sum_i p_i = 1$.
- In words, the probability that the outcome of the experiment is in the sample space is equal to 1, which, since S consists of all possible outcomes of the experiment, is the desired result.

Example

- Suppose the experiment consists of rolling a pair of fair dice.
- Let A be the event that the sum of the dice is equal to 7.
- Then
$$A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}.$$
- We have $P(A) = \frac{6}{36} = \frac{1}{6}$.
- Let B be the event that the sum is 8.
- Then
$$B = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}.$$
- So we have
$$P(B) = p(2,6) + p(3,5) + p(4,4) + p(5,3) + p(6,2) = \frac{5}{36}.$$

Example

- Consider a horse race between three horses.
- Let A denote the event that horse number 1 wins.
- Then
$$A = \{(1,2,3),(1,3,2)\}.$$
Moreover, we have
$$P(A) = p_{1,2,3} + p_{1,3,2}.$$

Complement of an Event

- For any event A, we let A^c, called the **complement** of A, be the event containing all those outcomes in S that are not in A.
- That is, A^c occurs if and only if A does not.
- We can show that
$$P(A^c) = 1 - P(A).$$
We have
$$P(A) + P(A^c) = \sum_{i \in A} p_i + \sum_{i \in A^c} p_i = \sum_i p_i = 1.$$
- So the probability that the outcome is not in A is 1 minus the probability that it is in A.

The Null Event

- The complement of the sample space S is the **null event** $\emptyset$, which contains no outcomes.
- We have
$$P(\emptyset) = 0.$$
In fact,
$$P(\emptyset) = 1 - P(\emptyset^c) = 1 - P(S) = 1 - 1 = 0.$$
- Alternatively, since $\emptyset$ contains no outcomes,
$$P(\emptyset) = \sum \emptyset = 0.$$

Union and Intersection

- For any events A and B, we define the **union** of A and B,

$$A \cup B,$$

as the event consisting of all outcomes that are in A, or in B, or in both A and B.

- Also, we define the **intersection** of A and B,

$$AB \quad \text{(sometimes written } A \cap B\text{)},$$

as the event consisting of all outcomes that are both in A and in B.

Example

- Let the experiment consist of rolling a pair of dice.
- Let:
  - A be the event that the sum is 10;
  - B be the event that both dice land on even numbers greater than 3.
- Then we have:
  - $A = \{(4,6),(5,5),(6,4)\}$;
  - $B = \{(4,4),(4,6),(6,4),(6,6)\}$.
- Therefore:

$$\begin{aligned} A \cup B &= \{(4,4),(4,6),(5,5),(6,4),(6,6)\}; \\ AB &= \{(4,6),(6,4)\}. \end{aligned}$$

Addition Theorem

- For any events A and B, we can write

$$P(A \cup B) = \sum_{i \in A \cup B} p_i, \quad P(A) = \sum_{i \in A} p_i, \quad P(B) = \sum_{i \in B} p_i.$$

Proposition (Addition Theorem of Probability)

If A and B are events, then

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

That is, the probability that the outcome of the experiment is in A or in B equals the probability that it is in A, plus the probability that it is in B, minus the probability that it is in both A and B.

- Every outcome in both A and B is counted twice in $P(A) + P(B)$ and only once in $P(A \cup B)$.

Example

- Suppose the probabilities that the Dow-Jones stock index increases today is 0.54, that it increases tomorrow is 0.54, and that it increases both days is 0.28.

  What is the probability that it does not increase on either day?

  Define the following events:
  - A is the event that the index increases today;
  - B is the event that it increases tomorrow.

  Then the probability that it increases on at least one of these days is

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(AB) \\ &= 0.54 + 0.54 - 0.28 = 0.80. \end{aligned}$$

  The probability that it increases on neither day is $1 - 0.80 = 0.20$.

