Topic 1: General Probability

SAMPLE SPACES AND EVENTS

Sample point and sample space: A sample point is the simple outcome of a random experiment. The sample space is the collection of all sample points related to a specified experiment.

Mutually exclusive outcomes: Outcomes are mutually exclusive if they cannot occur simultaneously. They are also referred to as disjoint outcomes.

Exhaustive outcomes: Outcomes are exhaustive if they combine to be the entire sample space, or equivalently, if at least one of the outcomes must occur whenever the experiment is performed.

Event: Any collection of sample points, or any subset of the sample space is referred to as an event.

Union of events A and B: $A \cup B$ denotes the union of events A and B, and consists of all sample points that are in either A or B.

Union of events $A_1, A_2, ..., A_n$: $A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i$ denotes the union of the events $A_1, A_2, ..., A_n$, and consists of all sample points that are in at least one of the A_i's. This definition can be extended to the union of infinitely many events.

Intersection of events $A_1, A_2, ..., A_n$: $A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i$ denotes the intersection of the events $A_1, A_2, ..., A_n$, and consists of all sample points that are simultaneously in all of the A_i's. $A \cap B$ is also denoted $A \cdot B$ or AB.

Mutually exclusive events $A_1, A_2, ..., A_n$: Two events are mutually exclusive if they have no sample points in common, or equivalently, if they have **empty intersection**. Events $A_1, A_2, ..., A_n$ are mutually exclusive if $A_i \cap A_j = \emptyset$ for all $i \neq j$, where $\emptyset$ denotes the empty set with no sample points.

Exhaustive events $B_1, B_2, ..., B_n$: If $B_1 \cup B_2 \cup \cdots \cup B_n = S$, the entire sample space, then the events $B_1, B_2, ..., B_n$ are referred to as exhaustive events.

Complement of event A: The complement of event A consists of all sample points in the sample space that are **not in A**. The complement is denoted $\overline{A}$, $\sim A$, A' or A^c and is equal to $\{x : x \notin A\}$.

Subevent (or subset) A of event B: If event B contains all the sample points in event A, then A is a subevent of B, denoted $A \subset B$. The occurrence of event A implies that event B has occurred.

Partition of event A: Events $C_1, C_2, ..., C_n$ form a partition of event A if $A = \bigcup_{i=1}^{n} C_i$ and the C_i's are mutually exclusive.

DeMorgan's Laws:

(i) $(A \cup B)' = A' \cap B'$ and

$$\left(\bigcup_{i=1}^{n} A_i \right)' = (A_1 \cup A_2 \cup \cdots \cup A_n)' = A_1' \cap A_2' \cap \cdots \cap A_n' = \bigcap_{i=1}^{n} A_i'$$

(ii) $(A \cap B)' = A' \cup B'$ and

$$\left(\bigcap_{i=1}^{n} A_i \right)' = (A_1 \cap A_2 \cap \cdots \cap A_n) = A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i$$

Indicator function for event A: The function $I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$ is the indicator function for event A, where x denotes a sample point.

Example 1: Suppose that an "experiment" consists of tossing a six-faced die. The **sample space** of outcomes consists of the set $\{1, 2, 3, 4, 5, 6\}$, each number being a **sample point** representing the number of spots that can turn up when the die is tossed. The outcomes 1 and 2 (or more formally, $\{1\}$ and $\{2\}$) are **mutually exclusive** when tossing a die. The outcomes (sample points) 1 to 6 are **exhaustive** for the experiment of tossing a die. The collection $\{2, 4, 6\}$ represents the **event** of tossing an even number when tossing a die. If $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$ then $A \cup B = \{1, 2, 3, 4, 6\}$ and $A \cap B = \{2\}$. The events $A =$ "a number less than 4 is tossed" $= \{1, 2, 3\}$ and $C =$ "a 4 is tossed" $= \{4\}$ are **mutually exclusive** since they have no sample points in common - A $\cap C = \emptyset$. If $A = \{1, 2, 3\}$, then $A' = \{4, 5, 6\}$.
If $D =$ "a 2 is tossed" $= \{2\}$ and $B =$ "an even number is tossed" $= \{2, 4, 6\}$, then $D \subset B$. The events $E =$ "a 2 or 4 is tossed" $= \{2, 4\}$ and $F =$ "a 6 is tossed" $= \{6\}$ form a **partition** of the event $B =$ "an even number is tossed" $= \{2, 4, 6\}$. For the die-tossing experiment, if $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$, then $A' = \{4, 5, 6\}$ and $B' = \{1, 3, 5\}$, and $A \cup B = \{1, 2, 3, 4, 6\}$, so that $(A \cup B)' = \{5\} = A' \cap B'$. □

Some rules concerning operations on events:

(i) $A \cap (B_1 \cup B_2 \cup \cdots \cup B_n) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)$ and
$A \cup (B_1 \cap B_2 \cap \cdots \cap B_n) = (A \cup B_1) \cap (A \cup B_2) \cap \cdots \cap (A \cup B_n)$ for any events $A, B_1, B_2, ..., B_n$

(ii) If $B_1, B_2, ..., B_n$ are exhaustive events ($\bigcup_{i=1}^{n} B_i = S$, the entire sample space), then for any event A, $A = A \cap (B_1 \cup B_2 \cup \cdots \cup B_n)$. As a special case of this, for any events A and B, $A \cap (B \cup B') = (A \cap B) \cup (A \cap B') = A$, so that $A \cap B$ and $A \cap B'$ form a partition of event A

(iii) For any event A, $A \cup A' = S$, the entire sample space, and $A \cap A' = \emptyset$

(iv) $A \cap B' = \{x : x \in A \text{ and } x \notin B\}$ is sometimes denoted $A - B$, and consists of all sample points that are in event A but not in event B

(v) If $A \subset B$ then $A \cup B = B$ and $A \cap B = A$.

PROBABILITY

Probability function: $P[a_i]$ or p_i denotes the probability that sample point (or outcome) a_i occurs; P must satisfy the following two conditions:

(i) $0 \leq P[a_i] \leq 1$ for each a_i in the sample space, and

(ii) $P[a_1] + P[a_2] + \cdots = \sum_i P[a_i] = 1$; this definition applies to both finite and infinite sample spaces;

Uniform probability function: If a sample space has a finite number of sample points, say k points - $a_1, a_2, ..., a_k$, then the probability function is said to be uniform if each sample point has the same probability of occurring - $P[a_i] = \frac{1}{k}$ for each $i = 1, 2, ..., k$.

Probability of event A: $P[A] = \sum_{a_i \in A} P[a_i]$, the sum of $P[a_i]$ over all sample points in event A.

Example 2: In tossing a "fair" die, it is assumed that each of the six faces has the same chance of $\frac{1}{6}$ of turning up. If this is true, then the probability function $P(j) = \frac{1}{6}$ for $j = 1, 2, 3, 4, 5, 6$ is a uniform probability function on the sample space $\{1, 2, 3, 4, 5, 6\}$.
The event "an even number is tossed" is $A = \{2, 4, 6\}$, and has probability
$P[A] = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$. $\square$

Some rules concerning operations on events:

(i) $P[\emptyset] = 0$

(ii) $P[S] = 1$ if S is the entire sample space

(iii) if events $A_1, A_2, ..., A_n$ are mutually exclusive then
$P[\bigcup_{i=1}^{n} A_i] = P[A_1 \cup A_2 \cup \cdots \cup A_n] = P[A_1] + P[A_2] + \cdots + P[A_n] = \sum_{i=1}^{n} P[A_i]$;
this extends to infinitely many mutually exclusive events

(iv) for any event A, $0 \leq P[A] \leq 1$

(v) if $A \subset B$ then $P[A] \leq P[B]$

(vi) for any events A, B and C, $P[A \cup B] = P[A] + P[B] - P[A \cap B]$ (since $P[A] + P[B]$ counts $P[A \cap B]$ twice), and
$P[A \cup B \cup C] = P[A] + P[B] + P[C] - P[A \cap B] - P[A \cap C] - P[B \cap C] + P[A \cap B \cap C]$;

(vii) for any events $A_1, A_2, ..., A_n$, $P[\bigcup_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} P[A_i]$, with equality holding if and only if the events are mutually exclusive ;

(viii) for any event A, $P[A'] = 1 - P[A]$;

(ix) for any events A and B, $P[A] = P[A \cap B] + P[A \cap B']$;

(x) for exhaustive events $B_1, B_2, ..., B_n$, $P[\bigcup_{i=1}^{n} B_i] = 1$,
if $B_1, B_2, ..., B_n$ are exhaustive and mutually exclusive, they form a partition of the entire sample space, and for any event A,

$$P[A] = P[A \cap B_1] + P[A \cap B_2] + \cdots + P[A \cap B_n] = \sum_{i=1}^{n} P[A \cap B_i]$$

(xi) if P is a uniform probability function on a sample space with k points, and if event A consists of m sample points, then $P[A] = \frac{m}{k}$

Example 3: A survey is made to determine the number of households having electric appliances in a certain city. It is found that 75% have radios (R), 65% have irons (I), 55% have electric toasters (T), 50% have (IR), 40% have (RT), 30% have (IT), and 20% have all three. Find the probability that a household has at least one of these appliances.

Solution:
$$P[R \cup I \cup T] = P[R] + P[I] + P[T] - P[R \cap I] - P[R \cap T] - P[I \cap T] + P[R \cap I \cap T]$$
$$= .75 + .65 + .55 - .5 - .4 - .3 + .2 = .95 \quad \square$$

Example 4: Let $P[A \cap B] = .2$, $P[A] = .6$, and $P[B] = .5$. Find $P[A' \cup B']$.

Solution: $P[A' \cup B'] = P[(A \cap B)'] = 1 - P[A \cap B] = .8$. $\square$

CONDITIONAL PROBABILITY AND INDEPENDENCE OF EVENTS

Conditional probability of event B given event A: If $P[A] > 0$, we define

$$P[B|A] = \frac{P[B \cap A]}{P[A]}$$

the conditional probability that event B occurs given that event A has occurred.

Bayes rule and Bayes Theorem: For any events A and B with $P[A] > 0$,

$$P[B|A] = \frac{P[A|B] \cdot P[B]}{P[A]}$$

If $B_1, B_2, ..., B_n$ form a partition of the entire sample space S, then

$$P[B_j|A] = \frac{P[A|B_j] \cdot P[B_j]}{\sum_{i=1}^{n} P[A|B_i] \cdot P[B_i]} \quad \text{for each} \quad j = 1, 2, ..., n$$

The values of $P[B_j]$ are called prior probabilities, and the value of $P[B_j|A]$ is called a posterior probability.

Independent events A and B: If events A and B satisfy the relationship $P[A \cap B] = P[A] \cdot P[B]$, then the events are said to be independent or stochastically independent or statistically independent. The independence of (non-empty) events A and B is equivalent to $P[A|B] = P[A]$ or $P[B|A] = P[B]$.

Mutually independent events $A_1, A_2, ..., A_n$: If events $A_1, A_2, ..., A_n$ satisfy the relationship

$$P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1] \cdot P[A_2] \cdots P[A_n] = \prod_{i=1}^{n} P[A_i]$$

then, events are said to be mutually independent.

Example 5: Suppose the die-tossing experiment is considered again. The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. We define the following events:

$A =$ "the number tossed is ≤ 3" $= \{1, 2, 3\}$

$B =$ "the number tossed is even" $= \{2, 4, 6\}$

$C =$ "the number tossed is a 1 or a 2" $= \{1, 2\}$

$D =$ "the number tossed doesn't start with the letters 'f' or 't'" $= \{1, 6\}$

The conditional probability of A given B is $P[A|B] = \frac{P[\{1,2,3\} \cap \{2,4,6\}]}{P[\{2,4,6\}]} = \frac{P[\{2\}]}{P[\{2,4,6\}]} = \frac{1/6}{1/2} = \frac{1}{3}$.

Events A and B are not independent, since $\frac{1}{6} = P[A \cap B] \neq P[A] \cdot P[B] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$, or alternatively, events A and B are not independent since $P[A|B] \neq P[A]$.

$P[A|C] = 1 \neq \frac{1}{2} = P[A]$, so that A and C are not independent.

$P[B|C] = \frac{1}{2} = P[B]$, so that B and C are independent

(alternatively, $P[B \cap C] = P[\{2\}] = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = P[B] \cdot P[C]$).

Both A and B are independent of D. □

Some rules concerning conditional probability and independence are:

(i) $P[A \cap B] = P[B|A] \cdot P[A] = P[A|B] \cdot P[B]$ for any independent events A and B

(ii) If $B_1, B_2, ..., B_n$ form a partition of the sample space S, then for any event A,

$$P[A] = \sum_{i=1}^{n} P[A|B_i] \cdot P[B_i] \; ;$$

as a special case, for any events A and B, $P[A] = P[A|B] \cdot P[B] + P[A|B'] \cdot P[B']$

(iii) If $P[A_1 \cap A_2 \cap \cdots \cap A_{n-1}] > 0$, then

$$P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1] \cdot P[A_2|A_1] \cdot P[A_3|A_1 \cap A_2] \cdots P[A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}]$$

(iv) $P[A'|B] = 1 - P[A|B]$

(v) if $A \subset B$ then $P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[A]}{P[B]}$, and $P[B|A] = 1$

(vi) if A and B are independent events then A' and B are independent events, A and B' are independent events, and A' and B' are independent events

(vii) since $P[\emptyset] = P[\emptyset \cap A] = 0 = P[\emptyset] \cdot P[A]$ for any event A, it follows that $\emptyset$ is independent of any event A

IMPORTANT NOTE: The following manipulation of event probabilities arises from time to time: $P[B] = P[B|A] \cdot P(A) + P[B|\overline{A}] \cdot P(\overline{A})$. If we are given conditional probabilities for event B given some other event A and its complement $\overline{A}$, and if we are given the (unconditional) probability of event A, then we can find the probability of event B. One of the important aspects of applying this relationship is the determination of the appropriate events A and B.

