Math 241: Introduction to Probability

Darshan Patel

Fall 2016

Contents

1 Probability

Definition 1.1. Experiment: anything that sets up a probabilistic situation

Definition 1.2. Outcome: any indivisible result of the experiment

Definition 1.3. Sample Space: the collection of all possible outcomes, represented by S

Example 1.1. Toss a coin. Outcomes: H, T; Sample space: S = $\{H, T\}$
Toss 2 coins. Outcomes: (H, H), (H, T), (T, H), (T, T); S = $\{(H, H), (H, T), (T, H), (T, T)\}$
Toss 2 dice. Outcomes: (1, 1), (1, 2), (1, 3),. . . , (6, 6); S = $\{(1, 1), (1, 2), (1, 3), \dots, (6, 6)\}$

Types of Sample Spaces

1. Finite: one where the number of outcomes is finite

2. Infinite Discrete: one where there is infinite outcomes but countable

   Example 1.2. Toss a coin until a head appears, S = $\{H, TH, TTH, TTH, \dots\}$

3. Infinite Nondiscrete: one where there is infinite outcomes

   Example 1.3. Select a real number between 0 and 1

Definition 1.4. Event: a collection of outcomes, represented by a letter

Example 1.4. Toss 2 coins. Let A be the event where the first coin is H. Then $A = \{(H, H), (H, T)\}$.
Toss 2 dice. Let B be the event where the sum equals 5. Then $B = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$.

Set Operations: Let A and B be two sets of outcomes:

1. $A \bigcup B$ is the set consisting of those outcomes that are in A or B or both

2. $A \bigcap B$ is the set consisting of those outcomes that are A and B

3. A' is the set consisting of those outcomes in S which are not in A, called the complement of A

Note: Intersection can be written as follows: $A \bigcap B$ or AB

Definition 1.5. Impossible event ($\emptyset$): the set consisting of zero outcomes

Example 1.5. Toss 2 dice. Let A be the event where the sum equal 1. Then $A = \{\} = \emptyset$.

Note: $\emptyset' = S, S' = \emptyset$
If every outcome in A is also an outcome in B, we say that A is a subset of B, $A \subset B$. This means that if A occurs, then B occurs.

Example 1.6. Toss 2 dice. Let A be the event that both dice are even and let B be the event where the sum is even. Then it is clear that $A \subset B$ but $B \not\subset A$.

If $A \subset B$ and $B \subset A$, then $A = B$.

Definition 1.6. A and B are disjoint or mutually exclusive if $A \bigcap B = \emptyset$.

Example 1.7. Select a card from an ordinary 52 card deck. Let A be the event a king is picked up, B be the event a queen is picked up and C be the event a heart is picked up. Then A and B are mutually exclusive and A and C are not mutually exclusive.

Properties:

1. Commutative Laws: $A \bigcup B = B \bigcup A$, $AB = BA$
2. Associative Laws: $(A \bigcup B) \bigcup C = A \bigcup (B \bigcup C)$, $(AB)C = A(BC)$
3. Distributive Laws: $A \bigcup (B \bigcap C) = (A \bigcup B) \bigcap (A \bigcup C)$, $A \bigcap (B \bigcup C) = (A \bigcap B) \bigcup (A \bigcap C)$
4. DeMorgan's Laws: $(A \bigcup B)' = A' \bigcap B'$, $(A \bigcap B)' = A' \bigcup B'$
5. $A \bigcup S = S$, $A \bigcap S = A$
6. $A \bigcup \emptyset = A$, $A \bigcap \emptyset = \emptyset$
7. $A \bigcup A' = S$, $A \bigcap A' = \emptyset$

Definition 1.7. Let S be the sample space of an experiment. The probability function is a function $P = S \to [0,1]$ which satisfies the following axioms:

1. If A is any event in S, $0 \leq P(A) \leq 1$
2. $P(S) = 1$
3. One of the following two:
   (a) If A and B are mutually exclusive, $P(A \bigcup B) = P(A) + P(B)$. This extends to any finite number of sets which are pair-wise mutually exclusive.
   (b) If $A_1, A_2, A_3, \dots$ is an infinite sequence of events which are pairwise mutually exclusive, $P(A_1 \bigcup A_2 \bigcup A_3 \dots) = P(A_1) + P(A_2) + P(A_3) + \dots$.

Theorem 1.1. $P(\emptyset) = 0$

Proof.

$$\begin{aligned} S \bigcap \emptyset &= \emptyset \\ P(S \bigcup \emptyset) &= P(S) + P(\emptyset) \\ P(S) &= P(S) + P(\emptyset) \\ 1 &= 1 + P(\emptyset) \\ P(\emptyset) &= 0 \end{aligned}$$

□

Theorem 1.2. $P(A') = 1 - P(A)$

Proof.

$$\begin{aligned} A \bigcap A' &= 0 \\ P(A \bigcup A') &= P(A) + P(A') \\ P(S) &= P(A) + P(A') \\ 1 &= P(A) + P(A') \\ P(A') &= 1 - P(A) \end{aligned}$$

□

Theorem 1.3. $P(AB') = P(A) - P(AB)$

Proof.

$$\begin{aligned} AB' \bigcap AB &= \emptyset \\ AB' \bigcup AB &= A \\ P(AB' \bigcup AB) &= P(A) \\ P(AB') + P(AB) &= P(A) \\ P(AB') &= P(A) - P(AB) \end{aligned}$$

□

Theorem 1.4. $P(A \bigcup B) = P(A) + P(B) - P(AB)$

Proof.

$$\begin{aligned} A \bigcup B &= AB' \bigcup AB \bigcup A'B \\ P(A \bigcup B) &= P(AB') + P(AB) + P(A'B) \\ P(A \bigcup B) &= P(A) - P(AB) + P(AB) + P(B) - P(AB) \\ P(A \bigcup B) &= P(A) + P(B) - P(AB) \end{aligned}$$

□

Example 1.8. A coin is tossed. Compute the probability it is heads.

Definition 1.8. Fairness assumption; every outcome in the sample space has an equal probability of occurrence. This will be noted by saying "fair." It will also be assumed from now on unless noted otherwise.

Example 1.9. A coin is tossed. Compute the probability it is heads.

$$\begin{aligned} H \bigcap T &= \emptyset \\ P(H \bigcup T) &= P(S) \\ P(H) + P(T) &= 1 \\ P(H) + P(H) &= 1 \\ 2P(H) &= 1 \\ P(H) &= \frac{1}{2} \end{aligned}$$

Example 1.10. A dice is tossed. Compute P(1). $P(1) = \frac{1}{6}$

Example 1.11. A dice is tossed. Compute P(odd).

$$\begin{aligned} P(\text{odd}) &= P(1 \text{ or } 3 \text{ or } 5) \\ P(\text{odd}) &= P(1) + P(3) + P(5) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{2} \end{aligned}$$

Theorem 1.5. If S contains N outcomes, each outcome having an equal probability of occurrence, and A is an event within S containing N_A outcomes, then $P(A) = \frac{N_A}{N}$.

Example 1.12. Select a card from an ordinary deck. Compute the probability it is a king or a heart.

$$\begin{aligned} P(K \bigcup H) &= P(K) + P(H) - P(KH) \\ &= \frac{4}{52} + \frac{13}{52} - \frac{1}{52} \\ &= \frac{16}{52} \end{aligned}$$

Example 1.13. Toss 2 dice. Compute the probability their sum equals 5. $P(\text{sum} = 5) = \frac{4}{36}$

Example 1.14. Toss 3 coins. Compute the probability that exactly 2 are heads. $P(A) = \frac{3}{8}$

Example 1.15. An urn contains 5 red balls numbered 1, 2, 3, 4, and 5, and 4 blue balls numbered 1, 2, 3, and 4. One ball is randomly selected. Compute the probability it is red or even. $P(R \bigcup E) = P(R) + P(E) - P(R \bigcap E) = \frac{5}{9} + \frac{4}{9} - \frac{2}{9} = \frac{7}{9}$

Example 1.16. 5 cards are selected from an ordinary deck. Compute the probability it is a royal flush. $P(RF) = \frac{4}{2598960}$

2 Combinatorial Methods

Definition 2.1. Fundamental Counting Principle: if an experiment has k parts and the i^{th} part has N_i possible outcomes, then the experiment has a total of $N = N_1 N_2 N_3 \dots N_i$ outcomes.

Example 2.1. A menu has 6 appetizers, 4 main courses and 5 desserts. How many different 3 course meals can be ordered?

$$\begin{aligned} k &= 3 \\ N_1 &= 6 \\ N_2 &= 4 \\ N_3 &= 5 \\ N &= N_1 N_2 N_3 = 120 \end{aligned}$$

Example 2.2. 4 dices are tossed. Compute the probability

1. all dice have odd values: $P = \frac{3^4}{6^4} = \frac{1}{16}$
2. all dice have the same value: $P = \frac{6\times1\times1\times1}{6^4} = \frac{1}{6^3}$
3. sum of the dices = 4: $P = \frac{1}{6^4}$
4. sum of the dices = 5: $P = \frac{4}{6^4}$

Definition 2.2. Permutations: an ordered selection of items without replacement

Example 2.3. An urn has 5 balls labeled 1, 2, 3, 4, 5. Select 3 balls without replacement. Some permutations are: $123, 231, 312, \dots, 345$ How many? $5 \cdot 4 \cdot 3 = 60$ permutations.

Example 2.4. A 3 digit number is formed by selecting 3 tags from 9 tags labeled 1, 2, 3, 4, 5, 6, 7, 8, 9 without replacement.

1. How many numbers can be formed: $9 \cdot 8 \cdot 7 = 504$
2. What is the probability that the number 123 is formed?: $\frac{1}{504}$
3. What is the probability that the number turns out to be less than 500?: $\frac{4\cdot8\cdot7}{504} = \frac{224}{504} = \frac{4}{9}$

4. What is the probability that the number turns out to be odd?: $\frac{8 \cdot 7 \cdot 5}{504} = \frac{5}{9}$

5. What is the probability that the number turns out to be odd and less than 500?: P(1 or 3, X, odd) + P(2 or 4, X, odd) = $\frac{2 \cdot 7 \cdot 4}{9 \cdot 8 \cdot 7} + \frac{2 \cdot 7 \cdot 5}{9 \cdot 8 \cdot 7} = \frac{126}{504} = \frac{1}{4}$

6. What is the probability that the number is odd OR less than 500?: $P(A \bigcup B) = P(A) + P(B) - P(AB) = \frac{5}{9} + \frac{4}{9} - \frac{1}{4} = \frac{3}{4}$

If k items are selected from n distinct items without replacement, the number of possible permutations is written as: $P(n, k)$ or P_k^n, where P stands for permutation, and it is evaluated as follows:

$$P_k^n = n(n-1)(n-2)\dots(n-k+1)$$

This equation has k terms.

Example 2.5. P(9, 3) = $9 \cdot 8 \cdot 7$, P(8, 5) = $8 \cdot 7 \cdot 6 \cdot 5 \cdot 4$

Note: If $k = n$, P(n,n) = $n(n-1)(n-2)\dot{1} = n!$. This means that $n!$ represents the number of possible rearrangements of n items without replacement.

Definition 2.3. Combinations: an unordered selection of items without replacement

Example 2.6. Select 3 letters from A, B, C, D, E without replacement. Some possible combinations are: ABC, ACB (both same), ABD, BCD, CBD (both same) ...
How many combinations are possible? $\frac{P_3^5}{3!} = \frac{60}{6} = 10$. All the possible combinations are: ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE.

Combinations can be written and computed as follows: C(n, k) = $C_k^n = \binom{n}{k} = \frac{P(n,k)}{n!} = \frac{n!}{k!(n-k)!}$.
The computational formula is: C(n, k) = $\frac{P(n,k)}{n!}$ while the theoretical formula is: $\frac{n!}{k!(n-k)!}$.

Example 2.7.

$$\binom{9}{3} = \frac{9!}{3!6!} = \frac{P_3^9}{3!}$$

Example 2.8. How many different 5 card hands are possible using an ordinary card deck? $\binom{52}{5} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5!} = \frac{P(52,5)}{5!} = 2,598,960$

1. P(royal flush) = $\frac{4}{2{,}598{,}960}$

2. P(4 aces) = $\frac{48}{2{,}598{,}960}$

3. P(4 of a kind) = $\frac{13 \cdot 48}{2{,}598{,}960}$

4. P(3 aces, 2 kinds) = $\frac{\binom{4}{3}\binom{4}{2}}{2{,}598{,}960} = \frac{4 \cdot 6}{2{,}598{,}960} = \frac{24}{2{,}598{,}960}$

5. P(full house) $= \frac{13\binom{4}{3}12\binom{4}{2}}{2{,}598{,}960} = \frac{3744}{2{,}598{,}960}$

6. P(flush) $= \frac{4\binom{13}{5}}{2{,}598{,}960}$

Theorem 2.1.

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

Proof.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k} = \frac{n!}{(n-k)!(n-(n-k))!} = \frac{n!}{(n-k)!\cdot k!} = \binom{n}{k}$$

□

Example 2.9.

$$\binom{100}{98} = \binom{100}{2} = \frac{100 \times 9}{2 \times 1} = 495$$

Theorem 2.2.

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

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

Proof.

$$\binom{n}{0} = \frac{n!}{0!(n-0)!} = \frac{n!}{1 \times n!} = 1$$

□

Theorem 2.3.