Mutually Exclusive or Disjoint Events

- Events A and B are **mutually exclusive** or **disjoint** if

$$AB = \emptyset.$$

- That is, events are mutually exclusive if they cannot both occur.
- If A and B are mutually exclusive,

$$P(A \cup B) = P(A) + P(B).$$

We have

$$\begin{array}{rcl} P(A \cup B) & = & P(A) + P(B) - P(AB) \\ & = & P(A) + P(B) - P(\emptyset) \\ & = & P(A) + P(B) - 0 \\ & = & P(A) + P(B). \end{array}$$

Subsection 2

Conditional Probability

Example: Conditional Probability

- Suppose that each of two teams is to produce an item.
- The items will be rated as either acceptable or unacceptable.
- The sample space of this experiment will then be

$$S = \{(a, a), (a, u), (u, a), (u, u)\},$$

  where (a, u) means, for instance, that:
  - The first team produced an acceptable item;
  - The second team produced an unacceptable item.
- Suppose that the probabilities of these outcomes are

$$P(a, a) = 0.54, \ \ P(a, u) = 0.28, \ \ P(u, a) = 0.14, \ \ P(u, u) = 0.04.$$

- Suppose we are told that exactly one of the items was acceptable.
- What is the probability that it was produced by the first team?

Example: Conditional Probability (Cont'd)

- To determine this probability, consider the following reasoning.
- Given that there was exactly one acceptable item produced, it follows that the outcome of the experiment was either (a, u) or (u, a).
- The outcome (a, u) was initially twice as likely as the outcome (u, a).
- So it should remain twice as likely given the information that one of them occurred.
- Therefore:
  - The probability that the outcome was (a, u) is $\frac{2}{3}$;
  - The probability that the outcome was (u, a) is $\frac{1}{3}$.

Conditional Probability

- Let $A = \{(a, u), (a, a)\}$ denote the event that the item produced by the first team is acceptable.
- Let $B = \{(a, u), (u, a)\}$ be the event that exactly one of the produced items is acceptable.
- The probability that the item produced by the first team was acceptable given that exactly one of the produced items was acceptable is called the **conditional probability of** A **given** B.
- It is denoted as
$$P(A|B).$$

Definition of Conditional Probability

- A general formula for $P(A|B)$ is obtained by an argument similar to the one given above.
- If the event B occurs, then, in order for the event A to occur, it is necessary that the occurrence be a point in both A and B, i.e., in AB.
- Now, since we know that B has occurred, it follows that B can be thought of as the new sample space.
- Hence, the probability that the event AB occurs will equal the probability of AB relative to the probability of B.
- That is,

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

.

Example

- A coin is flipped twice, with all four points in the sample space $S = \{(h,h),(h,t),(t,h),(t,t)\}$ equally likely.
  What is the conditional probability that both flips land on heads, given that:
  (a) The first flip lands on heads?
  (b) At least one of the flips lands on heads?
  Define the following events.
  - $A = \{(h,h)\}$ is the event that both flips land on heads;
  - $B = \{(h,h),(h,t)\}$ is the event that the first flip lands on heads;
  - $C = \{(h,h),(h,t),(t,h)\}$ is the event that at least one of the flips lands on heads.

  We have the following solutions.

$$\begin{aligned}
P(A|B) &= \frac{P(AB)}{P(B)} = \frac{P(\{(h,h)\})}{P(\{(h,h),(h,t)\})} = \frac{1/4}{2/4} = \frac{1}{2}; \\
P(A|C) &= \frac{P(AC)}{P(C)} = \frac{P(\{(h,h)\})}{P(\{(h,h),(h,t),(t,h)\})} = \frac{1/4}{3/4} = \frac{1}{3}.
\end{aligned}$$

Example (Cont'd)

- Some are surprised that the answers to (a) and (b) are not identical.
- We provide a brief explanation.
- Conditional on the first flip landing on heads, the second one is still equally likely to land on either heads or tails.

  So the probability in Part (a) is $\frac{1}{2}$.
- On the other hand, knowing that at least one of the flips lands on heads is equivalent to knowing that the outcome is not (t, t).

  Thus, given that at least one of the flips lands on heads, there remain the three equally likely possibilities (h, h), (h, t), (t, h).

  This shows that the answer to Part (b) is $\frac{1}{3}$.

The Multiplication Theorem of Probability

- **Multiplication Theorem of Probability**: Given events, A and B,

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

By the definition of $P(A|B)$,

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

Multiplying by $P(B)$ gives the Multiplication Formula.

- So the probability that both A and B occur is the probability that B occurs multiplied by the conditional probability that A occurs given that B occurred.

Example

- Suppose that two balls are to be withdrawn, without replacement, from an urn that contains 9 blue and 7 yellow balls.