Example 6: If $P[A] = \frac{1}{6}$ and $P[B] = \frac{5}{12}$, and $P[A|B] + P[B|A] = \frac{7}{10}$, find $P[A \cap B]$.

Solution: $P[B|A] = \frac{P[A \cap B]}{P[A]} = 6P[A \cap B]$ and $P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{12}{5} P[A \cap B]$

$$\rightarrow (6 + \tfrac{12}{5}) \cdot P[A \cap B] = \tfrac{7}{10} \rightarrow P[A \cap B] = \tfrac{1}{12} \; . \qquad \square$$

Example 7: Three dice have the following probabilities of throwing a "six": p, q, r, respectively. One of the dice is chosen at random and thrown (each is equally likely to be chosen). A "six" appeared. What is the probability that the die chosen was the first one?

Solution: The event " a 6 is thrown" is denoted by "6"

$$P[\text{die 1}|\text{"6"}] = \frac{P[(\text{die 1}) \cap (\text{"6"})]}{P[\text{"6"}]} = \frac{P[\text{"6"}|\text{die1}] \cdot P[\text{die1}]}{P[\text{"6"}]} = \frac{p \cdot \frac{1}{3}}{P[\text{"6"}]} \; .$$

But $P["6"] = P[("6") \cap (\text{die } 1)] + P[("6") \cap (\text{die } 2)] + P[("6") \cap (\text{die } 3)]$

$$= P["6"|\text{die } 1] \cdot P[\text{die } 1] + P["6"|\text{die } 2] \cdot P[\text{die } 2] + P["6"|\text{die } 3] \cdot P[\text{die } 3]$$

$$= p \cdot \tfrac{1}{3} + q \cdot \tfrac{1}{3} + r \cdot \tfrac{1}{3} = \tfrac{p+q+r}{3} \rightarrow P[\text{die } 1|"6"] = \frac{p \cdot \frac{1}{3}}{P["6"]} = \frac{p \cdot \frac{1}{3}}{(p+q+r) \cdot \frac{1}{3}} = \frac{p}{p+q+r} . \ \square$$

Example 8: Identical twins come from the same egg and hence are of the same sex. Fraternal twins have a 50-50 chance of being the same sex. Among twins, the probability of a fraternal set is p and an identical set is $q = 1 - p$. If the next set of twins are of the same sex, what is the probability that they are identical?

Solution: Let A be the event "the next set of twins are of the same sex", and let B be the event "the next sets of twins are identical". We are given

$P[A|B] = 1$, $P[A|B'] = .5$, $P[B] = q$, $P[B'] = p = 1 - q$. Then $P[B|A] = \frac{P[A \cap B]}{P[A]}$.

But

$$P[A \cap B] = P[A|B] \cdot P[B] = q \quad \text{and} \quad P[A \cap B'] = P[A|B'] \cdot P[B'] = .5p .$$

Thus, $P[A] = P[A \cap B] + P[A \cap B'] = q + .5p = q + .5(1 - q) = .5(1 + q)$, and

$$P[B|A] = \frac{q}{.5(1+q)} . \qquad \square$$

Example 9: Let events A and B be independent. Find the probability, in terms of $P[A]$ and $P[B]$, that exactly one of the events A and B occurs.

Solution: $P[\text{exactly one of } A \text{ and } B] = P[(A \cap B') \cup (B \cap A')]$.

Since $A \cap B'$ and $B \cap A'$ are mutually exclusive, it follows that

$$P[\text{exactly one of } A \text{ and } B] = P[A \cap B'] + P[B \cap A'] .$$

Since A and B are independent, it follows that A and B' are also independent, as are B and A'. Then

$$P[(A \cap B') \cup (B \cap A')] = P[A] \cdot P[B'] + P[B] \cdot P[A']$$
$$= P[A](1 - P[B]) + P[B](1 - P[A]) = P[A] + P[B] - 2P[A] \cdot P[B] \ \square$$

COMBINATORIAL PRINCIPLES - PERMUTATIONS AND COMBINATIONS

Factorial notation: $n!$ denotes the quantity

$$n! = n(n-1)(n-2)\cdots 2\cdot 1\,; \qquad 0! \text{ is defined to be equal to } 1.$$

Permutations:

(a) Given n distinct objects, the number of different ways in which the objects may be ordered (permuted) is $n!$. For example, the set of 3 letters $\{a,b,c\}$ can be ordered in $3! = 6$ ways - abc, acb, bac, bca, cab, cba.

The number of ways of choosing an ordered subset of size k from n objects, without replacement (i.e., after the first object is chosen, the next object is chosen from the remaining $n-1$, the next after that from the remaining $n-2$, etc.) is

$$\frac{n!}{(n-k)!} = n\cdot(n-1)\cdot\cdots(n-k+1)\text{, and is denoted } {}_nP_k \text{ or } P_{n,k} \text{ or } P(n,k).$$

Using the set $\{a,b,c\}$ again, the number of ways of choosing an ordered subset of size 2 is

$$\frac{3!}{(3-2)!} = \frac{6}{1} = 6 \text{ - } ab, ac, ba, bc, ca, cb.$$

(b) Given n objects, of which n_1 are of Type 1, n_2 are of Type 2, . . . , and n_t are of Type t ($t \geq 1$ is an integer), and $n = n_1 + n_2 + \cdots + n_t$, the number of ways of ordering all n objects (where objects of the same Type are indistinguishable) is

$$\frac{n!}{n_1!\cdot n_2!\cdots n_t!}\text{, which is sometimes denoted } \binom{n}{n_1\; n_2 \cdots n_t}.$$

Combinations:

(a) Given n distinct objects, the number of ways of choosing a subset of size $k \leq n$ without replacement (and without regard to the order in which the objects are chosen) is $\frac{n!}{k!\cdot(n-k)!}$, which is usually denoted $\binom{n}{k}$(or ${}_nC_k$, $C_{n,k}$ or $C(n,k)$) and is read "n choose k". $\binom{n}{k}$ is also called a **binomial coefficient** (and can be defined for any real number n and non-negative integer k). Note that if n is an integer and k is a non-negative integer, then

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!}, \qquad \binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{1} = \binom{n}{n-1} = n, \text{ and } \binom{n}{k} = \binom{n}{n-k}.$$

Using the set $\{a,b,c\}$ again, the number of ways of choosing a subset of size 2 without replacement is $\binom{3}{2} = \frac{3!}{2!\cdot(3-2)!} = 3$ - these being $\{a,b\}$, $\{a,c\}$, $\{b,c\}$.

(b) Given n objects, of which n_1 are of Type 1, n_2 are of Type 2, . . . , and n_t are of Type t ($t \geq 1$ is an integer), and $n = n_1 + n_2 + \cdots + n_t$, the number of ways of choosing a subset of size $k \leq n$ (without replacement). with k_1 objects of Type 1, k_2 objects of Type 2,..., and k_t objects of Type t, where $k = k_1 + k_2 + \cdots + k_t$ is

$$\binom{n_1}{k_1}\cdot\binom{n_2}{k_2}\cdots\binom{n_t}{k_t}$$

Binomial Theorem: In the power series expansion of $(1+t)^N$, the coefficient of t^k is $\binom{N}{k}$, so that

$$(1+t)^N = \sum_{k=0}^{\infty}\binom{N}{k}\cdot t^k = 1 + Nt + \frac{N(N-1)}{2}t^2 + \frac{N(N-1)(N-2)}{6}t^3 + \cdots$$

If N is an integer, then the summation stops at $k = N$ and the series is valid for any real number t, but if N is not an integer, then the series is valid if $|t| < 1$.

Multinomial Theorem: In the power series expansion of $(t_1 + t_2 + \cdots + t_s)^N$ where N is a positive integer, the coefficient of $t_1^{k_1} \cdot t_2^{k_2} \cdots t_s^{k_s}$ (where $k_1 + k_2 + \cdots + k_s = N$) is

$$\binom{N}{k_1\ k_2\ \cdots\ k_s} = \frac{N!}{k_1! \cdot k_2! \cdots k_s!}$$

Special note: In questions involving coin flips or dice tossing, it is understood, unless indicated otherwise, that successive flips or tosses are independent of one another.

In making a random selection of an object from a collection of n objects, it is understood that each object has the same chance of being chosen, $\frac{1}{n}$. In questions that arise involving choosing k objects at random from a total of n objects, or in constructing a random permutation of a collection of objects, it is understood that each of the possible choices or permutations is equally likely to occur. For instance, if a purse contains one quarter, one dime, one nickel and one penny, and two coins are chosen, there are $\binom{4}{2} = 6$ possible ways of choosing two coins without regard to order of choosing - these are Q-D , Q-N , Q-P , D-N , D-P , N-P (the choice Q-D is regarded as the same as D-Q, etc.). It would be understood that each of the possible ways are equally likely, and each has (uniform) probability of $\frac{1}{6}$ of occurring - the6 sample space would consist of the ways of choosing, and each sample point would have probability $\frac{1}{6}$. Then, the probabili6ty of a particular event occurring would be $\frac{j}{6}$, where j is the number of sample points in the event. If A is the event "one of the coins is either a quarter or a dime", then $P[A] = \frac{5}{6}$, since event A consists of the 5 of the sample points $\{$Q-D , Q-N , Q-P , D-N , D-P $\}$.

Example 10: An ordinary die and a die whose faces have $2,\ 3,\ 4,\ 6,\ 7,\ 9,$ dots are tossed independently of one another, and the total number of dots on the two dice is recorded as N. Find the probability that $N \geq 10$.

Solution: It is assumed that for each die, each face has a $\frac{1}{6}$ probability of turning up. If the number of dots turning up on die 1 and die 2 are d_1 and d_2, respectively, then the tosses that result in $N = d_1 + d_2 \geq 10$ are

$(1,9)$, $(2,9)$, $(3,7)$, $(3,9)$, $(4,6)$, $(4,7)$, $(4,9)$, $(5,6)$, $(5,7)$, $(5,9)$, $(6,4)$, $(6,6)$, $(6,7)$, $(6,9)$

14 combinations in $6 \times 6 = 36$ combinations that can possibly occur. Since each of the 36 (d_1, d_2) combinations is equally likely, the probability is $\frac{14}{36}$. □

Example 11: Three nickels, one dime and two quarters are in a purse. In picking three coins at one time (without replacement), what is the probability of getting a total of at least 35 cents?

Solution: In order to get at least 35 cents, at least one quarter must be chosen. The possible choices are 1Q + any 2 of the non-quarters , or 2Q + any 1 of the non-quarters.
The total number of ways of choosing three coins from the six coins is $\binom{6}{3} = 20$.
If we label the two quarters as Q_1 and Q_2 , then the number of ways of choosing the three coins so that only Q_1 (and not Q_2) is in the choice is $\binom{4}{2} = 6$ (this is the number of ways of choosing the other two coins from the three nickels and one dime) - and therefore, the number of choices that contain only Q_2 (and not Q_1) is also 6. The number of ways of choosing the three coins so that both Q_1 and Q_2 are in the choice is 4 (this is the number of ways of choosing the other coin from the three nickels and one dime). Thus, the total number of choices for which at least one of the three coins chosen is a quarter is 16. The probability in question is $\frac{16}{20}$.
Alternatively, the number of three coin choices that do not contain any quarters is $\binom{4}{3} = 4$ (the number of ways of choosing the three coins from the 4 non-quarters), so that number of choices that contain at least one quarter is $20 - 4 = 16$. $\square$

Example 12: A and B draw coins in turn without replacement from a bag containing 3 dimes and 4 nickels. A draws first. It is known that A drew the first dime. Find the probability that A drew it on the first draw.

Solution: $P[A\text{ draws dime on first draw}|A\text{ draws first dime}] = \frac{P[A\text{ draws dime on first draw}]}{P[A\text{ draws first dime}]}$

$$P[A\text{ draws dime on first draw}] = \frac{3}{7}.$$

Since there only 3 dimes, in order for A to draw the first dime, this must happen on A's first, second or third draw. Thus,

$$P[A\text{ draws first dime}] = P[A\text{ draws dime on first draw}] + P[A\text{ draws first dime on second draw}] + P[A\text{ draws first dime on third draw}] .$$

$P[A\text{ draws dime on second draw}] = \frac{4}{7} \cdot \frac{3}{6} \cdot \frac{3}{5} = \frac{6}{35}$, since A's first draw is one of the four non-dimes, and B's first draw is one of the three remaining non-dimes after A's draw, and A's second draw is one of the three dimes of the five remaining coins. In a similar way, $P[A\text{ draws first dime on third draw}] = \frac{4}{7} \cdot \frac{3}{6} \cdot \frac{2}{5} \cdot \frac{1}{4} \cdot 1 = \frac{1}{35}$. Then, $P[A\text{ draws first dime}] = \frac{3}{7} + \frac{6}{35} + \frac{1}{35} = \frac{22}{35}$, and

$$P[A\text{ draws dime on first draw}|A\text{ draws first dime}] = \frac{3/7}{22/35} = \frac{15}{22} . \quad \square$$

Example 13: Three people, X, Y and Z, in order, roll an ordinary die. The first one to roll an even number wins. The game continues until someone rolls an even number. Find the probability that X will win.