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

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

Proof.

$$\binom{n}{1} = \frac{n!}{1(n-1)!} = \frac{n!}{(n-1)!} = \frac{n}{1} = n$$

□

Theorem 2.4. Pascal's Theorem:

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$$

Proof.
$$\begin{aligned}
\binom{n}{k} + \binom{n}{k+1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-(k+1)!)} \\
&= \frac{n!}{k!(n-k)(n-k-1)!} + \frac{n!}{(k+1)k!(n-k-1)!} \\
&= \frac{n!}{k!(n-k-1)!}[\frac{1}{n-k} + \frac{1}{k+1}] \\
&= \frac{n!}{k!(n-k-1)!}\frac{k+1+n-k}{(n-k)(k+1)} \\
&= \frac{n!}{k!(n-k-1)!}\frac{(n+1)!}{(k+1)!((n+1)-(k+1))!} \\
&= \binom{n+1}{k+1}
\end{aligned}$$

$\square$

Theorem 2.5. Binomial Theorem:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

Example 2.10. $(x+y)^3 =$

$$(x+y)^3 = \binom{3}{0}x^3y^0 + \binom{3}{1}x^2y^1 + \binom{3}{2}x^1y^2 + \binom{3}{3}x^0y^3 = x^3 + 3x^2y + 3xy^2 + y^3$$

Example 2.11. $(2x+y)^4 =$

$$\begin{aligned}
(2x+y)^4 &= \binom{4}{0}(2x)^4y^0 + \binom{4}{1}(2x)^3y^1 + \binom{4}{2}(2x)^2y^2 + \binom{4}{3}(2x)^1y^3 + \binom{4}{4}(2x)^0y^4 \\
&= 16x^4 + 32x^3y + 24x^2y^2 + 8xy^3 + y^4
\end{aligned}$$

3 Conditional Probability and Independence

Example 3.1. An urn contains 5 red balls numbered 1 to 4, 6 blue balls numbered 1 to 6. Select 1 ball at random. Compute $P(2)$. $P(2) = \frac{2}{10} = \frac{1}{5}$

If you cheated and peeked for a red ball, compute $P(2)$. $P(2) = \frac{1}{4}$

The notation for conditional probability is $P[X|Y]$ where X is an event of something occurring and Y is the event of something occurring GIVEN.

Definition 3.1. If $P(B) \neq 0$, then $P[A|B] = \frac{P(AB)}{P(B)}$.

Example 3.2. An urn contains 5 red balls numbered 1 to 4, 6 blue balls numbered 1 to 6. Select 1 ball at random. Compute the probability where A: the ball is a 2 and B: the ball is red.

$$P[A|B] = \frac{P(AB)}{P(B)} = \frac{\frac{1}{10}}{\frac{4}{10}} = \frac{1}{4}$$

Note: Key words for conditional probability: GIVEN, IF

Example 3.3. Toss 2 coins. A: both coins are head, B: at least 1 coin is head

$$P[A|B] = \frac{P(AB)}{P(B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$
$$P[B|A] = \frac{P(BA)}{P(A)} = \frac{\frac{1}{4}}{\frac{1}{4}} = 1$$

Example 3.4. Toss 2 dice. A: exactly 1 die is a 1, B: sum = 7

$$P[A|B] = \frac{P(AB)}{P(B)} = \frac{\frac{2}{36}}{\frac{6}{36}} = \frac{1}{3}$$
$$P(A) = \frac{10}{36}$$

Example 3.5. Select 5 cards from an ordinary deck. A: 4 aces, B: exactly 1 picture card (KQJ)

$$P[A|B] = \frac{P(AB)}{P(B)} = \frac{\frac{4\times 3}{\binom{52}{5}}}{\frac{\binom{12}{1}\binom{40}{4}}{\binom{52}{5}}} = \frac{1}{\binom{40}{4}} = \frac{1}{91390}$$
$$P(A) = \frac{1}{52145}$$

If we know that $P[A|B] = \frac{P(AB)}{P(B)}$, then it is clear by rearrangement that

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

We also know that $P[B|A] = \frac{P(BA)}{P(A)}$ which then rearranges to

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

But since we know that $P(AB) = P(BA)$, then it is clear that:

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

Example 3.6. A urn has 7 red balls and 3 blue balls. Select 2 balls without replacement. Compute the probability that both balls are red.

$$P(R_1R_2) = P(R_2|R_1)P(R_1) = \frac{6}{9} \times 710 = \frac{42}{90}$$

Example 3.7. 2 defective calculators are mixed up with 3 good ones. They are going to be tested until the 2 defects are found. Compute the probability that the defects are found in the first 2 tests.

$$P(D_1D_2) = P(D_2|D_1)P(D_1) = \frac{1}{4} \times 25 = \frac{1}{10}$$

Definition 3.2. Let S be a sample space. A partition of S is a collection of mutually exclusive subsets whose union is all of S.

Example 3.8. Select a card. Let A: heads, B: clubs, C: diamonds, D: spades. Suppose A_0 be another event in S. Then:

$$A_0 = A_0A \bigcup A_0B \bigcup A_0C \bigcup A_0D$$

and

$$\begin{aligned} P(A_0) &= P(A_0A) + P(A_0B) + P(A_0C) + P(A_0D) \\ &= P(A_0|A)P(A) + P(A_0|B)P(B) + P(A_0|C)P(C) + P(A_0|D)P(D) \end{aligned}$$

Theorem 3.1. Theorem of Total Probability: In general, if $B_1, B_2, B_3, \dots B_n$ form a partition on S, and A is any event in S, then

$$P(A) = \sum_{k=1}^{n} P(A|B_k)P(B_k)$$

Example 3.9. Urn 1 has 7 red balls and 3 blue balls. Urn 2 has 4 red balls and 6 blue balls. Urn 3 has 2 red balls and 8 blue balls. Toss a die. If 1, 2, 3, pick a ball from urn 1. If 4, 5, pick a ball from urn 2. If 6, pick a ball from urn 3. Compute the probability that the ball is red.

$$\begin{aligned} P(R) &= P(R|U_1)P(U_1) + P(R|U_2)P(U_2) + P(R|U_3)P(U_3) \\ &= \frac{7}{10}\frac{3}{6} + \frac{4}{10}\frac{2}{6} + \frac{2}{10}\frac{1}{6} \\ &= \frac{21+8+2}{60} - \frac{31}{60} \end{aligned}$$

Compute the probability that the ball comes from urn 1 given that it is red.

$$P(U_1|R) = \frac{P(RU_1)}{P(R)} = \frac{P(R|U_1)P(U_1)}{P(R)} = \frac{\frac{21}{60}}{\frac{31}{60}} = \frac{21}{31}$$

Theorem 3.2. Baye's Theorem: If $B_1, B_2, B_3, \dots, B_n$ form a partition on S and A is any event in S, then

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum\limits_{i=1}^{n} P(A|B_i)P(B_i)}$$

Proof.

$$P(B_k|A) = \frac{P(B_kA)}{P(A)} = \frac{P(AB_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum\limits_{i=1}^{n} P(A|B_i)P(B_i)}$$

□

Definition 3.3. Let A and B be events in S. Suppose $P(A|B) = P(A)$ and $P(B|A) = P(B)$. Then we say that A and B are independent.

Since $P(A|B) = \frac{P(AB)}{P(B)}$ and A and B are independent, then $P(A) = \frac{P(AB)}{P(B)}$ and so $P(AB) = P(A)P(B)$. Similarly, since $P(B|A) = \frac{P(BA)}{P(A)}$ and A and B are independent, then $P(B) = \frac{P(BA)}{P(A)}$ and so $P(BA) = P(A)P(B)$.

This means that A and B are independent if

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

Recall: If A and B are mutually exclusive, then

$$P(A \bigcup B) = P(A) + P(B)$$

Note: To be mutually exclusive is when 2 events cannot occur simultaneously. To be independent is when 2 events cannot affect one another

Example 3.10. A coin is tossed and a die is thrown. Let A: heads on coin and B: 6 on the die. Prove that A and B are independent.

$$P(A) = \frac{1}{2}$$

$$P(B) = \frac{1}{6}$$

$$P(AB) = \frac{1}{2 \times 6} = \frac{1}{12}$$

Also,

$$P(AB) = P(A)P(B) = \frac{1}{2}\frac{1}{6} = \frac{1}{12}$$

This shows that A and B are independent.

Example 3.11. A coin is tossed 10 times. Compute the probability they are all heads.

$$P(HHH \dots H) = (\frac{1}{2})^1 0$$

Example 3.12. 5 coins are tossed. Compute the probability of obtaining 3 heads and 2 tails.

$$P(HHHTT) = (\frac{1}{2})^5$$

$$P(HTHTH) = (\frac{1}{2})^5$$

There are $\binom{5}{3}$ ways to position 3 heads. $P = \binom{5}{3}(\frac{1}{2})^5$. There are $\binom{5}{2}$ ways to position 2 tails. But $\binom{5}{3} = \binom{5}{2}$. Thus both will have equal probabilities.

Example 3.13. A pair of dice are tossed 8 times. Compute the probability their sum will be equal to 5 exactly 3 times.
On a single toss, $P(\text{sum} = 5) = \frac{4}{6^2} = \frac{1}{9}$. So, $P(\text{sum} \neq 5) = 1 - \frac{4}{6^2} = \frac{8}{9}$.
So let A: sum = 5 and A': sum $\neq$ 5. Then

$$P(AAAA'A'A'A'A') = (\frac{1}{9})^3(\frac{8}{9})^5$$

and the probability this will occur 3 times is:

$$P = \binom{8}{3}(\frac{1}{9})^3(\frac{8}{9})^5$$

4 Discrete Random Variables and Probability Functions

Definition 4.1. A random variable is a numerical function defined on a sample space

Definition 4.2. Let $f(x) = P[X = x]$. f is called either the probability function of X or probability distribution.

Example 4.1. Toss a coin.

$$X(H) = 1$$
$$X(T) = 0$$
$$f(1) = P[X = 1] = \frac{1}{2}$$
$$f(0) = P[X = 0] = \frac{1}{2}$$

x	$f(x)$
1	0
$\frac{1}{2}$	$\frac{1}{2}$

Example 4.2. Roll 2 dice. Let X = the sum of the dice.

$$X(2, 3) = 5$$
$$X(1, 5) = 6$$
$$X(i, j) = i + j$$

x	$f(x)$
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$

Example 4.3. Toss 3 coins. Let $X =$ the number of heads obtained.

$$X(HHT) = 2$$

$$X(THT) = 1 \text{ etc...}$$

x	f(x)
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

Note:

$$\sum f(x) = 1$$

Special Probability Distributions

Definition 4.3. Discrete Uniform Distribution:

$$f(x) = \begin{cases} \frac{1}{n} & \text{if } x = 1, 2, 3, \cdot, n \\ 0 & \text{otherwise} \end{cases}$$

Example 4.4. Toss a single dice.

x	$f(x)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

Definition 4.4. Bernoulli Distribution:

$$f(x) = \begin{cases} p & \text{if } x = 1 \\ q & \text{if } x = 0 \\ p + q = 1 \end{cases}$$

Example 4.5. Toss a single coin. If heads, it is a success (or X = 1). If tails, it is a failure (or X = 0).

$$f(1) = p = \frac{1}{2}$$

$$f(0) = q = \frac{1}{2}$$

Definition 4.5. Binomial Distribution: Consider any Bernoulli experiment (2 outcomes/ success or failure). Let $P(\text{success}) = p$ and $P(\text{failure}) = q$ and $p + q = 1$. Repeat the experiment n times under identical conditions (replication condition) in such a way that no repetition has any effect upon any other (independence condition). If X represents the number of successes in n trials, we say that X has a binomial distribution with parameters n and p.

Example 4.6. An urn contains 3 red balls and 2 blue balls. Select 3 balls with replacement.

Let x = number of red balls obtained. This means success is red, and failure is blue.

outcome	x	$f(x)$
BBB	0	$(\frac{2}{5})^3$
RBB	1	$(\frac{3}{5})(\frac{2}{5})^2$
BRB	1	$(\frac{3}{5})(\frac{2}{5})^2$
BBR	1	$(\frac{3}{5})(\frac{2}{5})^2$
BRR	2	$(\frac{3}{5})^2(\frac{2}{5})$
RBR	2	$(\frac{3}{5})^2(\frac{2}{5})$
RRB	2	$(\frac{3}{5})^2(\frac{2}{5})$
RRR	3	$(\frac{3}{5})^3$

Another way to write this in table form is:

x	$f(x)$
0	$\frac{8}{125}$
1	$\frac{36}{125}$
2	$\frac{54}{125}$
3	$\frac{27}{125}$

If we have n trials and x = number of successes, then $P[X = x] = \binom{n}{x} p^x q^{n-x}$ where p is the probability of success and q is the probability of failure.

Example 4.7. In the previous example,

$$f(0) = \binom{3}{0}(\frac{3}{5})^0(\frac{2}{5})^3 = \frac{8}{125}$$
$$f(1) = \binom{3}{1}(\frac{3}{5})^1(\frac{2}{5})^2 = \frac{36}{125}$$
$$f(2) = \binom{3}{2}(\frac{3}{5})^2(\frac{2}{5})^1 = \frac{54}{125}$$
$$f(3) = \binom{3}{3}(\frac{3}{5})^3(\frac{2}{5})^0 = \frac{27}{125}$$

Definition 4.6. Hypergeometric Distribution: Suppose n items are selected without replacement from N_1 items from one type and N_2 items of another type. The probability of selecting x items of the first type and $n-x$ items of the second type is

$$P = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N_1+N_2}{n}}$$

Example 4.8. An urn contains 4 red balls and 6 blue balls. 5 balls are selected without replacement. Find the probability function for x = number of red balls collected.

x	$f(x)$
0	$\frac{\binom{4}{0}\binom{6}{5}}{\binom{10}{5}}$
1	$\frac{\binom{4}{1}\binom{6}{4}}{\binom{10}{5}}$
2	$\frac{\binom{4}{2}\binom{6}{3}}{\binom{10}{5}}$
3	$\frac{\binom{4}{3}\binom{6}{2}}{\binom{10}{5}}$
4	$\frac{\binom{4}{4}\binom{6}{1}}{\binom{10}{5}}$

Theorem 4.1. If X has a binomial distribution, with parameters n and p,

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

for $x = 0, 1, 2, 3, \dots, n$

Observe that

1. $f(x) \geq 0$
2. $\sum\limits_{x=0}^{n} f(x) = \sum\limits_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = \sum\limits_{x=0}^{n} \binom{n}{x} q^{n-x} p^x = (q+p)^n = (q+p)^1 = 1^1 = 1$

If $N >> n$, the hypergeometric distribution $h(x, n, N_1, N_2)$ becomes nearly equal to the binomial distribution $b(x, n, p)$. That is to say, if there are many more selections made without replacement, then it resembles the probability of having selected with replacement and

$$h(x, n, N_1, N_2) \approx b(x, n, p)$$

where $p = \frac{N_1}{N_1+N_2}$.