  If each ball drawn is equally likely to be any of the balls in the urn at the time, what is the probability that both balls are blue?

  Let B_1 and B_2 denote, respectively, the events that the first and second balls withdrawn are blue.

  Clearly, $P(B_1) = \frac{9}{16}$.

  Given that the first ball withdrawn is blue, the second ball is equally likely to be any of the remaining 15 balls, of which 8 are blue.

  Therefore, $P(B_2|B_1) = \frac{8}{15}$.

  By the Multiplication Theorem,

$$P(B_1 B_2) = P(B_1)P(B_2|B_1) = \frac{9}{16}\frac{8}{15} = \frac{3}{10}.$$

Independent Events

- The conditional probability of A given that B has occurred is not generally equal to the unconditional probability of A.
- In general, knowing that the outcome of the experiment is an element of B changes the probability that it is an element of A.
- We say that A is **independent** of B if

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

- We know that $P(A|B) = \frac{P(AB)}{P(B)}$.
- So A is independent of B if

$$P(AB) = P(A)P(B).$$

- This relation is symmetric in A and B.
- Hence, whenever A is independent of B, B is also independent of A.
- We say A and B are **independent events**.

Example

- Suppose that:
  - The probability that the closing price of a stock is at least as high as the close on the previous day is 0.52;
  - The results for successive days are independent.

Find the probability that the closing price goes down in each of the next four days, but not on the following day.

Let A_i be the event that the closing price goes down on day i.

Then, by independence, we have

$$\begin{aligned} P(A_1A_2A_3A_4A_5^c) &= P(A_1)P(A_2)P(A_3)P(A_4)P(A_5^c) \\ &= (0.48)^4(0.52) = 0.0276. \end{aligned}$$

Subsection 3

Random Variables and Expected Values

Random Variables

- Numerical quantities whose values are determined by the outcome of the experiment are known as **random variables**.

  Example:

(a) The sum obtained when rolling dice is a random variable.

(b) The number of heads that result in a series of coin flips is a random variable.

- Since the value of a random variable is determined by the outcome of the experiment, we can assign probabilities to each of its possible values.

Example

- Let X denote the sum when a pair of fair dice are rolled.
- The possible values of X are $2, 3, \ldots, 12$.
- We have the following associated probabilities:

$$\begin{array}{rcl}
P\{X=2\} & = & P\{(1,1)\} = \frac{1}{36}; \\
P\{X=3\} & = & P\{(1,2),(2,1)\} = \frac{2}{36}; \\
P\{X=4\} & = & P\{(1,3),(2,2),(3,1)\} = \frac{3}{36}; \\
P\{X=5\} & = & P\{(1,4),(2,3),(3,2),(4,1)\} = \frac{4}{36}; \\
P\{X=6\} & = & P\{(1,5),(2,4),(3,3),(4,2),(5,1)\} = \frac{5}{36}; \\
P\{X=7\} & = & P\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} = \frac{6}{36}; \\
P\{X=8\} & = & P\{(2,6),(3,5),(4,4),(5,3),(6,2)\} = \frac{5}{36}; \\
P\{X=9\} & = & P\{(3,6),(4,5),(5,4),(6,3)\} = \frac{4}{36}; \\
P\{X=10\} & = & P\{(4,6),(5,5),(6,4)\} = \frac{3}{36}; \\
P\{X=11\} & = & P\{(5,6),(6,5)\} = \frac{2}{36}; \\
P\{X=12\} & = & P\{(6,6)\} = \frac{1}{36}.
\end{array}$$

Probability Distribution of a Random Variable

- Let X be a random variable with possible values $x_1, x_2, \ldots, x_n$.
- The **probability distribution** of X is the set of probabilities

$$P\{X = x_j\}, \quad j = 1, \ldots, n.$$

- X must assume one of the values $x_1, x_2, \ldots, x_n$.
- It follows that

$$\sum_{j=1}^{n} P\{X = x_j\} = 1.$$

Expected Value of a Random Variable

Definition

If X is a random variable whose possible values are $x_1, x_2, \ldots, x_n$, then the **expected value** of X, denoted by $E[X]$, is defined by

$$E[X] = \sum_{j=1}^{n} x_j P\{X = x_j\}.$$

Alternative names for $E[X]$ are the **expectation** or the **mean** of X.