Solution: Since X rolls first, fourth, seventh, etc. until the game ends, the probability that X will win is the probability that in throwing a die, the first even number will occur on the 1st, or 4th, or 7th, or . . . throw. The probability that the first even number occurs on the n-th throw is

$$(\tfrac{1}{2})^{n-1}(\tfrac{1}{2}) = \tfrac{1}{2^n} \quad n-1$$

of independence of successive throws, with $A_i =$ "throw i is even", the probability that the first even throw occurs on throw n is

$$P[A_1' \cap A_2' \cap \cdots \cap A_{n-1}' \cap A_n] = P[A_1'] \cdot P[A_2'] \cdots P[A_{n-1}'] \cdot P[A_n] \;\; = (\tfrac{1}{2})^{n-1}(\tfrac{1}{2}) = \tfrac{1}{2^n}$$

Thus, P[first even throw is on 1st, or 4th, or 7th, or ...] $= \frac{1}{2} + \frac{1}{2^4} + \frac{1}{2^7} + \cdots = \frac{1}{2}(1 + \frac{1}{8} + \frac{1}{8^2} + \cdots) = \frac{4}{7}$ · □

Example 14: Urn I contains 7 red and 3 black balls, and Urn II contains 4 red and 5 black balls. After a randomly selected ball is transferred from Urn I to Urn II, 2 balls are randomly drawn from Urn II without replacement. Find the probability that both balls drawn from Urn II are red.

Solution: Define the following events:

R_1: the ball transferred from Urn I to Urn II is red

B_1: the ball transferred from Urn I to Urn II is black

R_2: two red balls are selected from Urn II after the transfer from Urn I to Urn II .

Since R_1 and B_1 are mutually exclusive,

$$P[R_2] = P[R_2 \cap (R_1 \cup B_1)] = P[R_2 \cap R_1] + P[R_2 \cap B_1]$$
$$= P[R_2|R_1] \cdot P[R_1] + P[R_2|B_1] \cdot P[B_1] = \frac{\binom{5}{2}}{\binom{10}{2}} \cdot \frac{7}{10} + \frac{\binom{4}{2}}{\binom{10}{2}} \cdot \frac{3}{10} = \frac{44}{225} \,. \; \square$$

Example 15: A calculator has a random number generator button which, when pressed, displays a random digit $0, 1, ..., 9$. The button is pressed four times. Assuming that the numbers generated are independent of one another, find the probability of obtaining one "0", one "5", and two "9"'s in any order.

Solution: There are $10^4 = 10,000$ four-digit orderings that can arise, from 0-0-0-0 to 9-9-9-9 . From the notes above on permutations, if we have four digits, with one "0", one "5" and two "9"'s, the number of orderings is $\frac{4!}{1! \cdot 1! \cdot 2!} = 12$. The probability in question is then $\frac{12}{10,000}$.□

Topic 2: Univariate Random Variables

RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

Random variable X: A random variable is a function on a sample space S. This function assigns a real number $X(s)$ to each sample point $s \in S$. Often a random variable is simply equal to the sample point s, if the sample points are numerical values - for example, the sample space representing the number of spots that turn up when an ordinary die is tossed is $S = \{1, 2, 3, 4, 5, 6\}$, and $X(s) = s$ describes the random variable X which is the number of spots that turn up. Alternatively, suppose that a gamble based on the outcome of the toss of a die pays \$10 if an even number is tossed, and pays \$20 if an odd number is tossed. The payoff can be represented by the random variable Y, where $Y(s) = 10$ if s is even, and $Y(s) = 20$ if s is odd. A random variable is sometimes described in terms of the outcome of a random experiment (such as tossing a die), or may be described without explicit reference to the underlying random experiment or sample space (such as the prime rate of interest two years from now). Given a set of real numbers, A, then $P[X \in A]$ is defined to be the probability of the event represented by the related subset of the sample space $P\{s : X(s) \in A\}$. Using random variable Y from the \$10 for even, \$20 for odd die toss example, we have, as an example,
$P[Y \geq 12] = P[\{s : Y(s) \geq 12] = P[\{1, 3, 5\}]$, since these are the sample points s for which $Y(s) \geq 12$ (for a fair die, this probability is $\frac{1}{2}$).

Discrete random variable: The random variable X is discrete and is said to have a **discrete distribution** if it can take on values only from a finite or countable infinite sequence (usually the integers or some subset of the integers). As an example, consider the following two random variables related to successive tosses of a coin-

$X = 1$ if the first head occurs on an even-numbered toss, $X = 0$ if the first head occurs on an odd-numbered toss;

$Y = n$, where n is the number of the toss on which the first head occurs.

Both X and Y are discrete random variables, where X can take on only the values 0 or 1, and Y can take on any non-negative integer value. Both X and Y are based on the same sample space - the sample points are sequences of tail coin flips ending with a head coin flip:

$$S = \{H\,,\, TH\,,\, TTH\,,\, TTTH\,,\, TTTTH\,,\, ...\}.$$

Then,

$X(H) = 0$ (a head on flip one, an odd-numbered flip), $X(TH) = 1$, $X(TTH) = 0$, ...

$Y(H) = 1$ (first head on flip 1), $Y(TH) = 2$, $Y(TTH) = 3$, $Y(TTTH) = 4$, ...

Probability function of a discrete random variable: The probability function (p.f.) of a discrete random variable is usually denoted $f(x)$, $f_X(x)$, $p(x)$ or p_x, and is equal to $P[X = x]$. The probability function must satisfy

$$\text{(i)}\;\; 0 \le f(x) \le 1 \text{ for all } x, \text{ and } \quad \text{(ii)}\;\; \sum_x f(x) = 1 .$$

Given a set A of real numbers, $P[X \in A] = \sum_{x \in A} f(x)$.

Continuous random variable: A continuous random variable usually can assume numerical values from an interval of real numbers, perhaps the whole set of real numbers $\mathbb{R}$. As an example, the length of time between successive streetcar arrivals at a particular (in service) streetcar stop could be regarded as a continuous random variable (assuming that time measurement can be made perfectly accurate).

Probability density function: A continuous random variable X usually has a probability density function (p.d.f.) denoted $f(x)$ or $f_X(x)$ (or sometimes denoted $p(x)$), which is a continuous function except possibly at a finite number of points. Probabilities related to X are found by integrating the density function:

$$P[X \in (a,b)] = P[a < X < b] \text{ is defined to be equal to } \int_a^b f(x)\,dx .$$

$f(x)$ must satisfy

$$\text{(i)}\;\; f(x) \ge 0 \text{ for all } x, \text{ and } \quad \text{(ii)} \int_{-\infty}^{\infty} f(x)\,dx = 1 .$$

Often, the region of non-zero density is finite, and $f(x) = 0$ outside that interval.
If $f(x)$ is continuous except at a finite number of points, then probabilities are defined and calculated as if $f(x)$ was continuous everywhere (the discontinuities are ignored).

For example, suppose that X has density function

$$f(x) = \begin{cases} 2x \text{ for } 0<x<1 \\ 0, \text{ elsewhere} \end{cases}$$

Then f satisfies the requirements for a density function, since $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 2x\,dx = 1$.
Then, for example .

$$P[.2 < X < .5] = \int_{.2}^{.5} 2x\,dx = x^2\Big|_{.2}^{.5} = .21.$$

Mixed distribution: A random variable that has some points with non-zero **probability mass**, and with a continuous p.d.f. elsewhere is said to have a mixed distribution. The sum of the probabilities at the discrete points of probability plus the integral of the density function on the continuous region for X must be 1.

For example, suppose that X has probability of .5 at $X = 0$, and X is a continuous random variable on the interval $(0,1)$ with density function $f(x) = x$ for $0 < x < 1$, and X has no density or probability elsewhere. This satisfies the requirements for a random variable since

$$P[X = 0] + \int_0^1 f(x)\,dx = .5 + \int_0^1 x\,dx = .5 + .5 = 1.$$

Then, $P[0 < X < .5] = \int_0^{.5} x\,dx = .125$, and

$$P[0 \le X < .5] = P[X = 0] + P[0 < X < .5] = .5 + .125 = .625 .$$

Cumulative distribution function (and survival function): Given a random variable X, the cumulative distribution function of X (also called the **distribution function**, or c.d.f.) is $F(x) = P[X \le x]$ (also denoted $F_X(x)$). The survival function is the complement of the distribution function, $S(x) = 1 - F(x) = P[X > x]$. The event $X > x$ is referred to as a "tail" of the distribution. For a discrete random variable with probability function $f(x)$

$$F(x) = \sum_{w \le x} f(w)$$

In this case $F(x)$ is a "step function". It has a jump (or step increase) at each point with non-zero probability, while remaining constant until the next jump.

If X has a continuous distribution with density function $f(x)$, then

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

and $F(x)$ is a continuous, differentiable, non-decreasing function such that

$$\frac{d}{dx} F(x) = F'(x) = f(x),$$

If X has a mixed distribution, then $F(x)$ is continuous except at the points of non-zero probability mass, where $F(x)$ will have a jump. For any c.d.f.

$$P[a < X \le b] = F(b) - F(a) \ , \ \lim_{x \to \infty} F(x) = 1 \ , \ \lim_{x \to -\infty} F(x) = 0 \ .$$

Examples of distribution functions:
$X =$ number turning up when tossing one fair die, so X has probability function $f_X(x) = P[X = x] = \frac{1}{6}$ for $x = 1, 2, 3, 4, 5, 6$. X is a discrete random variable.

$$F_X(x) = P[X \le x] = \begin{cases} 0 \text{ if } x < 1 \\ \frac{1}{6} \text{ if } 1 \le x < 2 \\ \frac{2}{6} \text{ if } 2 \le x < 3 \\ \frac{3}{6} \text{ if } 3 \le x < 4 \\ \frac{4}{6} \text{ if } 4 \le x < 5 \\ \frac{5}{6} \text{ if } 5 \le x < 6 \\ 1 \text{ if } x \ge 6 \end{cases}$$

Y is a continuous random variable on the interval $(0, 1)$ with density function

$$f_Y(y) = \begin{cases} 3y^2 \text{ for } 0<x<1 \\ 0, \text{ elsewhere} \end{cases} .$$

Then

$$F_Y(y) = \begin{cases} 0 \text{ if } y < 0 \\ y^3 \text{ if } 0 \le y < 1 \\ 1 \text{ if } y \ge 1 \end{cases}$$

Z has a mixed distribution on the interval $[0,1)$. Z has probability of .5 at $Z=0$, and Z has density function $f_Z(z)=z$ for $0<z<1$, and Z has no density or probability elsewhere.

Then,

$$F_Z(z)=\begin{cases} 0 \text{ if } z<0 \\ .5 \text{ if } z=0 \\ .5+\frac{1}{2}z^2 \text{ if } 0<z<1 \\ 1 \text{ if } z\geq 1 \end{cases}$$

Some results and formulas relating to this section:

(i) For a continuous random variable X,

$$P[a<X<b]=P[a\leq X<b]=P[a<X\leq b]=P[a\leq X\leq b] \ ,$$

so that when calculating the probability for a continuous random variable on an interval, it is irrelevant whether or not the endpoints are included. For a continuous random variable,

$$P[X=a]=0 \ .$$

non-zero probabilities only exist over an interval, not at a single point. Also, for a continuous random variable, the hazard rate or failure rate is

$$h(x)=\frac{-f'(x)}{1-F(x)}=\frac{d}{dx}\,ln[1-F(x)]$$

(ii) If X has a mixed distribution, then $P[X=t]$ will be non-zero for some value(s) of t, and $P[a<X<b]$ will not always be equal to $P[a\leq X\leq b]$ (they will not be equal if X has a non-zero probability mass at either a or b).

(iii) $f(x)$ may be defined **piecewise**, meaning that $f(x)$ is defined by a different algebraic formula on different intervals.

(iv) A continuous random variable may have two or more different, but equivalent p.d.f.'s, but the difference in the p.d.f.'s would only occur at a finite (or countably infinite) number of points. The c.d.f. of a random variable of any type is always unique to that random variable.

Example 16: A die is loaded in such a way that the probability of the face with j dots turning up is proportional to j for $j=1,2,3,4,5,6$. What is the probability, in one roll of the die, that an even number of dots will turn up?

Solution: Let X denote the random variable representing the number of dots that appears when the die is rolled once. Then, $P[X=k]=R\cdot k$ for $k=1,2,3,4,5,6$, where R is the proportional constant. Since the sum of all of the probabilities of points in that can occur must be 1. it follows that $R\cdot[1+2+3+4+5+6]=1$, so that $R=\frac{1}{21}$.
Then, $P[\text{even number of dots turns up}]=P[2]+P[4]+P[6]=\frac{2+4+6}{21}=\frac{4}{7}$. $\square$

Example 17: An ordinary single die is tossed repeatedly until the first even number turns up. The random variable X is defined to be the number of the toss on which the first even number turns up. Find the probability that X is an even number.

Solution: X is a discrete random variable that can take on an integer value of 1 or more. The probability function for X is $f(x) = P[X = x] = (\frac{1}{2})^x$ (this is the probability of $x - 1$ succesive odd tosses followed by an even toss - the same as in Example 92 earlier in these notes). Then,

$$P[X \text{ is even}] = P[2] + P[4] + P[6] + \cdots = (\tfrac{1}{2})^2 + (\tfrac{1}{2})^4 + (\tfrac{1}{2})^6 + \cdots = \frac{(\frac{1}{2})^2}{1-(\frac{1}{2})^2} = \frac{1}{3}. \ \square$$

Example 18: The continuous random variable X has density function $f(x) = 3 - 48x^2$ for $-.25 \leq x \leq .25$ (and $f(x) = 0$ elsewhere). Find $P[\frac{1}{8} \leq X \leq \frac{5}{16}]$.

Solution: $P[.125 \leq X \leq .3125] = P[.125 \leq X \leq .25]$, since there is no density for X at points greater than .25. The probability is $\int_{.125}^{.25} (3 - 48x^2)\,dx = \frac{5}{32}$. $\square$

Example 19: Suppose that the continuous random variable X has the cumulative distribution function $F(x) = \frac{1}{1+e^{-x}}$ for $-\infty < x < \infty$. Find X's density function.