Definition 4.7. Geometric Distribution: Repeat an experiment over and over again until success occurs. The probability of this happening on the x^{th} trial is:

$$P[X = x] = pq^{x-1}$$

where p is the probability of success and q is the probability of failure.

Example 4.9. Roll a dice until the number 6 occurs. Let x = number of rolls for when this occurs. $P[X = 6] = f(6) = \frac{1}{6}(\frac{5}{6})^{6-1}$

Definition 4.8. Poisson Distribution: $f(x) = \frac{\lambda^x e^{-x}}{x!}$ where $x = 0, 1, 2, \dots$.

Theorem 4.2. Poisson's Limit Law: If $\lambda = np$, $\lim_{n\to\infty} b(n, p) = p(x, \lambda)$.

Proof.

$$\begin{aligned}
b(x, n, p) &= \binom{n}{x} p^x q^{1-x} \\
&= \frac{P_x^n}{x!} p^x (1-p)^{n-x} \\
&= \frac{n(n-1)\dots(n-x+1)}{x!} (\frac{\lambda}{n})^x (1 - \frac{\lambda}{n})^{n-x} \\
&= \frac{n(n-1)\dots(n-x+1)}{x!} \frac{\lambda^x}{n^x} (1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-x} \\
&= \frac{n(n-1)\dots(n-x+1)}{n^x} \frac{\lambda^x}{x!} (1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-x} \\
&= (\frac{n}{n})(\frac{n-1}{n})\dots(\frac{n-x+1}{n}) \frac{\lambda^x}{x!} (1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-x} \\
&= (1)(1 - \frac{1}{n})(1 - \frac{2}{n})\dots(1 - \frac{x-1}{n}) \frac{\lambda^x}{x!} (1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-x}
\end{aligned}$$

Let $n \to \infty$. Then $1 - \frac{1}{n} \to 1$, $1 - \frac{2}{n} \to 1$ and so forth up to $1 - \frac{x-1}{n} \to 1$. Also, $(1 - \frac{\lambda}{n})^{-x} \to 1$ and $(1 - \frac{\lambda}{n})^n = e^{-x}$. Thus,

$$\lim_{n\to\infty} b(x, n, p) = \frac{\lambda^x e^{-\lambda}}{x!}$$

□

Example 4.10. Let $n = 50,000$, $p = 0.0001$, and $x = 10$. Then:
Binomial: $b(10, 50000, 0.0001) = \binom{50000}{10}(0.0001)^10(0.9999)^49990 = 0.018130$
Poisson:$\lambda = np = (50000)(0.0001) = 5$ $p(10, 5) = \frac{5^10 e^{-5}}{10!} = 0.0018133$

In general, if n is large and p is small, the Poisson distribution gives an excellent approximation to the binomial distribution.

Definition 4.9. Cumulative Probability Distribution: Given a random variable X, we define $F(X) = P[X \leq x]$

Example 4.11. An urn has 10 balls: one 1, 2 twos, 4 threes, 3 fours. Select one ball and let X = value on the ball. Then:

$$f(2) = P[X = 2] = \frac{2}{10}$$

but

$$F(2) = P[X \leq 2] = \frac{3}{10}$$

x	$f(x)$	$F(x)$
1	0.1	0.1
2	0.2	0.3
3	0.4	0.7
4	0.3	1.0

Note: $f(2.5) = 0$ BUT $F(2.5) = \frac{3}{10}$.

$F(X)$ is a step function that is nondecreasing

- $\lim_{x\to\infty} F(X) = 1$
- $\lim_{x\to\infty} f(x) = 0$

Example 4.12. A pair of dice is rolled 1000 times. Compute the probability their sum is 5 between 100 and 125 times inclusive (including 100 and 125).
Binomial: $n = 1000, p = \frac{4}{36} = \frac{1}{9}$

$$\text{binomcdf}(1000, \frac{1}{9}, 125) - \text{binomcdf}(1000, \frac{1}{9}, 99) = 0.943 - 0.1203 = 0.804$$

Example 4.13. If X has a Poisson distribution with $\lambda = 7$, compute $P[X \geq 10]$.

$$P[X \geq 10] = 1 - \text{poissoncdf}(7, 9) = 1 - 0.7166 = 0.2034$$

Definition 4.10. A probability density function is a function which satisfies:

1. $f(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

if a random variable X has the probability density function $f(x)$:

$$P[a \leq x \leq b] = \int_a^b f(x)dx$$

Example 4.14. Let X be a random variable whose probability density function is

$$f(x) = \begin{cases} \frac{3}{4}(2x - x^2) & \text{if } 0 \leq x \leq 2 \\ 0 & \text{elsewhere} \end{cases}$$

Show that $f(x)$ is a valid probability density function and compute $P[0 \leq X \leq 1]$ and $P[X \geq \frac{1}{2}]$.

1. If $x < 0$ or $x > 2$, $f(x) = 0$. If $0 \leq x \leq 2$, $f(x) = \frac{3}{4}x(2-x)$. So, $f(x) \geq 0$ for all x.

2. $$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{0} f(x)dx + \int_{0}^{2} f(x)dx + \int_{2}^{\infty} f(x)dx \\ &= \int_{-\infty}^{0} 0dx + \int_{0}^{2} \frac{3}{4}(2x - x^2)dx + \int_{2}^{\infty} 0dx \\ &= \int_{0}^{2} \frac{3}{4}(2x - x^2)dx \\ &= \frac{3}{2}[x^2 - \frac{x^3}{3}]\Big|_0^2 \\ &= \frac{3}{4}(4 - \frac{8}{3}) = 1 \end{aligned}$$

$$P[0 \leq X \leq 1] = \int_0^1 \frac{3}{4}(2x - x^2)dx = \frac{3}{2}[x^2 - \frac{x^3}{3}]\Big|_0^1 = \frac{3}{4}(1 - \frac{1}{3}) = \frac{1}{2}$$

$$P[X \geq \frac{1}{2}] = \int_{\frac{1}{2}}^2 \frac{3}{4}(2x - x^3)dx = \frac{3}{4}[x^2 - \frac{x^3}{3}]\Big|_{\frac{1}{2}}^2 = \frac{3}{4}[(14 - \frac{8}{3}) - (\frac{1}{4} - \frac{1}{24})] = \frac{27}{32}$$

Example 4.15. Let $f(x) = \begin{cases} e^{-x} & \text{if} x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$ Is this a probability density function? Compute $P[0 \leq x \leq 1]$.

1. e^{-x} for $x > 0$ so $f(x) > 0$, 0 for $x < 0$. Thus $f(x) \geq 0$ for all x

2. $$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= \int_0^{\infty} e^{-x}dx \\ &= \lim_{t\to\infty} \int_0^t e^{-x}dx \\ &= \lim_{t\to\infty} -e^{-x}\Big|_0^t \\ &= \lim_{t\to\infty}(-e^{-t} + 1) = 1 \end{aligned}$$

$$P[0 \leq x \leq 1] = \int_0^1 e^{-x}dx = -e^{-x}\Big|_0^1 = -e^{-1} + 1 = 1 - \frac{1}{e}$$

Suppose X is a continuous random variable with the probability density function $f(x)$.

- $P[X = a] = \int_a^a f(x)dx = 0$
- $P[a \leq X \leq b] = P[a < X \leq b]$
- $P[a \leq X \leq b] = P[a \leq X < b]$
- $P[a \leq X \leq b] = P[a < X < b]$

Definition 4.11. Cumulative Probability Function: $F(x) = P[X \leq x] = \int_{-\infty}^{x} f(t)dt$

Example 4.16. Find $F(x)$ if $f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$ and $P[0.5 \leq X \leq 0.6]$.

- If $x < 0, F(x) = 0$
- If $x > 1, F(x) = 1$
- If $0 \leq x \leq 1, F(x) = \int_{-\infty}^{x} f(t)dt = \int_0^x 2tdt = t^2\Big|_0^x = x^2$

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

$$P[0.5 \leq X \leq 0.6] = F(0.6) - F(0.5) = 0.36 - 0.25 = 0.11$$

Example 4.17. Find $F(x)$ if $f(x) = \begin{cases} \frac{3}{4}(2x - x^2) & \text{if } 0 \leq x \leq 2 \\ 0 & \text{elsewhere} \end{cases}$ and $P[0 \leq X \leq 1]$.

- If $x < 0, F(x) = 0$
- If $x > 2, F(x) = 1$
- If $0 \leq x \leq 2, F(x) = \int_0^t \frac{3}{4}(2t - t^2)dt = \frac{3}{4}[t^2 - \frac{t^3}{3}]\Big|_0^x = \frac{3}{4}(x^2 - \frac{x^3}{3})$

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{3}{4}(x^2 - \frac{x^3}{3}) & \text{if } 0 \leq x \leq 2 \\ 1 & \text{if } x > 2 \end{cases}$$

$$P[0 \leq X \leq 1] = F(1) - F(0) = \frac{3}{4}(1 - \frac{1}{3}) - 0 = \frac{1}{2}$$

Properties of $F(x)$:

- continuous everywhere
- nondecreasing
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$
- $F'(x) = f(x)$ everywhere $F(x)$ is defined

5 Continuous Random Variables and Probability Density Functions

Important Continuous Distributions

Definition 5.1. Uniform Continuous Distribution:

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } \alpha \leq x \leq \beta \\ 0 & \text{elsewhere} \end{cases}$$

Example 5.1. Let X be a random variable and let $f(x) = \begin{cases} \frac{1}{4} & \text{if } 1 \leq x \leq 5 \\ 0 & \text{elsewhere} \end{cases}$. Compute $P[2 \leq x \leq 4]$ and $P[0 \leq x \leq 4]$.

$$P[2 \leq x \leq 4] = \int_2^4 \frac{1}{4} dx = \frac{1}{2}$$

$$P[0 \leq x \leq 4] = \int_0^1 0 dx + \int_1^4 \frac{1}{4} dx = \frac{3}{4}$$

Find $F(x)$.

$$F(x) = \int_{\infty}^x f(t) dt = \int_1^x \frac{1}{4} dt = \frac{1}{4} t \Big|_1^x = \frac{1}{4}(x-1)$$

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{4}(x-1) & 1 \leq x \leq 5 \\ 1 & x > 5 \end{cases}$$

$$P[2 \leq x \leq 4] = F(4) - F(2) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$$

$$P[0 \leq x \leq 4] = F(4) - F(0) = \frac{3}{4} - 0 = \frac{3}{4}$$

Definition 5.2. Exponential Distribution: parameter $\theta > 0$

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{if } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Is this a probability density function?

- $f(x) \geq 0$ for all x
- $\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} \frac{1}{\theta} e^{-\frac{x}{\theta}} dx = \lim_{t\to\infty} \int_0^t \frac{1}{\theta} e^{-\frac{x}{\theta}} dx = \lim_{t\to\infty} (-e^{-\frac{x}{\theta}}) \Big|_0^t = \lim_{t\to\infty}(-e^{-\frac{t}{\theta}} + e^0) = 0 + 1 = 1$

Yes, the exponential distribution is a probability density function.

Example 5.2. Suppose X has an exponential distribution with parameter $\theta = 2$. Compute $P[2 \leq x \leq 4]$.

$$P[2 \leq x \leq 4] = \int_2^4 \frac{1}{2} e^{-\frac{x}{2}} dx = -e^{-\frac{x}{2}} \Big|_2^4 = -e^{-2} + e^{-1} = \frac{1}{e} - \frac{1}{e^2} = 0.2325$$

Compute $F(x)$.

$$F(x) = \int_0^x \frac{1}{2} e^{-\frac{t}{2}} dt = -e^{-\frac{t}{2}} \Big|_0^x = -e^{-\frac{x}{2}} + 1 = 1 - e^{-\frac{x}{2}}$$

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\frac{x}{2}} & \text{if } x \geq 0 \end{cases}$$

$$P[2 \leq x \leq 4] = F(4) - F(2) = (1 - e^{-2}) - (1 - e^{-1}) = e^{-1} - e^{-2} = \frac{1}{e} - \frac{1}{e^2}$$

Definition 5.3. Normal Distribution: parameters μ, σ where $-\infty < \mu < \infty$ and $\sigma > 0$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \text{ for all } x$$

Definition 5.4. Standard Normal Distribution: $\mu = 0, \sigma = 1$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

Properties of the Standard Normal Distribution:

- Symmetry with respect to the y axis
- $f(x) > 0$ for all x
- $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1$

Example 5.3. Suppose X has a standard normal distribution. Compute $P[0 \leq x \leq 1]$.

$$P[0 \leq x \leq 1] = 0.341 \text{ use normalcdf}$$

Find x such that $P[x \leq X] = 0.7$.

$$x = 0.5244 \text{ use invNorm}$$

Manipulations:

- If $\mu = +x$, the graph shifts x units to the right
- If $\mu = -x$, the graph shifts x units to the left
- If $\sigma = +x$, the graph expands x times horizontally
- If $\sigma = -x$, the graph shrinks x times horizontally

Theorem 5.1. Standardization Theorem: If X has a normal distribution with parameters μ and σ, then

$$z = \frac{x - \mu}{\sigma}$$

has a standard distribution.