- In words, $E[X]$ is a weighted average of the possible values of X, where the weight given to a value is equal to the probability that X assumes that value.

Example: A Fair Bet

- Let the random variable X denote the amount that we win when we make a certain bet.

  Find $E[X]$ if there is:
  - A 60% chance that we lose 1;
  - A 20% chance that we win 1;
  - A 20% chance that we win 2.

  We have
  $$E[X] = -1(0.6) + 1(0.2) + 2(0.2) = 0.$$
  Thus, the expected amount that is won on this bet is equal to 0.
- A bet whose expected winnings is equal to 0 is called a **fair bet**.

Bernoulli Random Variables

- A **Bernoulli random variable with parameter** p is a random variable X, which is:
  - Equal to 1 with probability p;
  - Equal to 0 with probability $1 - p$.
- The expected value is

$$E[X] = 1(p) + 0(1 - p) = p.$$

Linearity of Expected Value

- For constants a and b,

$$E[aX+b]=aE[X]+b.$$

Let $Y=aX+b$.

Since Y will equal ax_j+b when $X=x_j$, it follows that

$$\begin{aligned} E[Y] &= \sum_{j=1}^n (ax_j+b)P\{X=x_j\} \\ &= \sum_{j=1}^n ax_j P\{X=x_j\}+\sum_{j=1}^n bP\{X=x_j\} \\ &= a\sum_{j=1}^n x_j P\{X=x_j\}+b\sum_{j=1}^n P\{X=x_j\} \\ &= aE[X]+b. \end{aligned}$$

Expectation of Sum of Random Variables

- An important result is that the expected value of a sum of random variables is equal to the sum of their expected values.

Proposition

For random variables $X_1, \ldots, X_k$,

$$E\left[\sum_{j=1}^{k} X_j\right] = \sum_{j=1}^{k} E[X_j].$$

Binomial Random Variables

- Consider n independent trials, each of which is a success with probability p.
- The random variable X, equal to the total number of successes that occur, is called a **binomial random variable** with parameters n, p.
- X has probability distribution

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, \ldots, n.$$

By independence, any sequence of trial outcomes resulting in i successes and $n - i$ failures has probability of occurrence

$$p^i (1-p)^{n-i}.$$

Moreover, there are $\binom{n}{i} = \frac{n!}{(n-i)!i!}$ such sequences.

Binomial Random Variables (Cont'd)

- We could compute the expected value of X by using the preceding to write
$$E[X] = \sum_{i=0}^{n} iP(X=i) = \sum_{i=0}^{n} i\binom{n}{i}p^i(1-p)^{n-i}$$
and then attempt to simplify it.
- It is easier to compute $E[X]$ by using $X = \sum_{j=1}^{n} X_j$, where X_j is defined to equal 1 if trial j is a success and to equal 0 otherwise.
- Using the preceding proposition, we obtain that
$$E[X] = E\left[\sum_{j=1}^{n} X_j\right] = \sum_{j=1}^{n} E[X_j] = np.$$
The final equality used the expected value of a Bernoulli random variable.

Equal Likelihood of Type of Success

Proposition

Consider n independent trials, each of which is a success with probability p. Then, given that there is a total of i successes in the n trials, each of the $\binom{n}{i}$ subsets of i trials is equally likely to be the set of trials that resulted in successes.

- Let T be any subset of size i of the set $\{1, \ldots, n\}$.

  Let A be the event that all of the trials in T were successes.

  Let X be the number of successes in the n trials.

  Then
  $$P(A|X = i) = \frac{P(A, X = i)}{P(X = i)}.$$

Equal Likelihood of Type of Success

- $P(A, X = i)$ is the probability that:
  - All trials in T are successes;
  - All trials not in T are failures.

  Using the independence of the trials, we obtain

$$P(A|X = i) = \frac{p^i(1-p)^{n-i}}{\binom{n}{i}p^i(1-p)^{n-i}} = \frac{1}{\binom{n}{i}}.$$

Independent Random Variables

- The random variables $X_1, \ldots, X_n$ are said to be **independent** if probabilities concerning any subset of them are unchanged by information as to the values of the others.

  Example: Consider the experiment of choosing randomly k balls from a set of N balls, n of which are red.