Solution: The density function for a continuous random variable is the first derivative of the cumulative distribution function. The density function of X is $f(x) = F'(x) = \frac{e^{-x}}{(1+e^{-x})^2}$. $\square$

Example 20: X is a random variable for which $P[X \leq x] = 1 - e^{-x}$ for $x \geq 1$, and $P[X \leq x] = 0$ for $x < 1$. Which of the following statements is true?

A) $P[X = 2] = 1 - e^{-2}$ and $P[X = 1] = 1 - e^{-1}$
B) $P[X = 2] = 1 - e^{-2}$ and $P[X \leq 1] = 1 - e^{-1}$
C) $P[X = 2] = 1 - e^{-2}$ and $P[X < 1] = 1 - e^{-1}$
D) $P[X < 2] = 1 - e^{-2}$ and $P[X < 1] = 1 - e^{-1}$
E) $P[X < 2] = 1 - e^{-2}$ and $P[X = 1] = 1 - e^{-1}$

Solution: Since $P[X \leq x] = 1 - e^{-x}$ for $x \geq 1$, it follows that $P[X \leq 1] = 1 - e^{-1}$. But $P[X \leq x] = 0$ if $x < 1$, and thus $P[X < 1] = 0$, so that $P[X = 1] = 1 - e^{-1}$ (since $P[X \leq 1] = P[X < 1] + P[X = 1]$). This eliminates answers C and D. Since the distribution function for X is continuous (and differentiable) for $x > 1$, it follows that $P[X = x] = 0$ for $x > 1$. This eliminates answers A, B and C.
This is an example of a random variable X with a mixed distribution - a point of probability at $X = 1$, and a continuous distribution for $X > 1$. $\square$

Example 21: A continuous random variable X has the density function

$$f(x) = \begin{cases} 2x\ , & 0 < x < \frac{1}{2} \\ \frac{4-2x}{3}\ , & \frac{1}{2} \le x < 2 \\ 0, & \text{elsewhere} \end{cases}$$

Find $P[.25 < X \le 1.25]$.

Solution: $P[.25 < X \le 1.25] = \int_{.25}^{1.25} f(x)\,dx = \int_{.25}^{.5} 2x\,dx + \int_{.5}^{1.25} \frac{4-2x}{3}\,dx = \frac{3}{4}$.
Note that since X is a continuous random variable, the probability $P[.25 \le X < 1.25]$ would be the same as $P[.25 < X \le 1.25]$. This is an example of a density function defined piecewise. Also, note that if the density function was defined to be

$$g(x) = \begin{cases} 2x\ , & 0 < x < \frac{1}{2} \\ 0\ , & x = 1/2 \\ \frac{4-2x}{3}, & \frac{1}{2} < x \le 2 \end{cases}$$

then all probabilities are unchanged (since the two density functions f and g differ at one point, probability calculations, which are based on integrals of the density function over an interval, are the same for both f and g). $\square$

EXPECTATION AND OTHER DISTRIBUTION PARAMETERS

Expected value of a random variable: For a random variable X, the expected value is denoted

$$E[X] \text{ or } \mu_X \text{ or } \mu.$$

For a discrete random variable, the expected value of X is

$$\boxed{E[X] = \sum x \cdot f(x)}$$

where the sum is taken over all points x at which X has non-zero probability. For instance, if is the result of one toss of a fair die, then

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

For a continuous random variable, the expected value is

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$

although the integral is written with lower limit $-\infty$ and upper limit ∞, the interval of integration is the interval of non-zero-density for f.

Note that f is the probability function in the discrete case, and f is the density function in the continuous case. The expected value of X is also called the **expectation of X**, or the **mean of X**. The expected value is the "average" over the range of values that X can be, or the "center" of the distribution.

Expectation of $h(x)$: If h is a function, then $E[h(X)]$ is equal to

$$\boxed{E[h(X)] = \sum_x h(x) \cdot f(x)} \quad \text{if } X \text{ is a discrete random variable,}$$

$$\boxed{E[h(X)] = \int_{-\infty}^{\infty} h(x) \cdot f(x)\,dx} \quad \text{if } X \text{ is a continuous random variable.}$$

Moments of a random variable: If $n \geq 1$ is an integer, then the ***n*-th moment of *X*** is $E[X^n]$.
If the mean of X is μ, then the ***n*-th central moment of *X* (about the mean *μ*)** is $E[(X-\mu)^n]$

Variance of *X*: The variance of X is denoted $Var[X]$, $V[X]$, σ_X^2 or σ^2. It is defined to be equal to

$$Var[X] = E[(X-\mu_X)^2] = E[X^2] - \mu_X^2 = E[X^2] - (E[X])^2$$

(the variance is the 2nd central moment of X about its mean). The variance is a measure of the "dispersion" of X about the mean - a large variance indicates significant levels of probability or density for points far from $E[X]$. The variance is always ≥ 0 (the variance of X is equal to 0 only if X has a discrete distribution with a single point and probability 1 at that point (not random at all).

Standard deviation of *X*: The standard deviation of the random variable X is the square root of the variance, and is denoted

$$\sigma_X = \sqrt{Var[X]}$$

Moment generating function of random variable *X*: The moment generating function of X (m.g.f.) is denoted $M_X(t)$, $m_X(t)$, $M(t)$ or $m(t)$, and it is defined to be

$$M_X(t) = E[e^{tX}] \text{ , which is either } \sum_x e^{tx} f(x) \quad \text{or} \quad \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

if X is discrete or continuous, respectively. It is always true that $M_X(0) = 1$. The moment generating function of X might not exist for all real numbers, but usually exists on some interval of real numbers. The function $ln[M_X(t)]$ is called the **cumulant generating function**.

Percentiles of a distribution: If $0 < p < 1$, then the $100p$-th percentile of the distribution of X is the number c_p which satisfies both of the following inequalities:

$$P[X \leq c_p] \geq p \qquad P[X \geq c_p] \geq 1-p$$

For a continuous random variable, it is sufficient to find the c_p for which $P[X \leq c_p] = p$.

If $p = .5$, **the 50-th percentile of a distribution is referred to as the median of the distribution.** It is the point M for which

$$\boxed{P[X \leq M] = .5 = P[X \geq M]} \text{ .}$$

The mode of a distribution: The mode is any point m at which the probability or density function $f(x)$ is maximized.

The skewness of a distribution: If the mean of random variable X is μ and the variance is σ^2 then the skewness is defined to be $E[(X-\mu)^3]/\sigma^3$.

Some results and formulas relating to this section:

(i) The mean of a random variable X might not exist, it might be $+\infty$ or $-\infty$, and the variance of X might be $+\infty$. For example, the continuous random variable X with p.d.f. $f(x) = \begin{cases} \frac{1}{x^2} & \text{for } x \geq 1 \\ 0, & \text{otherwise} \end{cases}$ has expected value $\int_1^\infty x \cdot \frac{1}{x^2}\, dx = +\infty$.

(ii) For any constants a_1, a_2 and b and functions h_1 and h_2,

$$E[a_1 h_1(X) + a_2 h_2(X) + b] = a_1 E[h_1(X)] + a_2 E[h_2(X)] + b$$

(iii) If a and b are constants, then ,

$$Var[aX + b] = a^2 Var[X].$$

(iv) If X is a random variable defined on the interval $[a, \infty)$ ($f(x) = 0$ for $x < a$), then $E[X] = a + \int_a^\infty [1 - F(x)]\, dx$, and if X is defined on the interval $[a, b]$, where $b < \infty$, then $E[X] = a + \int_a^b [1 - F(x)]\, dx$. This relationship is valid for any random variable, discrete, continuous or with a mixed distribution. As a special, case, if X is a non-negative random variable (defined on $[0, \infty)$ or $(0, \infty)$) then

$$E[X] = \int_0^\infty [1 - F(x)]\, dx$$

(v) **Chebyshev's inequality:** If X is a random va riable with mean μ_X and standard deviation σ , then for any real number $r > 0$,

$$P[|X - \mu \quad | > r] \leq \frac{\sigma_X^2}{r^2}.$$

(v) **Jensen's inequality:** If h is a function a nd X is a random variable such that $\frac{d^2}{dx^2} h(x) = h''(x) \geq 0$ at all points x with non-zero density or probability for X, then $E[h(X)] \geq h(E[X])$, and if $h'' > 0$ then

$$E[h(X)] \geq h(E[X]).$$

reverses if $h'' \leq 0$. For example, if $h(x) = x^2$, then $h''(x) = 2 \geq 0$ for any x, so that $E[X^2] \geq (E[X])^2$ (this is also true since $Var[X] = E[X^2] - (E[X])^2 \geq 0$ for any random variable X). As another example, if X is a positive random variable (i.e., X has non-zero density or probability only for $x \geq 0$), and $h(x) = \sqrt{x}$, then $h''(x) = \frac{-1}{4x^{3/2}} < 0$ for $x > 0$, and it follows from Jensen's inequality that $E[\sqrt{X}] < \sqrt{E[X]}$.

(vii) Suppose that for the random variable X, the moment generating function $M_X(t)$ exists in an interval containing the point $t = 0$. Then

$$\frac{d^n}{dt^n} M_X(t)\Big|_{t=0} = M_X^{(n)}(0) = E[X^n]$$

$$\frac{d}{dt} ln[M_X(t)]\Big|_{t=0} = \frac{M_X'(0)}{M_X(0)} = E[X], \text{ and } \frac{d^2}{dt^2} ln[M_X(t)]\Big|_{t=0} = Var[X].$$

The Taylor series expansion of $M_X(t)$ expanded about the point $t = 0$ is

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E[X^k] = 1 + t \cdot E[X] + \frac{t^2}{2} \cdot E[X^2] + \frac{t^3}{6} \cdot E[X^3] + \cdots$$

If X_1 and X_2 are random variables, and $M_{X_1}(t) = M_{X_2}(t)$ for all values of t in an interval containing $t = 0$, then X_1 and X_2 have identical probability distributions.

(viii) The median (50th percentile) and other percentiles of a distribution are not always unique. For example, if X is the discrete random variable with probability function $f(x) = .25$ for $x = 1, 2, 3, 4$, then the median of X would be any point from 2 to 3, but the usual convention is to set the median to be the midpoint between the two "middle" values of X, $M = 2.5$.

(ix) The distribution of the random variable X is said to be **symmetric about the point *c*** if $f(c+t) = f(c-t)$ for any value of t. It follows that the expected value of X and the median of X is c. Also, for a symmetric distribution, any odd-order central moments about the mean are 0, i.e. $E[(X-\mu)^k] = 0$ if $k \geq 1$ is an odd integer.

(x) If $E[X] = \mu$, $Var[X] = \sigma^2$ and Z is defined to be $Z = \frac{X-\mu}{\sigma}$, then

$$E[Z] = 0 \text{ and } Var[Z] = 1 .$$

Example 22: Let X equal the number of tosses of a fair die until the first "1" appears. Find $E[X]$.

Solution: X is a discrete random variable that can take on an integer value ≥ 1. The probability that the first 1 appears on the x-th toss is $f(x) = (\frac{5}{6})^{x-1}(\frac{1}{6})$ for $x \geq 1$ ($x-1$ tosses that are not 1 followed by a 1). This is the probability function of X. Then

$$E[X] = \sum_{k=1}^{\infty} k \cdot f(k) = \sum_{k=1}^{\infty} k \cdot (\tfrac{5}{6})^{k-1}(\tfrac{1}{6}) = (\tfrac{1}{6})[1 + 2(\tfrac{5}{6}) + 3(\tfrac{5}{6})^2 + \cdots] .$$

We use the general increasing geometric series relation $1 + 2r + 3r^2 + \cdots = \frac{1}{(1-r)^2}$,

so that $E[X] = (\frac{1}{6}) \cdot \frac{1}{(1-\frac{5}{6})^2} = 6$. □

Example 23: Given that the density function of X is $f(x : \theta) = \theta e^{-x\theta}$, for $x > 0$, and 0 elsewhere, find the n-th moment of X, where n is a non-negative integer (assuming that $\theta > 0$).

Solution: The n-th moment of X is $E[X^n] = \int_0^\infty x^n \cdot \theta e^{-x\theta} dx$. Applying integration by parts, this can be written as

$$\int_0^\infty x^n \, d(-e^{-x\theta}) = -x^n e^{-x\theta}\Big|_{x=0}^{x=\infty} - \int_0^\infty -nx^{n-1} e^{-x\theta}\,dx = \int_0^\infty nx^{n-1} e^{-x\theta} dx .$$

Repeatedly applying integration by parts results in $E[X^n] = \frac{n!}{\theta^n}$. It is worthwhile noting the general form of the integral that appears in this example - if $k \geq 0$ is an integer and $a > 0$, then by repeated applications of integration by parts, we have $\int_0^\infty t^k \, e^{-at} \, dt = \frac{k!}{a^{k+1}}$, so that in this example $\int_0^\infty x^n \theta e^{-x\theta} dx = \theta \int_0^\infty x^n e^{-x\theta} dx = \theta \cdot \frac{n!}{\theta^{n+1}} = \frac{n!}{\theta^n}$. □

Example 24: A fair die is tossed until the first 1 appears. Let x equal the number of tosses required, $x = 1, 2, 3, ...$ You are to receive $(.5)^x$ dollars if the 1 appears on the x-th toss. What is the expected amount that you will receive?