Example 5.4. Suppose X has a normal distribution with $\mu = 2$ and $\sigma = 4$. Compute $P[0 \leq x \leq 5]$.
Let $z = \frac{x-2}{4}$. If $x = 0$, $z = -\frac{1}{2}$. If $x = 5$, $z = \frac{3}{4}$. Then:

$$P[0 \leq x \leq 5] = P[-\frac{1}{2} \leq z \leq \frac{3}{4}]$$

6 Mathematical Expectation

Definition 6.1. Expected Value: If X is a discrete random variable whose probability density function is $f(x)$, the expected value of X is:

$$\text{E[X]} = \sum_x xf(x)$$

Example 6.1. An urn has four balls numbered 1, two balls numbered 2, one ball numbered 3, and three balls numbered 4. Select one ball and let X =value.

x	$f(x)$	$xf(x)$
1	0.4	0.4
2	0.2	0.4
3	0.1	0.4
4	0.3	1.2

$$\begin{aligned} \text{Average } &= \frac{(4)1+(2)2+(1)3+(3)4}{10} \\ &= \frac{(4)1}{10}+\frac{(2)2}{10}+\frac{(1)3}{10}+\frac{(3)4}{10} \\ &= 1(\frac{4}{10})+2(\frac{2}{10})+3(\frac{1}{10})+4(\frac{3}{10}) \\ &= \sum_{x=1}^{4} xf(x) = 2.3 \end{aligned}$$

Example 6.2. An urn contains 4 red balls and 1 blue ball. Two balls are randomly selected with replacement. Let X =the number of red balls obtained. Compute E[X].
This is a binomial distribution with $n = 2$ and $p = 0.8$.
Thus $f(x) = \binom{2}{x}(0.8)^x(0.2)^{2-x}$.

x	$f(x)$	$xf(x)$
0	0.4	0
1	0.32	0.32
2	0.64	1.28

$$\text{E[X]} = 0 + 0.32 + 1.28 = 1.60$$

Example 6.3. Two dices are rolled. Let X = sum. Compute E[X].

x	$f(x)$	$xf(x)$
2	$\frac{1}{36}$	$\frac{2}{36}$
3	$\frac{2}{36}$	$\frac{6}{36}$
4	$\frac{3}{36}$	$\frac{12}{36}$
5	$\frac{4}{36}$	$\frac{20}{36}$
6	$\frac{5}{36}$	$\frac{30}{36}$
7	$\frac{6}{36}$	$\frac{42}{36}$
8	$\frac{5}{36}$	$\frac{40}{36}$
9	$\frac{4}{36}$	$\frac{36}{36}$
10	$\frac{3}{36}$	$\frac{30}{36}$
11	$\frac{2}{36}$	$\frac{22}{36}$
12	$\frac{1}{36}$	$\frac{12}{36}$

$$\text{E[X]} = \frac{252}{36} = 7$$

Example 6.4. Two dices are rolled. Let $X =$ the absolute value of their difference. Compute E[X].

x	$f(x)$	$xf(x)$
0	$\frac{6}{36}$	$\frac{0}{36}$
1	$\frac{10}{36}$	$\frac{10}{36}$
2	$\frac{8}{36}$	$\frac{16}{36}$
3	$\frac{6}{36}$	$\frac{18}{36}$
4	$\frac{4}{36}$	$\frac{16}{36}$
5	$\frac{2}{36}$	$\frac{10}{36}$

$$\text{E[X]} = \frac{70}{36} = 1.944$$

Definition 6.2. If X is a continuous random variable with probability density function $f(x)$,

$$\text{E[X]} = \int_{-\infty}^{\infty} xf(x)dx$$

Example 6.5. Let X be a continuous random variable with probability density function $f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$. Compute E[X].

$$\text{E[X]} = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 2x^2dx = \frac{2}{3}x^3\Big|_0^1 = \frac{2}{3}$$

Example 6.6. Let X be a continuous random variable with probability density function $f(x) = \begin{cases} \frac{3}{4}(2x - x^2) & \text{if } 0 \leq x \leq 2 \\ 0 & \text{elsewhere} \end{cases}$. Compute E[X].

$$\text{E[X]} = \int_{-\infty}^{\infty} xf(x)dx = \int_0^2 \frac{3}{4}(2x - x^2)dx = \frac{3}{4}[\frac{2}{3}x^3 - \frac{x^4}{4}]\Big|_0^2 = \frac{3}{4}[\frac{16}{3} - \frac{16}{4}] = 4 - 3 = 1$$

Example 6.7. Suppose X has an exponential distribution with parameter $\theta = 3$. Compute E[X].

$$\begin{aligned} \text{E[X]} &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_0^{\infty} \frac{1}{3}xe^{-\frac{x}{3}}dx \\ &= \lim_{t \to infty} \int_0^t \frac{1}{3}xe^{-\frac{x}{3}}dx \\ &= \lim_{t \to \infty} [-xe^{-\frac{x}{3}} - 3e^{-\frac{x}{3}}]\Big|_0^t \\ &= \lim_{t \to \infty} [-te^{-\frac{t}{3}} - 3e^{-\frac{t}{3}} - 0 + 3] \\ &= 0 - 0 - 0 + 3 = 3 \end{aligned}$$

Theorem 6.1. Strong Law of Large Numbers: Let E be any experiment whose sample space if S. Let X be any random variable defined on S with $\text{E[X]} = \mu$. If the experiment is repeated under identical conditions and x_i denotes the value of x on the i^{th} repetition:

$$P[\lim_{n \to \infty} \frac{x_1 + x_2 + \cdots + x_n}{n} = \mu] = 1$$

In other words, as $x \to \infty$, $\frac{x_1+x_2+\cdots+x_n}{n} \to \mu$ with probability of 1. If n is large, $\frac{x_1+x_2+\cdots+x_n}{n} \approx \mu$ with large probabilities

Example 6.8. Toss a coin. $X_1 = 5, X_2 = 2.X_3 = 1$

$$\frac{x_1 + x_2 + x_3}{n} = \frac{5 + 2 + 1}{3} = 2.667$$

$$\text{E[X]} = 3.5$$

Definition 6.3. The game is fair if $\text{E[X]} = 0$.

Example 6.9. Pay 7 dollars to play a game. Roll 2 dice. You get an amount of money in dollars equal to the sum of the dice. Let X = your winnings.

x	$f(x)$	$xf(x)$
-5	$\frac{1}{36}$	$-\frac{5}{36}$
-4	$\frac{2}{36}$	$-\frac{8}{36}$
-3	$\frac{3}{36}$	$-\frac{9}{36}$
-2	$\frac{4}{36}$	$-\frac{8}{36}$
-1	$\frac{5}{36}$	$-\frac{5}{36}$
0	$\frac{6}{36}$	0
1	$\frac{5}{36}$	$\frac{5}{36}$
2	$\frac{4}{36}$	$\frac{8}{36}$
3	$\frac{3}{36}$	$\frac{9}{36}$
4	$\frac{2}{36}$	$\frac{8}{36}$
5	$\frac{1}{36}$	$\frac{5}{36}$

E[X] = 0 Game is fair.

Example 6.10. Roulette: 18 red slots, 18 blue slots, 2 green slots. Bet dollars on a color (red or blue). If your color shows up, you get back 2 dollars. Otherwise you get 0 dollars. Let X = winnings.

x	$f(x)$	$xf(x)$
-1	$\frac{20}{38}$	$-\frac{20}{38}$
1	$\frac{18}{38}$	$\frac{18}{38}$

E[X] $= -\frac{2}{38} \approx -0.0526$ This game is unfair.

Let X be a discrete random variable with probability density function $f(x)$. Then

$$E[g(x)] = \sum_x g(x)f(x)$$

If $g(x) = x$, $E(g(x)) =$ E[X].

Example 6.11. Toss 2 dice, one red and one green. Let X be the difference in the order red green. Compute $E(X^2)$.

x	x^2	$f(x)$	$x^2 f(x)$
-5	25	$\frac{1}{36}$	$\frac{25}{36}$
-4	16	$\frac{2}{36}$	$\frac{32}{36}$
-3	9	$\frac{3}{36}$	$\frac{27}{36}$
-2	4	$\frac{4}{36}$	$\frac{16}{36}$
-1	1	$\frac{5}{36}$	$\frac{5}{36}$
0	0	$\frac{6}{36}$	0
1	1	$\frac{5}{36}$	$\frac{5}{36}$
2	4	$\frac{4}{36}$	$\frac{16}{36}$
3	9	$\frac{3}{36}$	$\frac{27}{36}$
4	16	$\frac{2}{36}$	$\frac{32}{36}$
5	25	$\frac{1}{36}$	$\frac{25}{36}$

$\mathrm{E[X]} = 0$, $\mathrm{E[X^2]} = \frac{210}{36} \approx 5.833$

Note: Clearly

$$\mathrm{E[X^2]} \neq (\mathrm{E[X]})^2$$

If X is a continuous random variable with probability density function $f(x)$

$$\mathrm{E[g(x)]} = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Example 6.12. Let X have the probability density function:

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find $\mathrm{E[X^2]}$.

$$\mathrm{E[X^2]} = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 x^2 2x dx = \int_0^1 2x^3 dx = 2\frac{x^4}{4}\Big|_0^1 = \frac{1}{2}$$

Example 6.13. The radius of a circular disk produced by a certain machine is uniformly distributed between 1 inch and 1.02 inches. Compute its expected area. Let X = radius.

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} = \frac{1}{1.02-1} = 50 & \text{if } 1 \leq x \leq 1.02 \\ 0 & \text{elsewhere} \end{cases}$$

$$\begin{aligned} \text{E[A]} = \text{E}[\pi \text{X}^2] &= \int_{-\infty}^{\infty} \pi x^2 f(x) dx \\ &= \int_{1}^{1} .02\pi x^2 50 dx \\ &= 50\pi \frac{x^3}{3}\Big|_1^1 .02 = \frac{50\pi}{3}(1.02^3 - 1.0^3) \approx 1.02\pi \approx 32.05 \text{ in.}^2 \end{aligned}$$

Theorem 6.2.

$$\text{E[c]} = c$$

Proof. Discrete Case: $\text{E[c]} = \sum_x cf(x) = c\sum_x f(x) = c \times 1 = c$
Continuous Case: $\text{E[c]} = \int_{-\infty}^{\infty} cf(x)dx = c\int_{-\infty}^{\infty} f(x)dx = c \times 1 = c$ □

Theorem 6.3.

$$\text{E[aX]} = a\text{E[X]}$$

$$\text{E[g(x)]} = \sum_x g(x)f(x)$$

Proof. Discrete Case: $\text{E[aX]} = \sum_x axf(x) = a\sum_x xf(x) = a\text{E[X]}$ □

Theorem 6.4.

$$\text{E}[\text{aX}^2 + \text{bX} + \text{c}] = a\text{E}[\text{X}^2] + b\text{E[X]} + c$$

Proof. Discrete Case: $\text{E}[\text{aX}^2 + \text{bX} + \text{c}] = \sum_x (ax^2 + bx + c)f(x) = \sum_x (ax^2 f(x) + bxf(x) + cf(x)) = \sum_x ax^2 f(x) + \sum_x bxf(x) + \sum_x cf(x) = a\sum_x x^2 f(x) + b\sum_x xf(x) + c\sum_x f(x) = a\text{E}[\text{X}^2] + b\text{E[X]} + c$ □

This theorem extends to polynomials of higher degrees.

Definition 6.4. Variance of a Random Variable: Let X be a random variable where expected value $\text{E[X]} = \mu$.

$$\text{Var(X)} = \text{E}[(\text{x} - \mu)^2]$$

Example 6.14. An urn contains five balls numbered: 1, 3, 5, 7, 9. Select one ball. Let X =

its value. Find the variance of X. E[X] $= 5 = \mu$

x	$x-\mu$	$(x-\mu)^2$	$f(x)$	$(x-\mu)^2 f(x)$
1	-4	16	0.2	3.2
3	-2	4	0.2	0.8
5	0	0	0.2	0
7	2	4	0.2	0.8
9	4	16	0.2	3.2

$$\text{Var[X]} = \text{E}[(\text{x}-\mu)^2] = 8$$

Example 6.15. An urn contains five balls numbered: 3, 4, 5, 6, 7. Select one ball. Let $X =$ its value. Find the variance of X. E[X] $= 5 = \mu$

x	$x-\mu$	$(x-\mu)^2$	$f(x)$	$(x-\mu)^2 f(x)$
3	-2	4	0.2	0.8
4	-1	1	0.2	0.2
5	0	0	0.2	0
6	1	1	0.2	0.2
7	2	4	0.2	0.8

$$\text{Var[X]} = \text{E}[(\text{x}-\mu)^2] = 2$$

Example 6.16. An urn contains five balls numbered: 5, 5, 5, 5, 5. Select one ball. Let $X =$ its value. Find the variance of X. E[X] $= 5 = \mu$

x	$x-\mu$	$(x-\mu)^2$	$f(x)$	$(x-\mu)^2 f(x)$
5	0	0	1	0

$$\text{Var[X]} = \text{E}[(\text{x}-\mu)^2] = 0$$

One way to describe variance is average squared deviation.
It is always true that V[X] ≥ 0.