  Let
  $$X_i = \begin{cases} 1, & \text{if the } i\text{th ball chosen is red} \\ 0, & \text{if the } i\text{th ball chosen is black} \end{cases}.$$

  Then $X_1, \ldots, X_n$ would be:
  - Independent if each selected ball is replaced before the next selection is made.
  - Non independent if each selection is made without replacing previously selected balls.

Variance of a Random Variable

- Whereas the average of the possible values of X is indicated by its expected value, its spread is measured by its variance.

Definition

The **variance** of X, denoted by $\mathrm{Var}(X)$, is defined by

$$\mathrm{Var}(X) = E[(X - E[X])^2].$$

- In other words, the variance measures the average square of the difference between X and its expected value.

Variance of a Bernoulli Random Variable

Proposition

The variance of a Bernoulli random variable with parameter p is given by

$$\mathrm{Var}(X) = p - p^2.$$

- Because $E[X] = p$, we see that

$$(X - E[X])^2 = \begin{cases} (1-p)^2, & \text{with probability } p \\ p^2, & \text{with probability } 1-p \end{cases}$$

Hence,

$$\begin{aligned} \mathrm{Var}(X) &= E[(X - E[X])^2] \\ &= (1-p)^2 p + p^2(1-p) \\ &= p - p^2. \end{aligned}$$

Variance of Linear Expression

- If a and b are constants, then

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X).$$

We have

$$\begin{aligned}
\mathrm{Var}(aX+b) &= E[(aX+b-E[aX+b])^2] \\
&= E[(aX+b-aE[X]-b)^2] \\
&= E[(aX-aE[X])^2] \\
&= E[a^2(X-E[X])^2] \\
&= a^2\mathrm{Var}(X).
\end{aligned}$$

Variance of Sum of Independent Random Variables

- It is not generally true that the variance of the sum of random variables is equal to the sum of their variances.
- This holds, however, when the random variables are independent.

Proposition

If $X_1, \ldots, X_k$ are independent random variables, then

$$\mathrm{Var}\left(\sum_{j=1}^{k} X_j\right) = \sum_{j=1}^{k} \mathrm{Var}(X_j).$$

Variance of a Binomial Random Variable

Proposition

The variance of a binomial random variable X with parameters n and p is

$$\mathrm{Var}(X) = np(1-p).$$

- Recall that X represents the number of successes in n independent trials (each of which is a success with probability p).

  We can thus represent it as $X = \sum_{j=1}^{n} X_j$, where X_j is defined to equal 1 if trial j is a success and 0 otherwise.

  Hence,

$$\begin{aligned} \mathrm{Var}(X) &= \sum_{j=1}^{n} \mathrm{Var}(X_j) \\ &= \sum_{j=1}^{n} p(1-p) \\ &= np(1-p). \end{aligned}$$

Standard Deviation

- The square root of the variance is called the **standard deviation**.
- We will see, that a random variable tends to lie within a few standard deviations of its expected value.

Subsection 4

Covariance and Correlation

Covariance

- The **covariance** of any two random variables X and Y, denoted by $\mathrm{Cov}(X, Y)$, is defined by

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

- Upon multiplying the terms within the expectation, and then taking expectation term by term, it can be shown that

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y].$$

- A positive value of the covariance indicates that X and Y both tend to be large at the same time.
- A negative value indicates that when one is large the other tends to be small.
- Independent random variables have covariance equal to 0.

Example: Covariance for Bernoulli Random Variables

- Let X and Y both be Bernoulli random variables.
- Each takes on either the value 0 or 1.
- We use the identity $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$.
- Note that
$$XY = \begin{cases} 1, & \text{if } X = Y = 1 \\ 0, & \text{otherwise} \end{cases}.$$
So we get
$$\text{Cov}(X, Y) = P\{X = 1, Y = 1\} - P\{X = 1\}P\{Y = 1\}.$$
- From this, we see that
$$\begin{aligned} \text{Cov}(X, Y) > 0 \quad & \Leftrightarrow \quad P\{X = 1, Y = 1\} > P\{X = 1\}P\{Y = 1\} \\ & \Leftrightarrow \quad \tfrac{P\{X=1, Y=1\}}{P\{X=1\}} > P\{Y = 1\} \\ & \Leftrightarrow \quad P\{Y = 1 | X = 1\} > P\{Y = 1\}. \end{aligned}$$
- That is, the covariance of X and Y is positive if the outcome that $X = 1$ makes it more likely that $Y = 1$.