Solution: This is the same distribution as in Example 101 above, with the probability that the first 1 appears on the x-th toss being $(\frac{5}{6})^{x-1}(\frac{1}{6})$ for $x \geq 1$ ($x - 1$ tosses that are not 1 followed by a 1), and the amount received in that case is $h(x) = (.5)^x$. Then, the expected amount received is

$$E[h(X)] = E[(.5)^X] = \sum_{k=1}^{\infty} (.5)^k \cdot (\tfrac{5}{6})^{k-1}(\tfrac{1}{6}) = (\tfrac{1}{12})[1 + (\tfrac{5}{12}) + (\tfrac{5}{12})^2 + \cdots] = \tfrac{1}{7}. \square$$

Example 25: A continuous random variable X has density function

$$f(x) = \begin{cases} 1-|x| \text{ if } |x|<1 \\ 0, \text{ elsewhere} \end{cases}$$

Find $Var[X]$.

Solution: The density of X is symmetric about 0 (since $f(x) = f(-x)$), so that $E[X] = 0$. (this can be verified directly

$$E[X] = \int_{-1}^{1} x(1 - |x|)\,dx = \int_{-1}^{0} x(1 + x)\,dx + \int_{0}^{1} x(1 - x)\,dx = -\tfrac{1}{6} + \tfrac{1}{6} = 0).$$

Then,

$$Var[X] = E[X^2] - (E[X])^2 = E[X^2] = \int_{-1}^{1} x^2(1 - |x|)\,dx$$
$$= \int_{-1}^{0} x^2(1 + x)\,dx + \int_{0}^{1} x^2(1 - x)\,dx = \tfrac{1}{6}. \quad \square$$

Example 26: The moment generating function of X is $\frac{\alpha}{\alpha - t}$ for $t < \alpha$, where $\alpha > 0$. Find $Var[X]$.

Solution: $Var[X] = E[X^2] - (E[X])^2$. $E[X] = M_X'(0) = \frac{\alpha}{(\alpha-t)^2}\Big|_{t=0} = \frac{1}{\alpha}$, and

$$E[X^2] = M_X''(0) = \frac{2\alpha}{(\alpha-t)^3}\Big|_{t=0} = \frac{2}{\alpha^2} \rightarrow Var[X] = \frac{2}{\alpha^2} - (\frac{1}{\alpha})^2 = \frac{1}{\alpha^2}$$

Alternatively,

$$ln\, M_X(t) = ln(\frac{\alpha}{\alpha-t}) = ln\,\alpha - ln(\alpha - t) \rightarrow \frac{d}{dt}\, ln[M_X(t)] = \frac{1}{\alpha-t}$$

and

$$\frac{d^2}{dt^2}\, ln[M_X(t)] = \frac{1}{(\alpha-t)^2} \text{ so that } Var[X] = \frac{d^2}{dt^2}\, ln[M_X(t)]\Big|_{t=0} = \frac{1}{\alpha^2}. \quad \square$$

Example 27: The continuous random variable X has p.d.f. $f(x) = \frac{1}{2} \cdot e^{-|x|}$ for $-\infty < x < \infty$. Find the 87.5-th percentile of the distribution.

Solution: The 87.5-th percentile is the number b for which

$$.875 = P[X \leq b] = \int_{-\infty}^{b} f(x)\,dx = \int_{-\infty}^{b} \tfrac{1}{2} \cdot e^{-|x|}\,dx.$$

Note that this distribution is symmetric about 0, since $f(-x) = f(x)$, so the mean and median are both 0. Thus, $b > 0$, and so

$$\int_{-\infty}^{b} \tfrac{1}{2} \cdot e^{-|x|}\,dx = \int_{-\infty}^{0} \tfrac{1}{2} \cdot e^{-|x|}\,dx + \int_{0}^{b} \tfrac{1}{2} \cdot e^{-|x|}\,dx = .5 + \int_{0}^{b} \tfrac{1}{2} \cdot e^{-x}\,dx$$
$$= .5 + \tfrac{1}{2}(1 - e^{-b}) = .875 \rightarrow b = -ln(.25) = ln\,4. \quad \square$$

FREQUENTLY USED DISCRETE DISTRIBUTIONS

Uniform distribution on N points: We denote $X \sim$ *Uniform (1/N).* The probability function is

$$f(x) = \frac{1}{N} \text{ for } x = 1, 2, \ldots, N \text{, and } f(x) = 0 \text{ otherwise.}$$

$$E[X] = \frac{N+1}{2}\ ,\ Var[X] = \frac{N^2-1}{12}\ ,\ M_X(t) = \sum_{j=1}^{N} \frac{e^{jt}}{N} = \frac{e^t(e^{Nt}-1)}{N(e^t-1)} \text{ for any real } t.$$

Binomial distribution with parameters n and p ($n \geq 1$ integer and $0 \leq p \leq 1$):
A single trial of an experiment results in either success with probability p, or failure with probability $1 - p = q$. If n independent trials of the experiment are performed, and X is the number of successes that occur, then X is an integer between 0 and n. X is said to have a binomial distribution with parameters n and p (sometimes denoted $X \sim B(n, p)$).

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n\ ,$$

$$E[X] = np\ ,\ Var[X] = np(1-p) \quad M_X(t) = (1 - p + pe^t)^n\ .$$

In the special case of $n = 1$ (a single trial), the distribution is referred to as a **Bernoulli distribution**. If $X \sim B(n, p)$, then X is the sum of n independent Bernoulli random variables each with distribution $B(1, p)$.

Poisson distribution with parameter $\lambda > 0$: We denote $X \sim \mathcal{P}(\lambda)$ if the probability function of X is

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, 3, ...,$$
$$E[X] = Var[X] = \lambda\ ,\ M_X(t) = e^{\lambda(e^t-1)}$$

The Poisson distribution is often used as a model for counting the number of events of a certain type that occur in a certain period of time.

Suppose that X represents the number of customers arriving for service at bank in a 1 hour period, and that a model for X is the Poisson distribution with parameter λ. Under some reasonable assumptions (such as independence of the numbers arriving in different time intervals) it is possible to show that the number arriving in any time period also has a Poisson distribution with the appropriate parameter that is "scaled" from λ. Suppose that $\lambda = 40$ - meaning that X, the number of bank customers arriving in one hour, has a mean of 40. If Y represents the number of customers arriving in 2 hours, then Y has a Poisson distribution with a parameter of 80 - for any time interval of length t, the number of customers arriving in that time interval has a Poisson distribution with parameter $\lambda t = 40t$ - so the number of customers arriving during a 15-minute period ($t = \frac{1}{4}$ hour) will have a Poisson distribution with parameter $40 \cdot \frac{1}{4} = 10$.

Geometric distribution with parameter p ($0 \le p \le 1$): A single trial of an experiment results in either success with probability p, or failure with probability $1 - p = q$.

The experiment is performed with successi ve independent trials until the first success occurs. If X represents the number of failures until the first success, then X is a discrete random variable that can be $0, 1, 2, 3, \ldots$. X is said to have a geometric distribution with parameter p.

$$f(x) = (1-p)^x p \text{ for } x = 0, 1, 2, 3, \ldots,$$

$$E[X] = \frac{1-p}{p} = \frac{q}{p}, \quad Var[X] = \frac{1-p}{p^2} = \frac{q}{p^2}, \quad M_X(t) = \frac{p}{1-(1-p)e^t}.$$

The geometric distribution has the **lack of memory property:**

$$P[X = n + k | X \ge n] = P[X = k].$$

Another version of a geometric distribution is the random variable Y, the number of the experiment on which the first success occurs; $Y = X + 1$ and $P[Y = y] = P[X = y - 1]$.

Negative binomial distribution with parameters r and p ($r > 0$ and $0 < p \le 1$):

$$f(x) = \binom{r+x-1}{x} p^r (1-p)^x, \text{ for } x = 0, 1, 2, 3, \ldots,$$

$$E[X] = \frac{r(1-p)}{p}, \quad Var[X] = \frac{r(1-p)}{p^2}, \quad M_X(t) = \left[\frac{p}{1-(1-p)e^t}\right]^r$$

If r is an integer, then the negative binomial random variable X can be interpreted as being the number of failures until the r-th success occurs when successive trials of an experiment are performed for which the probability of success in a single particular trial is p (the distribution is defined even if r is not an integer).

The notation q is sometimes used to represent $1 - p$. The geometric distribution is a special case of the negative binomial with $r = 1$.

Multinomial distribution: This distribution is discussed later in these notes.

Hypergeometric distribution with integer parameters M, K and $0 \le K \le M$ and $1 \le n \le M$: In a group of M objects, K are of Type I and $M - K$ are of Type II.

If n objects are randomly chosen without replacement from the group of M, let

X denote the number that are of Type I in the group of n.

X is a non-negative integer that satisfies $X \le n$, $X \le K$, $0 \le X$ and $n - (M - K) \le X$.

X has a hypergeometric distribution: $X \sim H(M, n, K)$. The probability function is

$$f(x) = \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}}, \quad \text{for } max[0, n-(M-K)] \le x \le min[n, K]$$

(there are $\binom{M}{n}$ ways of choosing the objects from the group of M, and the number of choices that result in x objects of Type I and $n - x$ objects of type II is $\binom{K}{x}\binom{M-K}{n-x}$), and

$$E[X] = \frac{nK}{M}, \quad Var[X] = \frac{nK(M-K)(M-n)}{M^2 \cdot (M-1)}.$$

Recursive relationship for the binomial, Poisson and negative binomial distributions: The probability function for each these three distributions satisfies the following recursive relationship $\frac{p_k}{p_{k-1}} = a + \frac{b}{k}$ for $k = 1, 2, 3, \ldots$.

Poisson with parameter λ: $\frac{p_k}{p_{k-1}} = \frac{e^{-\lambda}\lambda^k/k!}{e^{-\lambda}\lambda^{k-1}/(k-1)!} = \frac{\lambda}{k} \rightarrow a = 0$, $b \Rightarrow \lambda$.

Binomial with parameters n and p: $a = -\frac{p}{1-p}$, $b = \frac{(n+1)p}{1-p}$.

Negative binomial with parameters r and p: $a = 1 - p$, $b = (r-1)(1-p)$.

Example 28: X is a discrete random variable that is uniformly distributed on the even integers $x = 0, 2, 4, \ldots, 22$, so that the probability function of X is $f(x) = \frac{1}{12}$ for each even integer x from 0 to 22. Find $E[X]$ and $Var[X]$.

Solution: The discrete uniform distribution described earlier in the notes is on the points $x = 1, 2, \ldots, N$. If we consider the transformation $Y = \frac{X+2}{2}$, then the random variable Y is distributed on the points $Y = 1, 2, \ldots, 12$, with probability function $f_Y(y) = \frac{1}{12}$ for each integer y from 1 to 12. Thus, Y has the discrete uniform distribution described earlier in the notes, and

$$E[Y] = \frac{12+1}{2} = \frac{13}{2} , \qquad Var[Y] = \frac{12^2-1}{12} = \frac{143}{12} .$$

But since $Y = \frac{X+2}{2}$, we can use rules for expectation and variance to get

$$E[Y] = \frac{E[X]+2}{2} \qquad E[X] = 2 \cdot E[Y] - 2 = 11$$

and

$$Var[Y] = Var[\frac{X+2}{2}] = \frac{1}{4} \cdot Var[X] \text{ , so that } Var[X] = 4 \cdot Var[Y] = \frac{143}{3} . \quad \square$$

Example 29: If X is the number of "6"'s that turn up when 72 ordinary dice are independently thrown, find the expected value of X^2.

Solution: X has a binomial distribution with $n = 72$ and $p = \frac{1}{6}$. Then $E[X] = np = 12$, and $Var[X] = np(1-p) = 10$. But $Var[X] = E[X^2] - (E[X])^2$, so that $E[X^2] = 10 + 12^2 = 154$. $\square$

Example 30: The number of hits, X, per baseball game, has a Poisson distribution. If the probability of a no-hit game is $\frac{1}{10{,}000}$, find the probability of having 4 or more hits in a particular game.

Solution: $P[X=0] = \frac{e^{-\lambda} \cdot \lambda^0}{0!} = e^{-\lambda} = \frac{1}{10{,}000} \rightarrow \lambda = ln\, 10{,}000$

$$P[X \geq 4] = 1 - (P[X=0] + P[X=1] + P[X=2] + P[X=3])$$

$$= 1 - (\frac{e^{-\lambda} \cdot \lambda^0}{0!} + \frac{e^{-\lambda} \cdot \lambda^1}{1!} + \frac{e^{-\lambda} \cdot \lambda^2}{2!} + \frac{e^{-\lambda} \cdot \lambda^3}{3!})$$

$$= 1 - (\frac{1}{10{,}000} + \frac{ln 10{,}000}{10{,}000} + \frac{(ln 10{,}000)^2}{2(10{,}000)} + \frac{(ln 10{,}000)^3}{6(10{,}000)}) = .9817. \quad \square$$

Example 31: In rolling a fair die repeatedly (and independently on successive rolls), find the probability of getting the third "1" on the t-th roll.

Solution: The negative binomial random variable X with parameters $r = 3$ and $p = \frac{1}{6}$ is the number of failures (rolling 2,3,4,5 or 6) until the 3rd success. The probability that the 3rd success (3rd "1") occurs on the t-th roll is the same as the probability of $x = t - 3$ failures before the 3rd success. Thus, with $X = T - 3$, X is a negative binomial random variable with parameters $r = 3$ and $p = \frac{1}{6}$. The probability is

$$P[X = t - 3] = \binom{3+t-3-1}{t-3}(\tfrac{1}{6})^3(\tfrac{5}{6})^{t-3} = \binom{t-1}{t-3}(\tfrac{1}{6})^3(\tfrac{5}{6})^{t-3}$$

$$= \binom{t-1}{2}(\tfrac{1}{6})^3(\tfrac{5}{6})^{t-3}$$ (the final equality follows from $\binom{n}{k} = \binom{n}{n-k}$). □

Example 32: An urn contains 6 blue and 4 red balls. 6 balls are chosen at random and without replacement from the urn. If X is the number of red balls chosen, find the standard deviation of X.