Definition 6.5. Standard Deviation of X:

$$\sigma = \sqrt{\text{V[X]}}$$

or

$$\sigma^2 = \text{Var[X]}$$

Example 6.17. Let X be a random variable whose probability density function is: $f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$. Find the variance of X.

$$\text{E[X]} = \int_0^1 xf(x)dx = \int_0^1 2x^2 dx = \frac{2}{3}x^3\Big|_0^1 = \frac{2}{3} = \mu$$

$$\begin{aligned}
\text{E[(x} - \mu)^2] &= \int_0^1 (x - \frac{2}{3})^2 f(x)dx \\
&= \int_0^1 (x - \frac{2}{3})^2 (2x)dx \\
&= \int_0^1 (x^2 - \frac{4}{3}x + \frac{4}{9})dx \\
&= \int_0^1 (2x^3 - \frac{8}{3}x^2 + \frac{8}{9}x)dx \\
&= [\frac{1}{2}x^4 - \frac{8}{9}x^3 + \frac{4}{9}x^2]\Big|_0^1 \\
&= \frac{1}{2} - \frac{8}{9} + \frac{4}{9} = \frac{1}{18} = \text{Var[x]}
\end{aligned}$$

Theorem 6.5.

$$\text{Var[X]} = \text{E[X}^2] - \text{E}^2\text{[X]}$$

Proof.

$$\begin{aligned}
\text{Var[X]} &= \text{E[(x} - \mu)^2] \\
&= \text{E[X}^2 - 2\mu\text{X} + \mu^2] \\
&= \text{E[X]} - \text{E[}2\mu\text{X]} + \text{E[}\mu^2] \\
&= \text{E[X}^2] - 2\mu\text{E[X]} + \mu^2 \\
&= \text{E[X}^2] - 2\mu^2 + \mu^2 = \text{E[X}^2] - \text{E}^2\text{[X]}
\end{aligned}$$

□

Example 6.18. Find the variance of random variable X whose probability density function is: $f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$.

$$\text{E[X}^2] = \int_0^1 x2xdx = \int_0^1 2x^3 dx = \frac{1}{2}x^4\Big|_0^1 = \frac{1}{4}$$

$$\text{Var[X]} = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{2} - \frac{4}{9} = \frac{9}{18} - \frac{8}{18} = \frac{1}{18}$$

Example 6.19. An urn contains five balls numbered: 1, 3, 5, 7, 9. Select one ball. Let $X =$ its value. Find the variance of X. E[X] = 5

x	$f(x)$	$xf(x)$	$x^2f(x)$
1	0.2	0.2	0.2
3	0.2	0.6	1.8
5	0.2	1.0	5.0
7	0.2	1.4	9.8
9	0.2	1.8	16.2

$$\text{Var[X]} = \text{E[X}^2] - \text{E}^2\text{[X]} = 33 - 5^2 = 8$$

Theorem 6.6.

$$\text{Var[aX]} = a^2\text{Var[X]}$$

Proof. Let $Y = aX$. Then Var[Y] = Var[aX] = aE[X]. Thus, let $\mu_Y = a\mu_X$.

$$\begin{aligned} \text{Var[Y]} &= \text{E[(Y} - \mu_\text{Y})^]} \\ &= \text{E[(aX} - \text{a}\mu_\text{X})^2] \\ &= \text{E[a}^2(\text{x} - \mu_\text{Y})^2] \\ &= a^2\text{E[(x} - \mu)^2] \\ &= a^2\text{Var[X]} \end{aligned}$$

□

Theorem 6.7.

$$\text{Var[X + b]} = \text{Var[X]}$$

Proof. Let $Y = X + b$. □

Theorem 6.8.

$$\text{Var[c]} = 0$$

Proof. Let $Y = c$. □

Example 6.20. Let X have a binomial distribution with parameters $n = 3$ and $p = 0.4$. Compute the expected value and variance of X.

x	$f(x)$	$xf(x)$	$x^2f(x)$
0	$\binom{3}{0}(0.4)^0(0.6)^3 = 0.216$	0.0	0.0
1	$\binom{3}{1}(0.4)^1(0.6)^2 = 0.432$	0.432	0.432
2	$\binom{3}{2}(0.4)^2(0.6)^1 = 0.288$	0.576	1.152
3	$\binom{3}{3}(0.4)^3(0.6)^0 = 0.064$	0.192	0.576

$$\text{E[X]} = 1.200$$
$$\text{E[X}^2] = 2.16$$
$$\text{Var[X]} = \text{E[X}^2] - \text{E}^2\text{[X]} = 2.16 - 1.200^2 = 2.16 - 1.44 = 0.72$$

Theorem 6.9. If X has a binomial distribution with parameters n and p,

$$\text{E[X]} = np$$

and

$$\text{Var[x]} = npq$$

Proof.

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

$$\begin{aligned}
\text{E[X]} &= \sum_{x=0}^{n} x f(x) \\
&= \sum_{x=0}^{n} x \binom{n}{x} p^x q^{1-x} \\
&= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x q^{1-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x q^{1-x} \\
&= \sum_{x=1}^{n} \frac{n(n-1)!}{(x-1)!(n-x)!} p p^{x-1} q^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} \\
&\text{Let y = x - 1 and m = n - 1} \\
&= np \underbrace{\sum_{y=0}^{m} \frac{m!}{y!(m-y)!} p^y q^{m-y}}_{(p+q)^m = 1} \\
&= np \cdot 1 \\
&= np
\end{aligned}$$

□

Proof.

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

$$\begin{aligned}
\text{Var[X]} &= \text{E[X}^2] - \text{E}^2\text{[X]} \\
&= \text{E[X(X} - 1) + \text{X]} - \text{E}^2\text{[X]} \\
&= \text{E[X(X} - 1)] + \text{E[X]} - \text{E}^2\text{[X]} \\
\text{E[X(X} - 1)] &= \sum_{x=0}^{n} x(x-1)f(x) \\
&= \sum_{x=0}^{n} x(x-1)\binom{n}{x}p^x q^{n-x} \\
&= \sum_{x=2}^{n} x(x-1)\binom{n}{x}p^x q^{n-x} \\
&= \sum_{x=2}^{n} x(x-1)\frac{n!}{x!(n-x)!}p^x q^{n-x} \\
&= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!}p^x q^{n-x} \\
&= \sum_{x=2}^{n} \frac{n(n-1)(n-2)!}{(x-2)!(n-x)!}p^2 p^{x-2} q^{n-x} \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!}p^{x-2} q^{n-x} \\
&\text{Let m = n - 2 and y = x - 2} \\
&= n(n-1)p \underbrace{\sum_{y=0}^{m} \frac{m!}{y!(m-y)!}p^y q^{m-y}}_{(p+q)^m = 1} \\
&= n(n-1)p^2 \\
\text{Var[X]} &= \text{E[X(X} - 1)] + \text{E[X]} - \text{E}^2\text{[X]} \\
&= n(n-1)p^2 + np - n^2p^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np - np^2 \\
&= np(1-p) \\
&= npq
\end{aligned}$$

□

Example 6.21. An urn has 7 red balls and 8 blue balls. Select 5 balls without replacement. Let X = the number of red balls obtained. Compute E[X] and Var[X]. (This is a binomial distribution.)

$$\text{E[X]} = np = 5 \cdot \frac{7}{15} = \frac{7}{3}$$

$$\text{Var[X]} = npq = 5 \cdot \frac{7}{15} \cdot \frac{8}{15} = \frac{56}{45}$$

Theorem 6.10. If X is uniformly distributed between α and β, then

$$\text{E[X]} = \frac{\alpha + \beta}{2}$$

and

$$\text{Var[X]} = \frac{(\beta - \alpha)^2}{12}$$

Proof.

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha \leq x \leq \beta \\ 0 & \text{elsewhere} \end{cases}$$

$$\begin{aligned} \text{E[X]} &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{\alpha}^{\beta} x \frac{1}{\beta - \alpha} dx \\ &= \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} x dx \\ &= \frac{1}{\beta - \alpha} \frac{1}{2} x \Big|_{\alpha}^{\beta} \\ &= \frac{1}{2(\beta - \alpha)} (\beta^2 - \alpha^2) \\ &= \frac{1}{2(\beta - \alpha)} (\beta + \alpha)(\beta - \alpha) \\ &= \frac{1}{2} (\beta + \alpha) \end{aligned}$$

□

Proof.

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha \leq x \leq \beta \\ 0 & \text{elsewhere} \end{cases}$$

$$\begin{aligned}
\text{Var[X]} &= \text{E[X}^2] - \text{E}^2\text{[X]} \\
\text{E[X}^2] &= \int_{-\infty}^{\infty} x^2 f(x)dx \\
&= \int_{\alpha}^{\beta} x^2 \frac{1}{\beta - \alpha} dx \\
&= \frac{1}{\beta - \alpha} \frac{x^3}{3}\Big|_{\alpha}^{\beta} \\
&= \frac{1}{3(\beta - \alpha)}(\beta^3 - \alpha^3) \\
&= \frac{1}{3(\beta - \alpha)}(\beta - \alpha)(\beta^2 + \alpha\beta + \alpha^2) \\
&= \frac{1}{3}(\beta^2 + \alpha\beta + \alpha^2) \\
\text{Var[X]} &= \text{E[X}^2] - \text{E}^2\text{[X]} \\
&= \frac{1}{3}(\beta^2 + \alpha\beta + \alpha^2) - \frac{\beta^2 - 2\alpha\beta + \beta^2}{4} \\
&= \frac{1}{12}(4\beta^2 + 4\alpha\beta + 4\alpha^3 - 3\beta^2 - 6\alpha\beta - 3\alpha^2) \\
&= \frac{1}{12}(\beta^2 - 2\alpha\beta + \alpha^2) \\
&= \frac{(\beta - \alpha)^2}{12}
\end{aligned}$$

□

Example 6.22. Compute the expected number of aces in a 5 cards poker hand. (This is a hypergeometric distribution.)

$$\text{E[X]} = \frac{nN_1}{N_1 + N_2} = \frac{5 \cdot 4}{4 + 48} = 0.38463$$

Example 6.23. Let X have a normal distribution with $\mu = 3$. Find σ if $P[x > 1] = 0.89435$.

$$P[X < 1] = 1 - P[X > 1] = 1 - 0.89435 = 0.010565$$

Let $z = \frac{x-\mu}{\sigma}$ where z has a standard normal distribution. Then:

$$z = \frac{x - 3}{\sigma} = \text{invNorm}(0.10565, 0, 1) = -1.245$$

This means

$$-1.245 = \frac{1 - 3}{\sigma} \to \sigma = \frac{-1}{-1.245} = 1.6$$

7 Multivariate Distribution

Definition 7.1. Bivariate Distribution: a probability distribution of two random variables either X, Y or X_1, X_2

Definition 7.2. Joint Probability Function (discrete case):

$$f(x,y) = P[X = x, Y = y]$$

Example 7.1. Two dices are tossed. Let X = the sum of 2 dices and Y = the absolute value of its difference. Construct the joint probability function.

$f(x,y)$	2	3	4	5	6	7	8	9	10	11	12
0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$
1	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0
2	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0
3	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0
4	0	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0	0
5	0	0	0	0	0	$\frac{2}{36}$	0	0	0	0	0

Example 7.2. Toss four coins. Let X = the number of heads and Y = the number of heads - the number of tails. Construct the joint probability function.

$f(x,y)$	0	1	2	3	4
-4	$\frac{1}{16}$	0	0	0	0
-3	0	0	0	0	0
-2	0	$\frac{4}{16}$	0	0	0
-1	0	0	0	0	0
0	0	0	$\frac{6}{16}$	0	0
1	0	0	0	0	0
2	0	0	0	$\frac{4}{16}$	0
3	0	0	0	0	0
4	0	0	0	0	$\frac{1}{16}$

Definition 7.3. Cumulative Joint Probability Function:

$$F(x,y) = P[X \leq x, Y \leq y]$$

Definition 7.4. Marginal Probability Function:

$$f_X(x) = P[X = x] = \sum_Y f(x,y)$$

$$f_Y(y) = P[Y = y] = \sum_X f(x,y)$$

Example 7.3. Suppose a certain experiment has the following joint probability distribution. Construct its marginal probability function.

$f(x,y)$	1	2	3	4	$f_Y(y)$
1	0.1	0.05	0.07	0.02	0.24
2	0.08	0.12	0.03	0.02	0.25
3	0.15	0.2	0.05	0.11	0.51
$f_X(x)$	0.33	0.37	0.15	0.15	1.00

$$F(3,2) = P[X \leq 3, Y \leq 2] = 0.45$$
$$F(3.4, 2.1) = P[X \leq 3.4, Y \leq 2.1] = 0.45$$
$$f_X(2) = 0.05 + 0.12 + 0.2 = 0.37$$
$$f_Y(2) = 0.08 + 0.12 + 0.03 + 0.02 = 0.25$$

Note: X and Y are independent if

$$f(x,y) = f_X(x)f_Y(y) \text{ for all } X, Y$$
$$P[X = x \text{ and } Y = y] = P[X = x]P[Y = y]$$
$$P(AB) = P(A)P(B)$$

Example 7.4. Roll a die and toss 2 coins. Let X = the value on the die and Y = the number of heads obtained. Construct the joint probability function.

$f(x,y)$	1	2	3	4	5	6	$f_Y(y)$
0	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{4}$
1	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{1}{2}$
2	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{4}$
$f_X(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

$$f(x,y) = f_X(x)f_Y(y) \text{ for all } x, y \to X, Y \text{ are independent.}$$

Properties of Single Summations:

1.

Theorem 7.1.
$$\sum_{x=1}^{n} cf(x) = c\sum_{x=1}^{n} f(x)$$

2.

Theorem 7.2.
$$\sum_{x=1}^{n}(f(x)+g(x)) = \sum_{x=1}^{n} f(x) + \sum_{x=1}^{n} g(x)$$

Definition 7.5. Double Summations:
$$\sum_{x=1}^{n}\sum_{y=1}^{m} f(x,y) = \sum_{x=1}^{n}(\sum_{y=1}^{m} f(x,y))$$

Properties of Double Summations:

1.