Properties of Covariance

- We can show that covariance has the following properties.
- For random variables X and Y, and constant c:

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathrm{Cov}(Y, X);\\
\mathrm{Cov}(X, X) &= \mathrm{Var}(X);\\
\mathrm{Cov}(cX, Y) &= c\mathrm{Cov}(X, Y);\\
\mathrm{Cov}(c, Y) &= 0.
\end{aligned}$$

Additivity of Covariance

- Covariance, like expected value, satisfies a linearity property:

$$\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y).$$

We have:

$$\begin{aligned}\mathrm{Cov}(X_1 + X_2, Y) &= E[(X_1 + X_2)Y] - E[X_1 + X_2]E[Y] \\ &= E[X_1 Y + X_2 Y] - (E[X_1] + E[X_2])E[Y] \\ &= E[X_1 Y] - E[X_1]E[Y] + E[X_2 Y] - E[X_2]E[Y] \\ &= \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y).\end{aligned}$$

- This can be generalized to:

$$\mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{Cov}(X_i, Y_j).$$

Variance of a Sum of Random Variables

- We get for the variance of the sum of random variables:

$$
\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) &= \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathrm{Cov}(X_i, X_i) + \sum_{i=1}^{n}\sum_{j\neq i} \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i=1}^{n}\sum_{j\neq i} \mathrm{Cov}(X_i, X_j).
\end{aligned}
$$

Correlation

- The **correlation** $\rho(X, Y)$ between random variables X and Y is defined by
$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$
- It reflects the degree to which large values of X tend to be associated with large values of Y.
- It can be shown that
$$-1 \leq \rho(X, Y) \leq 1.$$

Example

- If X and Y are linearly related by the equation $Y = a + bX$, then $\rho(X, Y)$ equals:
  - 1 when b is positive;
  - -1 when b is negative.

  We have:
  - $\text{Var}(Y) = b^2 \text{Var}(X)$;
  - $\text{Cov}(X, Y) = b\text{Var}(X)$.

  Therefore,

$$\rho(X, Y) = \frac{b\text{Var}(X)}{\sqrt{\text{Var}(X) b^2 \text{Var}(X)}} = \frac{b\text{Var}(X)}{|b|\text{Var}(X)} = \frac{b}{|b|}.$$

  This yields

$$\rho(X, Y) = \begin{cases} 1, & \text{if } b > 0 \\ -1, & \text{if } b < 0 \end{cases}.$$

Subsection 5

Conditional Expectation

Conditional Expectation

- For random variables X and Y, we define the conditional expectation of X given that $Y = y$ by

$$E[X|Y = y] = \sum_{x} xP(X = x|Y = y).$$

- Like the ordinary expectation of X, the conditional expectation of X given that $Y = y$ is a weighted average of the possible values of X.
- However, the value x is weighted not by the unconditional probability that $X = x$, but by its conditional probability given the information that $Y = y$.

Expected Value and Conditional Expectation

- The expected value of X is a weighted average of the conditional expectation of X given that $Y = y$.

Proposition

$$E[X] = \sum_y E[X|Y = y]P(Y = y).$$

- We have

$$\begin{aligned}\textstyle\sum_y E[X|Y = y]P(Y = y) & \\ &= \textstyle\sum_y \sum_x xP(X = x|Y = y)P(Y = y) \\ &= \textstyle\sum_y \sum_x xP(X = x, Y = y) \\ &= \textstyle\sum_x x \sum_y P(X = x, Y = y) \\ &= \textstyle\sum_x xP(X = x) \\ &= E[X].\end{aligned}$$

Conditional Expectation as a Function of Y

- Let $E[X|Y]$ be that function of the random variable Y which, when $Y = y$, is defined to equal $E[X|Y = y]$.
- The expected value of any function of Y , say $h(Y)$, can be expressed as
$$E[h(Y)] = \sum_y h(y)P(Y = y).$$
- So we have
$$E[E[X|Y]] = \sum_y E[X|Y = y]P(Y = y).$$
- Hence, the preceding proposition can be written as
$$E[X] = E[E[X|Y]].$$