Solution: This is a hypergeometric distribution with $M = 10$, $K = 4$ and $n = 6$.
The probability function of X is

$$f(x) = \frac{\binom{4}{x}\binom{6}{6-x}}{\binom{10}{6}}, \text{ for } x = 0, 1, 2, 3, 4 .$$

The variance is $Var[X] = \frac{nK(M-K)(M-n)}{M^2 \cdot (M-1)} = .64$. Standard deviation is $\sqrt{.64} = .8$. □

FREQUENTLY USED CONTINUOUS DISTRIBUTIONS

Uniform distribution on the interval (a,b) (where $-\infty < a < b < \infty$): The p.d.f. is

$f(x) = \frac{1}{b-a}$ for $a < x < b$, and $f(x) = 0$ otherwise.

$E[X] = \frac{a+b}{2}$, $Var[X] = \frac{(b-a)^2}{12}$, $M_X(t) = \frac{e^{bt}-e^{at}}{(b-a)\cdot t}$ for any real t ,

$E[X^n] = \frac{b^{n+1}-a^{n+1}}{(n+a)(b-a)}$.

This is a symmetric distribution about the mean, median $= \frac{b-a}{2}$.

Normal distribution with mean μ and variance σ^2 ($-\infty < \mu < \infty$, $\sigma^2 > 0$):

$$f(x) = \frac{1}{\sigma\cdot\sqrt{2\pi}} \cdot e^{-(x-\mu)^2/2\sigma^2} \text{ for } -\infty < x < \infty,$$

$$E[X] = \mu,\ Var[X] = \sigma^2,\quad M_X(t) = exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right].$$

If μ =0 and σ^2=1 , the distribution is referred to as a **standard normal distribution.** In this case, $F(x)$ is sometimes denoted $\phi(x)$.

Tables of values of $\Phi(x)$ may be found in most statistics textbooks and are provided with the exam. provided with the exam. If $0 < \alpha < 1$, the notation z_α refers to the point in the standard normal distribution Z such that

$P[Z > z_\alpha] = \alpha$ (z_α is the $100(1-\alpha)$ percentile of the standard normal distribution).

normal distribution). $X \sim N(\mu,\sigma^2)$ is used as notation describing X as a normal randomally distributed variable variable with mean μ and variance σ^2. An important rule concerning the normal distribution is that

If $X \sim N(\mu,\sigma^2)$, then $Y = \frac{X-\mu}{\sigma}$ $N(0,1)$

The normal distribution is symmetric ($f(x)$ is "bell-shaped", peaking at $x = \mu$) with

mean $=$ median $=$ mode $= \mu$.

From the standard normal table, it can be seen that $\Phi(1) = P[Z \leq 1] = .8413$. Because of the symmetry about 0 of the standard normal distribution it follows that

$$1 - \Phi(-1) = P[Z \geq -1] = .8413$$

and then

$$\Phi(-1) = P[Z < -1] = 1 - P[Z \geq -1] = 1 - .8413 = .1587 \cdot$$

In general, for $a > 0$, $\Phi(-a) = 1 - \Phi(a)$. Given any normal random variable $X \sim N(\mu,\sigma^2)$, it is possible to find $P[r < X < s]$ by first "standardizing": $Z = \frac{X-\mu}{\sigma}$

$$P[r < X < s] = P[\tfrac{r-\mu}{\sigma} < \tfrac{X-\mu}{\sigma} < \tfrac{s-\mu}{\sigma}] = \Phi(\tfrac{s-\mu}{\sigma}) - \Phi(\tfrac{r-\mu}{\sigma})$$

Approximating a distribution using a normal distribution: Given a random variable with mean μ and variance σ^2, probabilities related to the distribution of X are sometimes approximated by assuming the distribution of X is approximately $N(\mu,\sigma^2)$. If X is discrete and integer-valued then an "integer correction" is applied; the probability $P[n \leq X \leq m]$ is approximated by assuming that X is normal and then finding the probability $P[n - \frac{1}{2} \leq X \leq m + \frac{1}{2}]$.

Exponential distribution with mean $\frac{1}{\lambda} > 0$:

$$f(x) = \lambda e^{-\lambda x} \text{ for } x > 0, \text{ and } f(x) = 0 \text{ otherwise,}$$

$$F(x) = 1 - e^{-\lambda x} \text{ for } x \geq 0, \text{ and } P[X > x] = e^{-\lambda x},$$

$$E[X] = \frac{1}{\lambda}, \; Var[X] = \frac{1}{\lambda^2}, \quad E[X^k] = \int_0^\infty x^k \cdot \lambda e^{-\lambda x} dx = \frac{k!}{\lambda^k},$$

$$M_X(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda .$$

(note that an exponential distribution with mean μ has p.d.f. $f(x) = \frac{1}{\mu} e^{-x/\mu}$).

There are a few important properties satisfied by the exponential distribution:

(i) **lack of memory property**- for $x, y > 0$, $P[X > x + y | X > x] = P[X > y]$.

(ii) **link between the exponential distribution and Poisson distribution** - Suppose that X has an exponential distribution with mean $\frac{1}{\lambda}$ and we regard X as the time between successive occurrences of some type of event (say the event is the arrival of a new insurance claim at an insurance office), where time is measured in some appropriate units (second, minutes, hours or days, etc.). Now, we imagine that we choose some starting time (say labeled as $t = 0$), and from now we start recording times between successive events. Let N represent the number of events (claims) that have occurred when one unit of time has elapsed. Then N will be a random variable related to the times of the occurring events. The distribution of N is Poisson with parameter λ.

(iii) **the minimum of a collection of independent exponential random variables:**

Suppose that independent random variables $Y_1, Y_2, ..., Y_n$ have exponential distributions with means $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, ..., \frac{1}{\lambda_n}$ (parameters $\lambda_1, \lambda_2, ..., \lambda_n$) respectively.

Let $Y = min\{Y_1, Y_2, \ldots, Y_n\}$. Then Y has an exponential distribution with mean

$$\frac{1}{\lambda_1 + \lambda_2 + \cdots + \lambda_n} .$$

(iv) **a "mixture" of distributions:** Given any finite collection of independent random variables, $X_1, X_2, \ldots, X_k$ with density or probability functions, say $f_1(x), f_2(x), \ldots f_k(x)$, where k is a non-negative integer, and given a set of "weights", $\alpha_1, \alpha_2, \ldots, \alpha_k$, where $0 \leq \alpha_i \leq 1$ for each i and $\sum_{i=1}^{k} \alpha_i = 1$, it is possible to construct the density function $f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \cdots + \alpha_k f_k(x)$, which is a "weighted average" of the original density functions. It then follows that the resulting distribution X, whose density/probability function is f, has moments and moment generating function which are weighted averages of the original distribution moments and moment generating functions:

$$E[X^n] = \alpha_1 E[X_1^n] + \alpha_2 E[X_2^n] + \cdots + \alpha_k E[X_k^n]$$

$$M_X(t) = \alpha_1 M_{X_1}(t) + \alpha_2 M_{X_2}(t) + \cdots + \alpha_k M_{X_k}(t) .$$

One common application of this is seen in a distribution which is a "mixture of exponentials" - suppose that the continuous random variable X has density function

$$f(x) = e^{-3x} + 2e^{-4x} + \frac{1}{12}e^{-x/2}$$

After some consideration, it might be noticed that

$$f(x) = \frac{1}{3} \cdot 3e^{-3x} + \frac{1}{2} \cdot 4e^{-4x} + \frac{1}{6} \cdot \frac{1}{2}e^{-x/2} = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x)$$

where

$$\alpha_1 = \frac{1}{3}\ ,\ \alpha_2 = \frac{1}{2}\ ,\quad \alpha_3 = \frac{1}{6} \text{ and } f_1(x) = 3e^{-3x},\ f_2(x) = 4e^{-4x}\ ,\ f_3(x) = \frac{1}{2}e^{-x/2}.$$

Thus, X is a weighted average of three exponential distributions with parameters 3, 4 and 1/2, and weights $\frac{1}{3}$ $\frac{1}{2}$ and $\frac{1}{6}$. Then,

$$E[X] = \frac{1}{3} \cdot E[X_1] + \frac{1}{2} \cdot E[X_2] + \frac{1}{6} \cdot E[X_3] = \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{6} \cdot \frac{1}{1/2} = \frac{41}{72}\ ,$$

$$E[X^2] = \frac{1}{3} \cdot E[X_1^2] + \frac{1}{2} \cdot E[X_2^2] + \frac{1}{6} \cdot E[X_3^2] = \frac{1}{3} \cdot \frac{2}{9} + \frac{1}{2} \cdot \frac{2}{16} + \frac{1}{6} \cdot \frac{2}{1/4} = \frac{205}{144}\ ,$$

$$Var[X] = E[X^2] - (E[X])^2 = \frac{205}{144} - (\frac{41}{72})^2$$

$$M_X(t) = \frac{1}{3}M_{X_1}(t) + \frac{1}{2}M_{X_2}(t) + \frac{1}{6}M_{X_3}(t) = \frac{1}{3} \cdot \frac{3}{3-t} + \frac{1}{2} \cdot \frac{4}{4-t} + \frac{1}{6} \cdot \frac{\frac{1}{2}}{\frac{1}{2}-t}\ .$$

Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$:

$$f(x) = \frac{\beta^\alpha \cdot x^{\alpha-1} \cdot e^{-\beta x}}{\Gamma(\alpha)} \text{ for } x > 0\ , \text{ and} f(x) = 0 \text{ otherwise.}$$

$\Gamma(\alpha)$ is the **gamma function**, which is defined for $\alpha > 0$ to be $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} \cdot e^{-y}dy$ from which it follows that if n is a positive integer, $\Gamma(n) = (n-1)!$.

$$E[X] = \frac{\alpha}{\beta}\ ,\ Var[X] = \frac{\alpha}{\beta^2}\ ,\ M_X(t) = (\frac{\beta}{\beta-t})^\alpha \text{ for } 0 < t < \beta\ .$$

The exponential distribution with parameter λ is a special case of the gamma distribution with $\alpha = 1$ and $\beta = \lambda$.

Pareto distribution with parameters α, $x_0 > 0$: The Pareto distribution with parameters $\alpha > 0$ and $x_0 > 0$ has p.d.f.

$$f(x) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}}\ ,\ x > x_0\ ,\quad f(x) = 0 \quad \text{otherwise.}$$

$$E[X] = \frac{\alpha x_0}{\alpha-1} \qquad Var[X] = \frac{\alpha x_0^2}{(\alpha-2)(\alpha-1)^2}$$

Lognormal distribution with parameters m and $\sigma^2 > 0$ ($-\infty < m < \infty$):

If $W \sim N(m,\sigma^2)$, then $X = e^W$ has a lognormal distribution with parameters m and σ^2 (the *log* of X has a normal distribution $N(m,\sigma^2)$). The p.d.f. of X is

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}exp[-(log(x) - m)^2/2\sigma^2] \text{ for } x > 0 \text{ and } f(x) = 0 \text{ otherwise.}$$

$$E[X] = e^{m+\frac{1}{2}\sigma^2}\ ,\ Var[X] = (e^{\sigma^2} - 1)e^{2m+\sigma^2}\ .$$

Beta distribution with parameters $a > 0$ and $b > 0$: The **beta function** is defined to be $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$. The beta distribution with parameters $a, b > 0$ has p.d.f.

$$f(x) = \frac{1}{B(a,b)} \cdot x^{a-1}(1-x)^{b-1} \text{ for } 0 < x < 1 \text{, and } f(x) = 0 \text{ otherwise.}$$

$$E[X] = \frac{a}{a+b} \; , \; Var[X] = \frac{ab}{(a+b)^2(a+b+1)} \; .$$

The gamma function was defined earlier in the context of the gamma distribution. The beta function can be expressed in terms of the gamma function:

$$B(a,b) = \frac{\Gamma(a)\cdot\Gamma(b)}{\Gamma(a+b)} \text{ , and if } a \text{ and } b \text{ are integers } > 0 \text{, then } B(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!} \; .$$

Example 33: Suppose that X has a uniform distribution on the interval $(0, a)$, where $a > 0$. Find $P[X > X^2]$.

Solution: If $a \leq 1$, then $X > X^2$ is always true, so that $P[X > X^2] = 1$.
If $a > 1$, then $X > X^2$ only if $X < 1$, which has probability

$$P[X < 1] = \int_0^1 f(x)\,dx = \int_0^1 \frac{1}{a}\,dx = \frac{1}{a} \text{ . Thus, } P[X > X^2] = min[1, \frac{1}{a}] \text{ .} \quad \square$$

Example 34: The random variable T has an exponential distribution such that $P[T \leq 2] = 2 \cdot P[T > 4]$. Find $Var[T]$.

Solution: Suppose that T has mean $\frac{1}{\lambda}$. $P[T \leq 2] = 1 - e^{-2\lambda} = 2P[T > 4] = 2e^{-4\lambda}$

$$\rightarrow 2x^2 + x - 1 = 0 \text{ , where } x = e^{-2\lambda}$$

Solving the quadratic equation results in $x = \frac{1}{2}, -1$.

$$e^{-2\lambda} = \frac{1}{2} \text{ and } \lambda = \frac{1}{2}\,ln\,2 \,.$$

Then, $Var[T] = \frac{1}{\lambda^2} = \frac{4}{(ln\,2)^2}$. $\square$

Example 35: If for a certain normal random variable X, $P[X < 500] = .5$ and $P[X > 650] = .0227$, find the standard deviation of X.