Theorem 7.3.
$$\sum_{x}\sum_{y} cf(x,y) = c\sum_{x}\sum_{y} f(x,y)$$

2.

Theorem 7.4.
$$\sum_{x}\sum_{y}[f(x,y)+g(x,y)] = \sum_{x}\sum_{y} f(x,y) + \sum_{x}\sum_{y} g(x,y)$$

Example 7.5. Compute $\sum_{x=1}^{3}\sum_{y=1}^{4} xy^2$.
$$\sum_{x=1}^{3}\sum_{y=1}^{4} xy^2 = \sum_{x=1}^{3}(x+4x+9x+16x) = \sum_{x=1}^{3} 30x = 30\sum_{x=1}^{3} x = 30\cdot 6 = 180$$

Example 7.6. Compute $\sum_{y=1}^{4}\sum_{x=1}^{3} xy^2$.
$$\sum_{y=1}^{4}\sum_{x=1}^{3} xy^2 = \sum_{y=1}^{4}(y^2+2y^2+3y^2) = \sum_{y=1}^{4} 6y^2 = 6(1+4+9+16) = 6\cdot 30 = 180$$

Theorem 7.5. If $h(x,y) = f(x)g(y)$, then
$$\sum_{x}\sum_{y} h(x,y) = (\sum_{x} f(x))(\sum_{y} g(y))$$

Example 7.7. Compute $\sum_{x=1}^4 \sum_{y=1}^5 (x+y^2)$.

$$\begin{aligned}\sum_{x=1}^4 \sum_{y=1}^5 (x+y^2) &= \sum_{x=1}^4 \sum_{y=1}^5 x + \sum_{x=1}^4 \sum_{y=1}^5 y^2 \\ &= \sum_{x=1}^4 x \sum_{y=1}^5 1 + \sum_{y=1}^5 y^2 \sum_{x=1}^4 1 \\ &= 5\sum_{x=1}^4 x + 4\sum_{y=1}^5 y^2 \\ &= (5\cdot 10) + (4 \cdot 55) = 50 + 220 = 270\end{aligned}$$

Definition 7.6. Double Integrals:

$$\int_a^b \int_c^d f(x,y)dydx = \int_a^b [\int_c^d f(x,y)dy]dx$$

Example 7.8. Compute $\int_0^1 \int_0^2 xydydx$.

$$\int_0^1 \int_0^2 xydydx = \int_0^1 \frac{xy^2}{2}\Big|_{y=0}^{y=2} dx = \int_0^1 2xdx = x^2\Big|_0^1 = 1$$

Example 7.9. Compute $\int_0^1 \int_0^1 xy^3dydx$

$$\int_0^1 \int_0^1 xy^3dydx = \int_0^1 x\frac{y^4}{4}\Big|_{y=0}^{y=1} dx = \int_0^1 \frac{1}{4}xdx = \frac{x^2}{8}\Big|_0^1 = \frac{1}{24}$$

Example 7.10. Compute $\int_0^1 \int_0^x xy^3dydx$

$$\int_0^1 \int_0^x xy^3dydx = \int_0^1 x\frac{y^4}{4}\Big|_{y=0}^{y=x} dx = \int_0^1 \frac{1}{4}x^5dx = \frac{1}{24}x^6\Big|_0^1 = \frac{1}{24}$$

$f(x,y)$ is called a joint probability density function if:

- $f(x,y) \geq 0$ for all x, y
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dydx = 1$

If $f(x,y)$ is the joint probability density function of X and Y,

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

and

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

If X and Y have joint probability density function $f(x,y)$, we say that X and Y are independent if

$$f(x,y) = f_X(x)f_Y(y)$$

Example 7.11. Suppose X and Y have joint probability density function as follows: $f(x,y) = \begin{cases} \frac{3}{16}xy^2 & \text{if } 0 \leq x \leq 2, 0 \leq y \leq 2 \\ 0 & \text{elsewhere} \end{cases}$

1. Verify that $f(x,y)$ is a joint probability density function.

   - $f(x,y) \geq 0$ for all x, y
   - $$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dydx &= \int_0^2\int_0^2 \frac{3}{16}xy^2dydx \\ &= \int_0^2 \frac{x}{16}y^3\Big|_{y=0}^{y=2}dx \\ &= \int_0^2 \frac{1}{2}xdx \\ &= \frac{x^2}{4}\Big|_0^2 \\ &= 1 \end{aligned}$$

2. Compute $P[0 \leq X \leq 1, 0 \leq Y \leq 1]$.

$$\begin{aligned} P[0 \leq X \leq 1, 0 \leq Y \leq 1] &= \int_0^1\int_0^1 \frac{3}{16}xy^2dydx \\ &= \int_0^1 \frac{1}{16}xy^3\Big|_{y=0}^{y=1}dx \\ &= \int_0^1 \frac{1}{16}xdx \\ &= \frac{1}{32}x^2\Big|_0^1 \\ &= \frac{1}{32} \end{aligned}$$

3. Compute $P[Y \leq X]$.

$$\begin{aligned} P[Y \leq X] &= \int_0^2\int_0^x \frac{3}{16}xy^2dydx \\ &= \int_0^2 \frac{1}{16}xy^3\Big|_{y=0}^{y=x}dx \\ &= \int_0^2 \frac{1}{16}x^4dx \\ &= \frac{1}{80}x^5\Big|_0^2 \\ &= \frac{32}{80} \end{aligned}$$

4. Compute $f_X(x)$ and $f_Y(y)$.

$$\begin{aligned} f_X(x) &= \int_0^2 xy^2 dy \\ &= \frac{1}{16}xy^3\Big|_{y=0}^{y=2} \\ &= \frac{1}{2}x \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_0^2 \frac{3}{16}xy^2 dx \\ &= \frac{3}{32}x^2y^2\Big|_{x=0}^{x=2} \\ &= \frac{3}{8}y^2 \end{aligned}$$

5. Show that X and Y are independent.

$$\frac{3}{16}xy^2 = (\frac{1}{2}x)(\frac{3}{8}y^2) = \frac{3}{16}xy^2$$

8 Functions of Several Random Variables; Central Limit Theorem

Example 8.1. Suppose X and Y have joint probability density function as shown in the following table:

$f(x,y)$	1	2	3
1	0.1	0.1	0.2
2	0.5	0.03	0.12
3	0.2	0.15	0.05

Find the probability function of $Z = X + Y$ and $W = XY$

Z	$f_Z(z)$
2	0.1
3	0.15
4	0.43
5	0.27
6	0.05

W	$f_W(w)$
1	0.1
2	0.15
3	0.4
4	0.03
6	0.27
9	0.05

Definition 8.1.

$$\mathrm{E[g(x, y)]} = \sum_x \sum_y g(x,y) f(x,y)$$

$$\mathrm{E[g(x, y)]} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dy dx$$

Theorem 8.1.

$$\mathrm{E[X + Y]} = \sum_x \sum_y (x+y) f(x,y)$$

Proof.

$$\begin{aligned} \mathrm{E[X + Y]} &= \sum_x \sum_y x f(x,y) + y f(x,y) \\ &= \sum_x \sum_x x f(x,y) + \sum_y \sum_y y f(x,y) \\ &= \mathrm{E[X] + E[Y]} \end{aligned}$$

□

Theorem 8.2.

$$\mathrm{E[X + Y] = E[X] + E[Y]}$$

$$\mathrm{E[X - Y] = E[X] - E[Y]}$$

It is not always the case that $\mathrm{E[XY] = E[X]E[Y]}$

Theorem 8.3. If X and Y are independent, then

$$\mathrm{E[XY] = E[X]E[Y]}$$

Proof.

$$\begin{aligned}
\mathrm{E[XY]} &= \sum_x \sum_y (xy) f(x,y) \\
&= \sum_x \sum_y (xy) f_X(x) f_Y(y) \\
&= \sum_x x f_X(x) + \sum_y y f_Y(y) \\
&= \mathrm{E[X]E[Y]}
\end{aligned}$$

□

Definition 8.2. Let X and Y be random variables with means μ_X and μ_Y respectively. The covariance of X with Y is

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

Theorem 8.4.

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

Proof.

$$\begin{aligned}
\mathrm{Cov[X,Y]} &= \mathrm{E[(X - \mu_X)(Y - \mu_Y)]} \\
&= \mathrm{E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y]} \\
&= \mathrm{E[XY] - E[\mu_X Y] - E[\mu_Y X] + E[\mu_X \mu_Y]} \\
&= \mathrm{E[XY]} - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y \\
&= \mathrm{E[XY]} - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\
&= \mathrm{E[XY]} - \mu_X \mu_Y \\
&= \mathrm{E[XY] - E[X]E[Y]}
\end{aligned}$$

□

Theorem 8.5. If X and Y are independent, then

$$\mathrm{Cov[X,Y]} = 0$$

Note: If $\mathrm{Cov[X,Y]} = 0$, then X and Y may or may not be independent. Thus, if $\mathrm{Cov[X,Y]} = 0$, we say that X and Y are uncorrelated.
In general,

$$-\infty < \mathrm{Cov[X,Y]} < \infty$$

Definition 8.3. Correlation Coefficient:

$$\rho(X,Y) = \frac{\mathrm{Cov[X,Y]}}{\mu_X \mu_Y}$$

Properties of the Correlation Coefficient:

1. $-1 \leq \rho(X,Y) \leq 1$

2. If $\rho(X,Y) = 1$, $Y = aX + b$ with $a > 0$
3. If $\rho(X,Y) = -1$, $Y = aX + b$ with $a < 0$

Theorem 8.6.

$$\text{Var[X + Y]} = \text{Var[X]} + \text{Var[Y]} + 2\text{Cov[X, Y]}$$
$$\text{Var[X − Y]} = \text{Var[X]} + \text{Var[Y]} - 2\text{Cov[X, Y]}$$

Proof. Let $Z = X + Y$

$$\text{E[Z]} = \text{E[X]} + \text{E[Y]}$$

$$\begin{aligned}\text{Var[Z]} &= \text{E}[\text{Z}^2] - \text{E}^2[\text{Z}] \\ &= \text{E}[\text{X} + \text{Y}]^2 - [\text{E[X]} + \text{E[Y]}]^2 \\ &= \text{E}[\text{X}^2 + 2\text{XY} + \text{Y}^2] - [\text{E}^2[\text{X}] + 2\text{E[X]E[Y]} + \text{E}^2[\text{Y}]] \\ &= \text{E}[\text{X}^2] + 2\text{E[XY]} + \text{E}[\text{Y}^2] - \text{E}^2[\text{X}] - 2\text{E[X]E[Y]} \\ &= \text{E}[\text{X}^2] - \text{E}^2[\text{X}] + \text{E}[\text{Y}^2] - \text{E}^2[\text{Y}] + 2[\text{E[XY]} - \text{E[X]E[Y]}] \\ &= \text{Var[X]} + \text{Var[Y]} + 2\text{Cov[X, Y]}\end{aligned}$$

□

If X and Y are independent,

$$\text{Var[X + Y]} = \text{Var[X]} + \text{Var[Y]}$$

$$\text{Var[X − Y]} = \text{Var[X]} + \text{Var[Y]}$$

Example 8.2. Suppose X and Y have joint probability function as follows:

$f(x,y)$	1	2	3
1	0.1	0.1	0.2
2	0.5	0.03	0.12
3	0.2	0.15	0.05

$$\begin{aligned}\text{E[X + Y]} &= \sum_{x=1}^{3}\sum_{y=1}^{3}(x+y)f(x,y) \\ &= (1+1)(0.1) + (1+2)(0.1) + (1+3)(0.2) \\ &+ (2+1)(0.05) + (2+2)(0.03) + (2+3)(0.12) \\ &+ (3+1)(0.2) + (3+2)(0.03) + (3+3)(0.05) \\ &= 4.02\end{aligned}$$

$$\begin{aligned}
\text{E[X]} &= \sum_{x=1}^{3} xf(x,y) \\
&= (1)(0.1) + (2)(0.1) + (3)(0.2) \\
&+ (1)(0.05) + (2)(0.03) + (3)(0.12) \\
&+ (1)(0.2) + (2)(0.15) + (3)(0.05) \\
&= 2.02
\end{aligned}$$

$$\text{E[X]} = \sum_{x=1}^{3} xf_X(x) = (1)(0.35) + (2)(0.28) + (3)(0.37) = 2.02$$

$$\text{E[Y]} = 2.00$$

$$\begin{aligned}
\text{E[XY]} &= (1\cdot 1)(0.1) + (1\cdot 2)(0.0.1) + (1\cdot 3)(0.2) \\
&+ (2\cdot 1)(0.05) + (2\cdot 2)(0.03) + (2\cdot 3)(0.12) \\
&+ (3\cdot 1)(0.2) + (3\cdot 2)(0.15) + (3\cdot 3)(0.05) \\
&= 3.79
\end{aligned}$$

$$\text{Cov[X, Y]} = \text{E[XY]} - \text{E[X]E[Y]} = 3.79 - (2.02)(2) = -0.25$$

Theorem 8.7. If $X_1, X_2, \dots, X_n$ are independent,

$$\text{Var}[\text{X}_1, \text{X}_2, \dots, \text{X}_\text{n}] = \text{Var}[\text{X}_1] + \text{Var}[\text{X}_2] + \cdots + \text{Var}[\text{X}_\text{n}]$$

Note: if 1 or more of the plus signs on the left hand side are minuses, the result remains the same.