Solution: The normal distribution is symmetric about its mean, with $P[X < \mu] = .5$ for any normal random variable. Thus, for this normal X, $\mu = 500$. Then,

$$P[X > 650] = .0227 = P\left[\frac{X-500}{\sigma} > \frac{150}{\sigma}\right]$$

Since $\frac{X-500}{\sigma}$ has a standard normal distribution, it follows from the table for the standard normal distribution that $\frac{150}{\sigma} = 2.00$ and $\sigma = 75$. $\square$

Example 36: Verify algebraically the validity of properties (i) (lack of memory) and (iii) (minimum of a collection of independent exponential distributions) of the exponential distribution listed earlier in these notes.

Solution: (i) Suppose that X has an exponential distribution with parameter λ. Then

$$P[X > x + y|X > x] = \frac{P[(X>x+y)\cap(X>x)]}{P[X>x]} = \frac{P[X>x+y]}{P[X>x]} = \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} = e^{-\lambda y},$$

and $P[X > y] = e^{-\lambda y}$.

(iii) Suppose that independent random variables $Y_1, Y_2, ..., Y_n$ have exponential distributions with means $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, ..., \frac{1}{\lambda_n}$ (parameters $\lambda_1, \lambda_2, ..., \lambda_n$) respectively.

Let $Y = min\{Y_1, Y_2, \ldots, Y_n\}$. Then,

$$\begin{aligned} P[Y > y] &= P[Y_i > y \text{ for all } i = 1, 2, ..., n] \\ &= P[(Y_1 > y) \cap (Y_2 > y) \cap \cdots \cap (Y_n > y)] \\ &= P[Y_1 > y] \cdot P[Y_2 > y] \cdots P[Y_n > y] \text{ (because of independence of the } Y_i\text{'s)} \\ &= (e^{-\lambda_1 y})(e^{-\lambda_2 y}) \cdots (e^{-\lambda_n y}) = e^{-(\lambda_1+\lambda_2+\cdots+\lambda_n)y} . \end{aligned}$$

The c.d.f of Y is then

$$F_Y(y) = P[Y \le y] = 1 - P[Y > y] = 1 - e^{-(\lambda_1+\lambda_2+\cdots+\lambda_n)y}$$

and the p.d.f. of Y is

$$f_Y(y) = F_Y'(y) = (\lambda_1 + \lambda_2 + \cdots + \lambda_n)e^{-(\lambda_1+\lambda_2+\cdots+\lambda_n)y}$$

which is the p.d.f. of an exponential distribution with parameter $\lambda_1 + \lambda_2 + \cdots + \lambda_n$. □

Example 37: Suppose that X has a binomial distribution based on 100 trials with a probability of success of .2 on any given trial. Find the approximate probability $P[15 \le X \le 25]$.

Solution: The mean and variance of X are $E[X] = 100(.2) = 20$, $Var[X] = 100(.2)(.8) = 16.$ Using the normal approximation with integer correction, we assume that X is approximately normal and find

$$P[14.5 \le X \le 25.5] = P[\frac{14.5-20}{\sqrt{16}} \le \frac{X-20}{\sqrt{16}} \le \frac{25.5-20}{\sqrt{16}}] = P[-1.375 \le Z \le 1.375] ,$$

where Z has a standard normal distribution.

$$P[-1.375 \le Z \le 1.375] = \Phi(1.375) - \Phi(-1.375) = \Phi(1.375) - [1 - \Phi(1.375)] = 2\Phi(1.375) - 1 .$$

From the standard normal table we have $\Phi(1.3) = .9032$ and $\Phi(1.4) = .9192$. Using linear interpolation (since 1.375 is $\frac{3}{4}$ of the way from 1.3 to 1.4) we have

$$\Phi(1.375) = (.25)\Phi(1.3) + (.75)\Phi(1.4) = .9152$$

and then the probability in question is $2(.9152) - 1 = .8304$. □

Topic 3: Multivariate Random Variables

JOINT, MARGINAL, AND CONDITIONAL DISTRIBUTIONS

Joint distribution of random variables X and Y: A joint distribution of two random variables has a probability function or probability density function $f(x,y)$ that is a function of two variables (sometimes denoted $f_{X,Y}(x,y)$).

If X and Y are discrete random variables, then $f(x,y)$ must satisfy

$$\text{(i) } 0 \le f(x,y) \le 1 \quad \text{and} \quad \text{(ii) } \sum_x \sum_y f(x,y) = 1 \, .$$

If X and Y are continuous random variables, then $f(x,y)$ must satisfy

$$\text{(i) } f(x,y) \ge 0 \quad \text{and} \quad \text{(ii) } \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dy\, dx = 1.$$

It is possible to have a joint distribution in which one variable is discrete and one is continuous, or either has a mixed distribution. The joint distribution of two random variables can be extended to a joint distribution of any number of random variables.
If A is a subset of two-dimensional space, then $P[(X,Y) \in A]$ is the double summation (discrete case) or double integral (continuous case) of $f(x,y)$ over the region A.

Cumulative distribution function of a joint distribution: If random variables X and Y have a joint distribution, then the cumulative distribution function is

$$F(x,y) = P[(X \le x) \cap (Y \le y)] \, .$$

In the continuous case, $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\, dt\, ds$,

and in the discrete case, $F(x,y) = \sum_{s=-\infty}^{x} \sum_{t=-\infty}^{y} f(s,t)$.

In the continuous case, $\frac{\partial^2}{\partial x\, \partial y} F(x,y) = f(x,y)$.

Expectation of a function of jointly distributed random variables: If $h(x,y)$ is a function of two variables, and X and Y are jointly distributed random variables, then the **expected value of $h(X,Y)$** is defined to be

$$E[h(X,Y)] = \sum_x \sum_y h(x,y) \cdot f(x,y) \text{ in the discrete case, and}$$

$$E[h(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) \cdot f(x,y)\, dy\, dx \text{ in the continuous case.}$$

Marginal distribution of X found from a joint distribution of X and Y:
If X and Y have a joint distribution with joint density or probability function $f(x,y)$, then the **marginal distribution of X** has a probability function or density function denoted $f_X(x)$, which is equal to

$$f_X(x) = \sum_y f(x,y) \text{ in the discrete case, and is equal to } f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

in the continuous case. The density function for the marginal distribution of Y is found in a similar way:

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

If the cumulative distribution function of the joint distribution of X and Y is $F(x,y)$, then

$$F_X(x) = \lim_{y\to\infty} F(x,y) \text{ and } F_Y(y) = \lim_{x\to\infty} F(x,y) .$$

This can be extended to define the marginal distribution of any one (or subcollection) variable in a multivariate distribution.

Independence of random variables X and Y: Random variables X and Y with cumulative distribution functions $F_X(x)$ and $F_Y(y)$ are said to be independent (or stochastically independent) if the cumulative distribution function of the joint distribution $F(x,y)$ can be factored in the form

$$F(x,y) = F_X(x) \cdot F_Y(y) \text{ for all } (x,y).$$

This definition can be extended to a multivariate distribution of more than 2 variables.
If X and Y are independent, then $f(x,y) = f_X(x) \cdot f_Y(y)$, (but the reverse implication is not always true, i.e. if the joint distribution probability or density function can be factored in the form $f(x,y) = f_X(x) \cdot f_Y(y)$ then X and Y are usually, but not always, independent).

Conditional distribution of Y given $X = x$: Suppose that the random variables X and Y have joint density/probability function $f(x,y)$, and the density/probability function of the marginal distribution of X is $f_X(x)$. The density/probability function of the conditional distribution of Y given $X = x$ is

$$f_{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)} \text{ if } f_X(x) \neq 0 .$$

The conditional expectation of Y given $X = x$ is $E[Y|X = x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|X = x)\,dy$ in the continuous case, and $E[Y|X = x] = \sum_x y \cdot f_{Y|X}(y|X = x)$ in the discrete case.

If X and Y are independent random variables, then

$$f_{X|Y=y}(x) = f_X(x) \text{ and } f_{Y|X=x}(y) = f_Y(y) .$$

Covariance between random variables X and Y: If random variables X and Y are jointly distributed with joint density/probability function $f(x,y)$, then the covariance between X and Y is $Cov[X,Y] = E[(X - E[X])(Y - E[Y])] = E[(X - \mu_X)(Y - \mu_Y)]$.
Note that $Cov[X,X] = Var[X]$.

Coefficient of correlation between random variables X and Y:
The coefficient of correlation between random variables X and Y is

$$-1 \leq \rho_{X,Y} = \frac{Cov[X,Y]}{\sigma_X \ \sigma_Y} \leq 1$$

where σ_X and σ_Y are the standard deviations of X and Y respectively.

Moment generating function of a joint distribution: Given jointly distributed random variables X and Y, the moment generating function of the joint distribution is

$$M_{X,Y}(t_1, t_2) = E[e^{t_1 X + t_2 Y}] \, .$$

This definition can be extended to the joint distribution of any number of random variables.

Multinomial distribution with parameters n, p_1, p_2, . . ., p_k (where n is a positive integer and $0 \leq p_i \leq 1$ for all $i = 1, 2, ..., k$ and $p_1 + p_2 + \cdots p_k = 1$):
Suppose that an experiment has k possible outcomes, with probabilities p_1, p_2, ..., p_k respectively. If the experiment is performed n successive times (independently), let X_i denote the number of experiments that resulted in outcome i, so that

$$X \ + X \ + \cdots + X \ = n$$

The multivariate probability function is

$$f(x_1, x_2, ..., x_k) = \frac{n!}{x_1! \cdot x_2! \cdots x_k!} \cdot p_1^{x_1} \cdot p_2^{x_2} \cdots p_k^{x_k}$$

and

$$E[X_i] = np_i \, , \ Var[X_i] = np_i(1 - p_i) \, , \ Cov[X_i X_j] = -np_i p_j \, .$$

For example, the toss of a fair die results in one of $k = 6$ outcomes, with probabilities $p_i = \frac{1}{6}$ for $i = 1, 2, 3, 4, 5, 6$. If the die is tossed n times, then with

X_i = # of tosses resulting in face "i" turning up,

the multivariate distribution of $X_1, X_2, ..., X_6$ is a multinomial distribution.

Some results and formulas related to this section are:

(i) $E[h_1(X,Y) + h_2(X,Y)] = E[h_1(X,Y)] + E[h_2(X,Y)]$, and in particular,

$$E[X+Y] = E[X] + E[Y] \text{ and } E[\sum X_i] = \sum E[X_i]$$

(ii) $\lim_{x \to -\infty} F(x,y) = \lim_{y \to -\infty} F(x,y) = 0$

(iii) $P[(x_1 < X \leq x_2) \cap (y_1 < Y \leq y_2)] = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$

(iv) $P[(X \leq x) \cap (Y \leq y)] = F \ (x) + F \ (y) - F(x,y) \leq 1$.

(v) If X and Y are independent, then for any functions g and h,

$$E[g(X)\cdot h(Y)] = E[g(X)]\cdot E[h(Y)]$$

and in particular, .

$$E[X\cdot Y] = E[X]\cdot E[Y]$$

(vi) The density/probability function of jointly distributed variables X and Y can be written in the form

$$f(x,y) = f_{Y|X}(y|X=x)\cdot f_X(x) = f_{X|Y}(x|Y=y)\cdot f_Y(y)$$

$Cov[X,Y] = E[X\cdot Y] - \mu_X\cdot\mu_Y = E[XY] - E[X]\cdot E[Y] = Cov[Y,X]\cdot$

If X and Y are independent, then $E[X\cdot Y] = E[X]\cdot E[Y]$ and $Cov[X,Y]=0.$

and $Cov[X,Y]=0$. For constants $a,\ b,\ c,\ d,\ e,\ f$ and random variables $X,\ Y,\ Z$ and W,

$$\begin{aligned} &Cov[aX+bY+c\ ,\ dZ+eW+f] \\ &\qquad = adCov[X,Z] + aeCov[X,W] + bdCov[Y,Z] + beCov[Y,W] \end{aligned}$$

(viii) For any jointly distributed random variables X and Y, $-1 \le \rho_{XY} \le 1$

(ix)
$$\begin{aligned} Var[X+Y] &= E[(X+Y)^2] - (E[X+Y])^2 \\ &= E[X^2+2XY+Y^2] - (E[X]+E[Y])^2 \\ &= E[X^2]+E[2XY]+E[Y^2] - (E[X])^2 - 2E[X]E[Y] - (E[Y])^2 \\ &= Var[X]+Var[Y]+2\cdot Cov[X,Y] \end{aligned}$$

If X and Y are independent, then $Var[X+Y] = Var[X]+Var[Y]$.

For any X,Y, $Var[aX+bY+c] = a^2Var[X]+b^2Var[Y]+2ab\cdot Cov[X,Y]$

(x) $M_{X,Y}(t_1,0) = E[e^{t_1X}] = M_X(t_1)$ and $M_{X,Y}(0,t_2) = E[e^{t_2Y}] = M_Y(t_2)$

(xi)
$$\left.\frac{\partial}{\partial t_1} M_{X,Y}(t_1,t_2)\right|_{t_1=t_2=0} = E[X],\quad \left.\frac{\partial}{\partial t_2} M_{X,Y}(t_1,t_2)\right|_{t_1=t_2=0} = E[Y]$$

$$\left.\frac{\partial^{r+s}}{\partial^r t_1\,\partial^s t_2} M_{X,Y}(t_1,t_2)\right|_{t_1=t_2=0} = E[X^r\cdot Y^s]$$

(xii) If

$$M(t_1,t_2) = M(t_1,0)\cdot M(0,t_2)\ ,$$

for t_1 and t_2 in a region about $(0,0)$, then X and Y are independent.

(xiii) If $Y = aX+b$ then $M_Y(t) = e^{bt}M_X(at)$.