Example 8.3. Let $f(x,y) = \begin{cases} \frac{3}{16}xy^2 & \text{if } 0 \le x, y \le 2 \\ 0 & \text{elsewhere} \end{cases}$. Compute the following:

- $\text{E[X]} = \int_0^2 \int_0^2 x\frac{3}{16}xy^2 dydx = \int_0^2 \int_0^2 \frac{3}{16}x^2y^2 dydx = \frac{4}{3}$
- $\text{E[Y]} = \int_0^2 \int_0^2 y\frac{3}{16}xy^2 dydx = \int_0^2 \int_0^2 \frac{3}{16}xy^3 dydx = \frac{3}{2}$
- $\text{E[X}^2\text{]} = \int_0^2 \int_0^2 x^2\frac{3}{16}xy^2 dydx = \int_0^2 \int_0^2 \frac{3}{16}x^3y^2 dydx = 2$
- $\text{E[Y}^2\text{]} = \int_0^2 \int_0^2 y^2\frac{3}{16}xy^2 dydx = \int_0^1 \int_0^2 \frac{3}{16}xy^4 dydx = \frac{12}{5}$
- $\text{E[XY]} = \int_0^2 \int_0^2 xy\frac{3}{16}xy^2 dydx = \int_0^1 \int_0^1 \frac{3}{16}x^2y^3 dydx = 2$
- $\text{Var[X]} = \text{E[X}^2\text{]} - \text{E}^2\text{[X]} = 2 - (\frac{4}{3})^2 = \frac{2}{9}$
- $\text{Var[Y]} = \text{E[Y}^2\text{]} - \text{E}^2\text{[Y]} = \frac{12}{5} - (\frac{3}{2})^2 = \frac{3}{20}$
- $\text{E[X + Y]} = \text{E[X]} + \text{E[Y]} = \frac{4}{3} + \frac{3}{2} = \frac{17}{6}$
- $\text{E[X} - \text{Y]} = \text{E[X]} - \text{E[Y]} = \frac{4}{3} - \frac{3}{2} = -\frac{1}{6}$
- $\text{Cov[X, Y]} = \text{E[XY]} - \text{E[X]E[Y]} = 2 - (\frac{4}{3})(\frac{3}{2}) = 0$

- $\text{Var[X + Y]} = \text{Var[X]} + \text{Var[Y]} + 2\text{Cov[X, Y]} = \frac{2}{9} + \frac{3}{10} + 2(0) = \frac{67}{180}$
- $\text{Var[X - Y]} = \text{Var[X]} + \text{Var[Y]} - 2\text{Cov[X, Y]} = \frac{2}{9} + \frac{3}{10} - 2(0) = \frac{67}{180}$

Definition 8.4. Random Variables X and Y are identically distributed if they have the same probability distributions with identical parameters.

Example 8.4. Toss a single dice twice. Let X = the result of the first toss and Y = the result of the second toss.

x	$f(x)$	y	$f(y)$
1	$\frac{1}{6}$	1	$\frac{1}{6}$
2	$\frac{1}{6}$	2	$\frac{1}{6}$
3	$\frac{1}{6}$	3	$\frac{1}{6}$
4	$\frac{1}{6}$	4	$\frac{1}{6}$
5	$\frac{1}{6}$	5	$\frac{1}{6}$
6	$\frac{1}{6}$	6	$\frac{1}{6}$

With respect to X, this is a uniform distribution with parameter 6. With respect to Y, this is also a uniform distribution with parameter 6. Thus, X and Y are identically distributed. X and Y are also IID.

Definition 8.5. If X and Y are identically distributed and are independent, we say they are IID.

Theorem 8.8. If $X_1, X_2, \dots, X_n$ are IID, with mean μ and standard deviation σ:

1. If $S = X_1 + X_2 + \cdots + X_n$,
$$\text{E[S]} = n\mu$$
$$\text{Var[S]} = n\sigma^2$$
$$\sigma_S = \sigma\sqrt{n}$$
2. If $\bar{X} = \frac{X_1+X_2+\cdots+X_N}{n}$,
$$\text{E}[\bar{\text{X}}] = \mu$$
$$\text{Var}[\bar{\text{X}}] = \frac{\sigma^2}{n}$$
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Proof.

$$\bar{X} = \frac{1}{n}S$$

$$\mathrm{E}[\bar{\mathrm{X}}] = \mathrm{E}[\frac{1}{\mathrm{n}}\mathrm{S}] = \frac{1}{n}\mathrm{E}[\mathrm{S}] = \frac{1}{n}n\mu = \mu$$

$$\mathrm{Var}[\bar{\mathrm{X}}] = \mathrm{Var}[\frac{1}{\mathrm{n}}\mathrm{S}] = \frac{1}{n^2}\mathrm{Var}[\mathrm{S}] = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

□

Example 8.5. A pair of dice is tossed 100 times. a) Compute the expected value and standard deviation of the sum of the tosses.
If X represents the sum of the dices on the i^{th} toss, then $S = X_1 + X_2 + \cdots + X_{100}$. Furthermore, $X_1, X_2, \dots, X_{100}$ are IID.

x	$f(x)$
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$

$$\mathrm{E}[\mathrm{X}] = 7$$

$$\mathrm{Var}[\mathrm{X}] = \mathrm{E}[\mathrm{X}^2] - \mathrm{E}^2[\mathrm{X}] = 54.833 - 7^2 = 5.833$$

$$\mathrm{E}[\mathrm{S}] = 100\mathrm{E}[\mathrm{X}] = 100 \cdot 7 = 700$$

$$\mathrm{Var}[\mathrm{S}] = 100\mathrm{Var}[\mathrm{X}] = 100 \cdot 5.833 = 583.3$$

$$\sigma_S = \sigma_X\sqrt{100} = 10\sqrt{5.833} = 2.415$$

b) Compute $\mathrm{E}[\bar{\mathrm{x}}]$, $\mathrm{Var}[\bar{\mathrm{X}}]$ and $\sigma_{\bar{X}}$.

$$\mathrm{E}[\bar{\mathrm{X}}] = \mu = 7$$

$$\text{Var}[\bar{X}] = \frac{\text{Var}[X]}{n} = \frac{5.833}{100} = 0.05833$$
$$\sigma_{\bar{X}} = \frac{\sigma_x}{\sqrt{n}} = \frac{2.415}{\sqrt{100}} = 0.2415$$

Theorem 8.9. If $X_1, X_2, \ldots, X_n$ are independent random variables having Bernoulli distribution, with parameter p, then $S = X_1 + X_2 + \cdots + X_n$ has a binomial distribution with parameters n and p such that
$$\text{E[S]} = n\text{E[X]} = np$$
$$\text{Var[S]} = n\text{Var[X]} = npq$$

Proof.

x	$f(x)$
1	p
0	q

$S = X_1 + X_2 + \cdots + X_n =$ number of successes in n Bernoulli trials □

Theorem 8.10. If X and Y are independent random variables having Poisson distribution, with parameters λ_X and λ_Y, then $X + Y$ also has a Poisson distribution with parameter $\lambda_X + \lambda_Y$. This extends to any number of random variables having a Poisson distribution.

Example 8.6. The number of cars coming at a certain toll booth in a one minute time interval obeys a Poisson distribution with $\lambda = 3$. a) Compute the probability that no more than 2 cars will arrive between 6:00 and 6:01. Let $X =$ the number of cars.

$$P[X \leq 2] = \text{poissoncdf}(3, 2) = 0.42319$$

b) Compute the probability that no more than 160 cars will arrive between 6:00 and 7:00. Let $X_1 =$ the number of cars arriving in the first minute, $X_2 =$ the number of cars arriving in the second minute, $\ldots$, $X_{60} =$ the number of cars arriving in the sixtieth minute.

$$S = X_1 + X_2 + \cdots + X_{60}$$

By the previous theorem, S is a Poisson distribution with parameter

$$\lambda_S = 60\lambda = 60 \cdot 3 = 180$$

$$P[S \leq 160] = \text{poissoncdf}(180, 160) = 0.07101$$

Theorem 8.11. If X and Y are independent random variables having Poisson distributions with parameters λ_X and λ_Y, then $X+Y$ has a Poisson distribution with parameter $\lambda_X + \lambda_Y$.

Proof. Let $Z = X + Y$ such that $f_Z(z) = P[Z = z] = P[X + Y = z]$.
If $X = k$, then $Y = Z - k$ in order for $X + Y = Z$. Thus:

$$\begin{aligned}
f_Z(z) &= \sum_{k=0}^{z} P[X = k, Y = Z - k] \\
&= \sum_{k=0}^{z} P[X = k] \cdot P[Y = Z - k] \\
&= \sum_{k=0}^{z} \frac{\lambda_X^k e^{-\lambda_X}}{k!} \cdot \frac{\lambda_Y^{Z-k} e^{-\lambda_Y}}{(Z-k)!} \\
&= e^{-\lambda_X} e^{-\lambda_Y} \sum_{k=0}^{z} \frac{\lambda_X^k \lambda_Y^{Z-k}}{k!(Z-k)!} \\
&= \frac{e^{-(\lambda_X + \lambda_Y)}}{z!} \sum_{k=0}^{z} \frac{z!}{k!(Z-k)!} \lambda_X^k \lambda_Y^{Z-k} \\
&= \frac{e^{-(\lambda_X + \lambda_Y)}}{z!} (\lambda_X + \lambda_Y)^z
\end{aligned}$$

If $\lambda = \lambda_X + \lambda_Y$, $f_Z(z) = \frac{e^{-\lambda}\lambda^z}{z!}$. This is the Poisson distribution with parameter $\lambda = \lambda_1 + \lambda_2$. □

Theorem 8.12. If X and Y are independent random variables having normal distributions with parameters μ_X, σ_X and μ_Y, σ_Y respectively, then $X + Y$ has a normal distribution with parameters $\mu_{X+Y} = \mu_X + \mu_Y$ and $\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$. In addition, $X - Y$ has a normal distribution with parameters $\mu_{X-Y} = \mu_X - \mu_Y$ and $\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$.

Proof.

$$\mu_{X+Y} = \text{E[X + Y]} = \text{E[X]} + \text{E[Y]} = \mu_X + \mu_Y$$

$$\mu_{X-Y} = \text{E[X − Y]} = \text{E[X]} - \text{E[Y]} = \mu_X - \mu_Y$$

$$\text{Var[X + Y]} = \text{Var[X]} + \text{Var[Y]} \to \sigma_{X+Y} = \sqrt{\text{Var[X]} + \text{Var[Y]}} = \sqrt{\sigma_X^2 + \sigma_Y^2}$$

$$\text{Var[X − Y]} = \text{Var[X]} + \text{Var[Y]} \to \sigma_{X+Y} = \sqrt{\text{Var[X]} + \text{Var[Y]}} = \sqrt{\sigma_X^2 + \sigma_Y^2}$$

□

Example 8.7. To get to an important meeting, Bill has to drive on two highways to catch a train to Philadelphia. His travel time X on the first highway is a normally distributed random variable with mean 30 minutes and standard deviation 6 minutes. His travel time Y on the second highway is also a normally distributed random variable with mean 45 minutes and standard deviation 8 minutes. If he leaves his home at 8:00, what is the probability he will miss his train which leaves promptly at 9:20? Assume independence.
Let $Z = X + Y$. Then Z has a normal distribution and

$$\mu_Z = \mu_X + \mu_Y = 30 + 45 = 75$$

$$\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2} = \sqrt{6^2 + 8^2} = 10$$
$$P[Z > 80] = \text{normalcdf}(80, E99, 75, 10) = 0.3085$$

Theorem 8.13. Central Limit Theorem: Let $X_1, X_2, \dots, X_n$ be IID random variables, each having mean μ and standard deviation σ. Let $\bar{X} = \frac{X_! + X_2 + \cdots + X_n}{n}$ and let $X* = \frac{\bar{X} - mu}{\frac{\sigma}{\sqrt{n}}}$. As $n \to \infty$, the distribution of $X*$ approaches a standard normal distribution.
Consequences:

1. If n is large, $X*$ is approximately normal with $\mathrm{E[X*]} = 0$ and $\sigma_{X*} = 1$
2. If n is large, $\bar{X}$ is approximately normal with $\mathrm{E[\bar{X}]} = \mu$ and $\sigma_{\bar{X}} = \frac{\mu}{\sqrt{n}}$
3. If n is large, $S = X_1 + X_2 + \cdots + X_n$ is approximately normal with $\mathrm{E[S]} = n\mu$ and $\sigma_S = \sigma\sqrt{n}$

Recall that

$$\begin{aligned}
\mathrm{E[S]} &= \mathrm{E[X_1 + X_2 + \cdots + X_n]} \\
&= \mathrm{E[X_1] + E[X_2] + \cdots + E[X_n]} \\
&= \mu + \mu + \cdots + \mu = n\mu \\
\mathrm{Var[S]} &= \mathrm{Var[X_1 + X_2 + \cdots + X_n]} \\
&= \mathrm{Var[X_1] + Var[X_2] + \cdots + Var[X_n]} \\
&= \sigma^2 + \sigma^2 + \cdots + \sigma^2 = n\sigma^2 \\
\sigma_S &= \sigma\sqrt{n} \\
\mathrm{E[\bar{X}]} &= \mathrm{E[\frac{S}{n}]} \\
&= \frac{1}{n}\mathrm{E[S]} \\
&= \frac{1}{n} \cdot n\mu = \mu \\
\mathrm{Var[\bar{X}]} &= \mathrm{Var[\frac{S}{n}]} \\
&= \frac{1}{n^2}\mathrm{Var[S]} \\
&= \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n} \\
\sigma_{\bar{X}} &= \frac{\sigma}{\sqrt{n}}
\end{aligned}$$

Note: For most applications, "large" means $n \geq 30$.

Example 8.8. The life of a lithium battery has an exponential distribution with parameter $\theta = 10$ years.
a) Compute the probability that a single battery will last between 10 and 12 years.

$$f(x) = \begin{cases} \frac{1}{10}e^{-\frac{x}{10}} & \text{if } x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

$$P[10 \leq x \leq 12] = \int_{10}^{12} \frac{1}{10} e^{-\frac{x}{10}} dx = -e^{-\frac{x}{10}} \Big|_{10}^{12} = -e^{-12} + e^{-1} = 0.0667$$

b) Compute the probability that 100 batteries will have an average between 10 and 12 years.

$$\text{E[X]} = \theta = 10$$

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{100}}{100}$$

$$\mu_{\bar{X}} = \mu_X = \text{E[X]} = 10$$

$$\text{Var[X]} = \theta^2 = 100$$

$$\sigma_X = 10$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = 1$$

$$P[10 \leq x \leq 12] = \text{normalcdf}(10, 12, 10, 1) = 0.477$$

If $X_1, X_2, \dots, X_n$ have Bernoulli distributions, with parameter p, assuming independence, the random variable $S = X_1 + X_2 + \cdots + X_n$ will have a binomial distribution with parameters n and p. Furthermore, S has a mean of np and a standard deviation of $\sqrt{npq}$. Since S is the sum of IID random variables, S can be approximated by a normal distribution having the same mean and standard deviation (assuming n is large).

Definition 8.6. Continuity Correction Factor: a factor of half a number placed in addition at both ends of a probability distribution calculation

Example 8.9. Binomial distribution: $n = 16$, $p = \frac{1}{2}$
Exact Binomial probability:

$$P[5 \leq x \leq 11] = \text{binomcdf}(16, \frac{1}{2}, 11) - \text{binomcdf}(16, \frac{1}{2}, 4) = 0.9232$$

Normal Distribution without CCF: $\mu = np = 8$, $\sigma = \sqrt{npq} = 2$

$$P[5 \leq x \leq 11] = \text{normalcdf}(5, 11, 8, 2) = 0.8664$$

$$\text{Percent Error} = \frac{0.9232 - 0.8664}{0.9232} = 0.062 = 6.2\%$$

Normal Distribution with CCF:

$$P[5 \leq x \leq 11] = \text{normalcdf}(4.5, 11.5, 8, 2) = 0.91922$$

$$\text{Percent Error} = \frac{0.9232 - 0.9192}{0.9232} = 0.004 = 0.4\%$$

9 Elements of Statistical Inference

Definition 9.1. Consider a population of size N

$$P = \{x_1, x_2, \dots, x_n\}$$

Population Mean:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Population Variance:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

Population Standard Deviation:

$$\sigma$$

Definition 9.2. Select a random sample of size n where $n << N$

$$\text{Sample} = \{x_1, x_2, \dots, x_n\}$$

Sample Mean:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

We would like to use $\bar{x}$ to approximate μ (mean of the population).

- Point Estimate: $\mu = \bar{x}$ - AVOID
- Interval Estimate: $\mu \in (\bar{x} - \epsilon, \bar{x} + \epsilon)$

The larger the value of ϵ, the more likely our estimate is correct. However, if ϵ is too large, our estimate is useless. Ideally, we sample without replacement. However, if $n << N$, we can assume sampling is with replacement.

Definition 9.3. Sample Standard Deviation:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

Theorem 9.1.

$$\mathrm{E}[\mathrm{s}^2] = \sigma^2$$

s^2 is called an unbiased estimator of σ^2.

Proof. Verification for $n = 3$:

$$\begin{aligned}
s^2 &= \frac{1}{2}\sum_{i=1}^{3}(x_i - \bar{x})^2 \\
&= \frac{1}{2}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2] \\
&= \frac{1}{2}[(x_1 - \frac{x_1 + x_2 + x_3}{3})^2 + (x_2 - \frac{x_1 + x_2 + x_3}{3})^2 + (x_3 - \frac{x_1 + x_2 + x_3}{3})^2] \\
&= \frac{1}{2}[(\frac{2x_1 - x_2 - x_3}{3})^2 + (\frac{2x_2 - x_1 - x_3}{3})^2 + (\frac{2x_3 - x_1 - x_2}{3})^2] \\
&= \frac{1}{18}[(2x_1 - x_2 - x_3)^2 + (2x_2 - x_1 - x_3)^2 + (2x_3 - x_1 - x_2)^2] \\
&= \frac{1}{18}[6x_1^2 + 6x_2^2 + 6x_3^2 - 6x_1x_2 - 6x_2x_3 - 6x_1x_3] \\
&= \frac{1}{3}[x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_2x_3 - x_1x_3] \\
\mathrm{E[s^2]} &= \frac{1}{3}[\mathrm{E[x_1^2] + E[x_2^2] + E[x_3^2] - E[x_1x_2] - E[x_2x_3] - E[x_1x_3]}] \\
&\text{Assuming that } x_1, x_2, x_3 \text{ are independent} \\
&= \frac{1}{3}[\mathrm{E[x_1^2] + E[x_2^2] + E[x_3^2] - E[x_1]E[x_2] - E[x_2]E[x_3] - E[x_1]E[x_3]}] \\
&= \frac{1}{3}[\mathrm{E[x_1^2] + E[x_2^2] + E[x_3^2]} - \mu - \mu - \mu] \\
&= \frac{1}{3}[\mathrm{Var[x_1] + Var[x_2] + Var[x_3]}] \\
&= \frac{1}{3}[\sigma^2 + \sigma^2 + \sigma^2] \\
&= \sigma^2
\end{aligned}$$

□

Four Cases:

1. Assume the population is normally distributed. We know that $\bar{x}$ will have a normal distribution regardless of the size of n.

$$\mathrm{E[\bar{x}] = E[x]} = \mu(\mu_{\bar{x}} = \mu)$$

$$\mathrm{Var[\bar{x}]} = \frac{1}{n}\mathrm{Var[x]}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

We would like to mathematically determine ϵ so that

$$P[\bar{x} - \epsilon < \mu < \bar{x} + \epsilon] = 1 - \alpha$$

If α is known, we can compute $-z_{\frac{\alpha}{2}}$ and $z_{\frac{\alpha}{2}}$ so that the area between them is $1 - \alpha$ on the standard normal curve.

Example 9.1. $\alpha = 0.05, \frac{\alpha}{2} = 0.025, 1-\alpha = 0.95$

$$-z_{\frac{\alpha}{2}} = \text{invNorm}(0.025, 0, 1) = -1.96$$

$$z_{\frac{\alpha}{2}} = 1.96$$

Let $z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}$.

$$P[-1.96 < z < 1.96] = 0.95$$
$$P[-1.96 < \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} < 1.96] = 0.95$$
$$P[-1.96 < \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < 1.96] = 0.95$$
$$P[-1.96\frac{\sigma}{\sqrt{n}} < \bar{x}-\mu < 1.96\frac{\sigma}{\sqrt{n}}] = 0.95$$
$$P[-1.96\frac{\sigma}{\sqrt{n}} < \mu-\bar{x} < 1.96\frac{\sigma}{\sqrt{n}}] = 0.95$$
$$P[\bar{x} - \underbrace{1.96\frac{\sigma}{\sqrt{n}}}_{-\epsilon} < \mu < \bar{x} + \underbrace{1.96\frac{\sigma}{\sqrt{n}}}_{\epsilon}] = 0.95$$

$$(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}})$$

is called the 95% confidence interval for μ.

Example 9.2. A population is normally distributed and is known to have a standard deviation $\sigma = 10$. A random sample of size 25 is selected and $\bar{x}$ is computed to be 80. Find the 95% confidence interval for μ.

$$(80 - 1.96\frac{10}{\sqrt{25}}, 80 + 1.96\frac{10}{\sqrt{25}})$$

$$(76.08, 83.92)$$

Example 9.3. Repeat for the 98% confidence interval. $\alpha = 1 - 0.98 = 0.02$

$$z_{\frac{\alpha}{2}} = z_{0.01} = \text{invNorm}(0.01, 0, 1) = 2.33$$

$$(80 - 2.33\frac{10}{\sqrt{25}}, 80 + 2.33\frac{10}{\sqrt{25}})$$
$$(75.34, 84.66)$$

2. If the population is not normally distributed, $\bar{x}$ will have an approximately normal distribution provided that n is large ($n \geq 30$) by the Central Limit Theorem. Proceed the same way as (1).

3. The population standard deviation is unknown. If n is large, we can use the t- distribution with $n-1$ degrees of freedom. Then the confidence interval is:

$$(\bar{x} - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}})$$

Example 9.4. Find the t value for the 95% confidence interval if $n = 30$ and $df = 29$.

$$-t_{\frac{\alpha}{2}} = \text{invT}(0.025, 29) = -2.045$$
$$t_{\frac{\alpha}{2}} = 2.045$$

Example 9.5. In a study of 100 randomly selected NYC high school students, the mean number of hours per week they study was found to be 16.6 with a standard deviation of 2.8. Construct the 95% confidence interval for the mean study time of all NYC high school students.

$$\bar{x} = 16.6, s = 2.8, df = 99, \frac{\alpha}{2} = 0.025 \to t_{\frac{\alpha}{2}} = 1.984$$
$$(16.6 - 1.984\frac{2.8}{\sqrt{100}}, 16.6 + 1.984\frac{2.8}{\sqrt{100}})$$
$$(16.044, 17.156)$$

4. The population is assumed to be normally distributed. Sample size does not matter. Use the t- distribution with $n-1$ degrees of freedom.

   Example 9.6. The prices, in dollars, for a particular model of camera at 10 dealers randomly selected in NYC are: 225, 240, 215, 206, 211, 210, 193, 250, 225, 202. Assuming the population is normal, construct a 99% confidence interval for the mean price of all dealers in NYC. Using T-Interval:

$$\bar{x} = 217.7, S_x = 17.486, n = 10$$
$$(207.56, 227.84)$$

Suppose we would like to estimate the mean μ of a population with an absolute error less than ϵ:

1. Case 1: If σ is known and the population is normal OR σ is known and the sample size is large

$$\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$
$$|\text{error}| < \epsilon$$
$$z_{\frac{\alpha}{2}}\sigma < \mu\sqrt{n}$$
$$\sqrt{n} < \frac{z_{\frac{\alpha}{2}}\sigma}{\epsilon}$$
$$n > (\frac{z_{\frac{\alpha}{2}}\sigma}{\epsilon})^2$$

   Example 9.7. Consider the example of NYC high school students with $n = 100, \bar{x} = 16.6$ hours of study time and $s = 2.8$.
   At the 95% confidence level, $(16.044 < \mu < 17.158)$, determine an appropriate sample size so that at the 95% confidence level, $|\text{error}| < 0.5$.

$$n > (\frac{(1.96)(2.08)}{0.5}) = 120.47 \to n \geq 121$$

2. Case 2: If σ is unknown but $n \geq 30$, we replace σ by s and proceed as in case 1.
$$n > (\frac{z_{\frac{\alpha}{2}}\sigma}{\epsilon})^2$$

3. Case 3: If the population is normal, σ is unknown and the sample size is not necessarily small. Use case 2 with $z_{\frac{\alpha}{2}}$ in place of $t_{\frac{\alpha}{2}}$. Then compute n as in case 2, and then check the confidence interval using $t_{\frac{\alpha}{2}}$ with $n-1$ degrees of freedom. If $|\text{error}| < \epsilon$, then use this value of n. Otherwise increase n until $|\text{error}| < \epsilon$.

10 Estimation

Suppose we have 2 large populations X and Y which have means μ_X and μ_Y respectively. From population X, we select a random sample of size n_x and compute $\bar{x}$ and s_x. From population Y, we select a sample of size n_y and compute $\bar{y}$ and s_y. We would like to use $\bar{x} - \bar{y}$ to estimate the difference in the means of the 2 populations or $\sigma_X - \sigma_Y$.
$$\begin{aligned} \text{E}[\bar{\text{x}} - \bar{\text{y}}] &= \text{E}[\bar{\text{x}}] - \text{E}[\bar{\text{y}}] = \sigma_x - \sigma_y \\ \mu_{\bar{x}-\bar{y}} &= \mu_X - \mu_Y \\ \text{Var}[\bar{\text{x}} - \bar{\text{y}}] &= \text{Var}[\bar{\text{x}}] + \text{Var}[\bar{\text{y}}] = \frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \end{aligned}$$

Here we are using s_x and s_y in place of σ_X and σ_Y provided the sample size is large enough.
$$\sigma_{\bar{x}-\bar{y}} = \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$$

We proceed to find a confidence interval using the same procedure as for a single population but using $\bar{x} - \bar{y}$ in place of just $\bar{x}$, $\mu_{\bar{x}-\bar{y}}$ in place of μ_x and $s_{\bar{x}-\bar{y}}$ in place of $s_{\bar{x}}$.

Example 10.1. A random sample of 169 pages typed by Ms. X showed a mean of 3.1 mistakes and standard deviation of 0.65. A random sample of 121 pages typed by Ms. Y showed an average of 2.7 mistakes per page with a standard deviation of 0.66. Find a 95% confidence interval for the difference in the mean number of mistakes for all pages typed by the two typists.
$$\begin{gathered} \bar{x} = 3.1, \bar{y} = 2.7 \\ s_x = 0.65, s_y = 0.66 \\ n_x = 1.69, n_y = 1.21 \\ z_{\frac{\alpha}{2}} = 1.96 \\ (\bar{x} - \bar{y}) - z_{\frac{\alpha}{2}} s_{\bar{x}-\bar{y}} < \mu_x - \mu_y < (\bar{x} - \bar{y}) + z_{\frac{\alpha}{2}} s_{\bar{x}-\bar{y}} \\ \bar{x} - \bar{y} = 0.4 \\ s_{\bar{x}-\bar{y}} = \sqrt{\frac{0.65^2}{169} + \frac{0.66^2}{121}} = 0.0781 \\ 0.4 < (1.96)(0.0781) < \mu_x - \mu_y < 0.4 + (1.96)(0.0781) \\ 0.2469 < \mu_x - \mu_y < 0.5531 \end{gathered}$$