(xiv) If X and Y are jointly distributed, then for any y, $E[X|Y=y]$ depends on y, say $E[X|Y=y] = h(y)$. It can then be shown that $E[h(Y)] = E[X]$; this is more

$$E[\ E[X|Y]\ ] = E[X].$$

It can also be shown that

$$Var[X] = E[\ Var[X|Y]\ ] + Var[\ E[X|Y]\ ].$$

(xv) A random variable X can be defined as a combination of two (or more) random variables X_1 and X_2, defined in terms of whether or not a particular event A occurs.

$$X = \begin{cases} X_1 & \text{if event } A \text{ occurs (probability } p) \\ X_2 & \text{if event } A \text{ does not occur (probability } 1-p) \end{cases}$$

Then, Y can be the indicator random variable $I_A = \begin{cases} 1 & \text{if } A \text{ occurs (prob. } p) \\ 0 & \text{if } A \text{ doesn't occur (prob. } 1-p) \end{cases}$

Probabilities and expectations involving X can be found by "conditioning" over Y:

$$P[X \leq c] = P[X \leq c | A \text{ occurs}] \cdot P[A \text{ occurs}] + P[X \leq c | A' \text{ occurs}] \cdot P[A' \text{ occurs}]$$
$$= P[X_1 \leq c] \cdot p + P[X_2 \leq c] \cdot (1-p),$$
$$E[X^k] = E[X_1^k] \cdot p + E[X_2^k] \cdot (1-p) \, , \; M_X(t) = M_{X_1}(t) \cdot p + M_{X_2}(t) \cdot (1-p)$$

This is really an illustration of a mixture of the distributions of X_1 and X_2, with $\alpha_1 = p$ and $\alpha_2 = 1 - p$.

As an example, suppose there are two urns containing balls - Urn I contains 5 red and 5 blue balls and Urn II contains 8 red and 2 blue balls. A die is tossed, and if the number turning up is even then 2 balls are picked from Urn I, and if the number turning up is odd then 3 balls are picked from Urn II. X is the number of red balls chosen. Event A would be $A =$ die toss is even. Random variable X_1 would be the number of red balls chosen from Urn I and X_2 would be the number of red balls chosen from Urn II, and since each urn is equally likely to be chosen, $\alpha_1 = \alpha_2 = \frac{1}{2}$.

(xvi) If X and Y have a joint distribution which is uniform on the two dimensional region R (usually R will be a triangle, rectangle or circle in the (x, y) plane), then the conditional distribution of Y given $X = x$ has a uniform distribution on the line segment (or segments) defined by the intersection of the region R with the line $X = x$. The marginal distribution of Y might or might not be uniform.

Example 116: If $f(x,y) = K(x^2 + y^2)$ is the density function for the joint distribution of the continuous random variables X and Y defined over the unit square bounded by the points $(0,0)$, $(1,0)$, $(1,1)$ and $(0,1)$, find K.

Solution: The (double) integral of the density function over the region of density must be 1, so that $1 = \int_0^1 \int_0^1 K(x^2 + y^2)\, dy\, dx = K \cdot \frac{2}{3} \rightarrow K = \frac{3}{2}$. □

Example 117: The cumulative distribution function for the joint distribution of the continuous random variables X and Y is $F(x,y) = (.2)(3x^3y + 2x^2y^2)$, for $0 \leq x \leq 1$ and $0 \leq y \leq 1$. Find $f(\frac{1}{2}, \frac{1}{2})$.

Solution: $f(x,y) = \frac{\partial^2}{\partial x\, \partial y} F(x,y) = (.2)(9x^2 + 8xy) \rightarrow f(\frac{1}{2}, \frac{1}{2}) = \frac{17}{20}$. □

Example 118: X and Y are discrete random variables which are jointly distributed with the following probability function $f(x,y)$:

		X		
		-1	0	1
Y	1	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{6}$
	0	$\frac{1}{9}$	0	$\frac{1}{6}$
	-1	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{9}$

Find $E[X \cdot Y]$.

Solution: $E[XY] = \sum_x \sum_y xy \cdot f(x,y) = (-1)(1)(\frac{1}{18}) + (-1)(0)(\frac{1}{9}) + (-1)(-1)(\frac{1}{6})$

$$+ (0)(1)(\tfrac{1}{9}) + (0)(0)(0) + (0)(-1)(\tfrac{1}{9})$$

$$+ (1)(1)(\tfrac{1}{6}) + (1)(0)(\tfrac{1}{6}) + (1)(-1)(\tfrac{1}{9}) = \tfrac{1}{6}$$ □

Example 119: Continuous random variables X and Y have a joint distribution with density function $f(x,y) = \frac{3(2-2x-y)}{2}$ in the region bounded by $y = 0$, $x = 0$ and $y = 2 - 2x$. Find the density function for the marginal distribution of X for $0 < x < 1$.

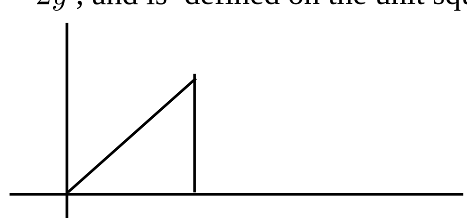
Solution: The region of joint density is illustrated in the graph below. Note that X must be in the

interval $(0,1)$ and Y must be in the interval $(0,2)$. Since $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$, we note that given a value of x in $(0,1)$, the possible values of y (with non-zero density for $f(x,y)$) must satisfy $0 < y < 2 - 2x$, so that

$$f_X(x) = \int_0^{2-2x} f(x,y)\,dy = \int_0^{2-2x} \frac{3(2-2x-y)}{2}\,dy = 3(1-x)^2 .$$ □

Example 120: Suppose that X and Y are independent continuous random variables with the following density functions: $f_X(x) = 1$ for $0 < x < 1$ and $f_Y(y) = 2y$ for $0 < y < 1$. Find $P[Y < X]$.

Solution: Since X and Y are independent, the density function of the joint distribution of X and Y is $f(x,y) = f_X(x) \cdot f_Y(y) = 2y$, and is defined on the unit square. The graph below



right illustrates the region for the probability in question. $P[Y < X] = \int_0^1 \int_0^x 2y\,dy\,dx = \frac{1}{3}$ □

Example 121: Continuous random variables X and Y have a joint distribution with density function $f(x,y) = x^2 + \frac{xy}{3}$ for $0 < x < 1$ and $0 < y < 2$.
Find $P[X > \frac{1}{2}|Y > \frac{1}{2}]$.

Solution: $P[X > \frac{1}{2}|Y > \frac{1}{2}] = \frac{P[(X>\frac{1}{2})\cap(Y>\frac{1}{2})]}{P[Y>\frac{1}{2}]}$.

$$P[(X > \tfrac{1}{2}) \cap (Y > \tfrac{1}{2})] = \int_{1/2}^{1}\int_{1/2}^{2}[x^2 + \tfrac{xy}{3}]\,dy\,dx = \tfrac{43}{64} .$$

$$P[Y > \tfrac{1}{2}] = \int_{1/2}^{2} f_Y(y)\,dy = \int_{1/2}^{2}[\int_0^1 f(x,y)\,dx]\,dy = \int_{1/2}^{2}\int_0^1 [x^2 + \tfrac{xy}{3}]\,dx\,dy = \tfrac{13}{16}$$

$$\rightarrow \quad P[X > \tfrac{1}{2}|Y > \tfrac{1}{2}] = \frac{43/64}{13/16} = \tfrac{43}{52} . \qquad \square$$

Example 122: Continuous random variables X and Y have a joint distribution with density function $f(x,y) = \frac{\pi}{2}\left(sin\,\frac{\pi}{2}y\right)e^{-x}$ for $0 < x < \infty$ and $0 < y < 1$.
Find $P[X > 1|Y = \frac{1}{2}]$.

Solution: $P[X > 1|Y = \frac{1}{2}] = \frac{P[(X>1)\cap(Y=\frac{1}{2})]}{f_Y(\frac{1}{2})}$.

$$P[(X > 1) \cap (Y = \tfrac{1}{2})] = \int_1^\infty f(x,\tfrac{1}{2})\,dx = \int_1^\infty \tfrac{\pi}{2}\left(sin\,\tfrac{\pi}{4}\right)e^{-x}dx = \tfrac{\pi\sqrt{2}}{4} \cdot e^{-1} .$$

$$f_Y(\tfrac{1}{2}) = \int_0^\infty f(x,\tfrac{1}{2})\,dx = \int_0^\infty \tfrac{\pi}{2}\left(sin\,\tfrac{\pi}{4}\right)e^{-x}dx = \tfrac{\pi\sqrt{2}}{4}$$

$$\rightarrow \quad P[X > 1|Y = \tfrac{1}{2}] = e^{-1} . \qquad \square$$

Example 123: X is a continuous random variable with density function $f_X(x) = x + \frac{1}{2}$ for $0 < x < 1$. X is also jointly distributed with the continuous random variable Y, and the conditional density function of Y given $X = x$ is

$$f_{Y|X}(y|X = x) = \frac{x+y}{x+\frac{1}{2}} , \qquad 0 < x < 1 , \qquad 0 < y < 1.$$

Find $f_Y(y)$ for $0 < y < 1$.

Solution: $f(x,y) = f(y|x) \cdot f_X(x) = \frac{x+y}{x+\frac{1}{2}} \cdot (x + \frac{1}{2}) = x + y$ Then, $f_Y(y) = \int_0^1 f(x,y)\,dx = y + \frac{1}{2}$. $\square$

Example 124: Find $Cov[X,Y]$ for the jointly distributed discrete random variables in Example 118 above.

Solution: $Cov[X,Y] = E[XY] - E[X] \cdot E[Y]$. In Example 118 it was found that $E[XY] = \frac{1}{6}$. The marginal probability function for X is $P[X = 1] = \frac{1}{6} + \frac{1}{6} + \frac{1}{9} = \frac{4}{9}$, $P[X = 0] = \frac{2}{9}$ and $P[X = -1] = \frac{1}{3}$, and the mean of X is

$$E[X] = (1)(\tfrac{4}{9}) + (0)(\tfrac{2}{9}) + (-1)(\tfrac{1}{3}) = \tfrac{1}{9} .$$

In a similar way, the probability function of Y is found to be $P[Y = 1] = \frac{1}{3}$, $P[Y = 0] = \frac{5}{18}$, and $P[Y = -1] = \frac{7}{18}$, with a mean of $E[Y] = -\frac{1}{18}$.
Then, $Cov[X,Y] = \frac{1}{6} - (\frac{1}{9})(-\frac{1}{18}) = \frac{14}{81}$. $\square$

Example 125: The coefficient of correlation between random variables X and Y is $\frac{1}{3}$, and $\sigma_X^2 = a$, $\sigma_Y^2 = 4a$. The random variable Z is defined to be $Z = 3X - 4Y$, and it is found that $\sigma_Z^2 = 114$. Find a.

Solution: $\sigma_Z^2 = Var[Z] = 9Var[X] + 16Var[Y] - 2 \cdot (3)(4)\,Cov[X,Y]$.
Since $Cov[X,Y] = \rho[X,Y] \cdot \sigma_X \cdot \sigma_Y = \frac{1}{3} \cdot \sqrt{a} \cdot \sqrt{4a} = \frac{2a}{3}$, it follows that
$114 = \sigma_Z^2 = 9a + 16(4a) - 24(\frac{2a}{3}) = 57a \rightarrow a = 2$. □

Example 126: Suppose that X and Y are random variables whose joint distribution has moment generating function $M(t_1, t_2) = \left(\frac{1}{4}e^{t_1} + \frac{3}{8}e^{t_2} + \frac{3}{8}\right)^{10}$, for all real t_1 and t_2.
Find the covariance between X and Y.

Solution: $Cov[X,Y] = E[XY] - E[X] \cdot E[Y]$.

$$E[XY] = \frac{\partial^2}{\partial t_1 \partial t_2} M_{X,Y}(t_1,t_2)\Big|_{t_1=t_2=0}$$
$$= (10)(9)\left(\frac{1}{4}e^{t_1} + \frac{3}{8}e^{t_2} + \frac{3}{8}\right)^8 \left(\frac{1}{4}e^{t_1}\right)\left(\frac{3}{8}e^{t_2}\right)\Big|_{t_1=t_2=0} = \frac{135}{16} ,$$

$$E[X] = \frac{\partial}{\partial t_1} M_{X,Y}(t_1,t_2)\Big|_{t_1=t_2=0} = (10)\left(\frac{1}{4}e^{t_1} + \frac{3}{8}e^{t_2} + \frac{3}{8}\right)^9 \left(\frac{1}{4}e^{t_1}\right)\Big|_{t_1=t_2=0} = \frac{5}{2} ,$$
$$E[Y] = \frac{\partial}{\partial t_2} M_{X,Y}(t_1,t_2)\Big|_{t_1=t_2=0} = (10)\left(\frac{1}{4}e^{t_1} + \frac{3}{8}e^{t_2} + \frac{3}{8}\right)^9 \left(\frac{3}{8}e^{t_2}\right)\Big|_{t_1=t_2=0} = \frac{15}{4} ,$$
$\rightarrow Cov[X,Y] = \frac{135}{16} - (\frac{5}{2})(\frac{15}{4}) = -\frac{15}{16}$. □

Example 127: Suppose that X has a continuous distribution with p.d.f. $f_X(x) = 2x$ on the interval $(0,1)$, and $f_X(x) = 0$ elsewhere. Suppose that Y is a continuous random variable such that the conditional distribution of Y given $X = x$ is uniform on the interval $(0,x)$. Find the mean and variance of Y.

Solution: This problem can be approached in two ways.

(i) The first approach is to determine the unconditional (marginal) distribution of Y. We are given $f_X(x) = 2x$ for $0 < x < 1$, and $f_{Y|X}(y|X = x) = \frac{1}{x}$ for $0 < y < x$.
Then, $f(x,y) = f(y|x) \cdot f_X(x) = \frac{1}{x} \cdot 2x = 2$ for $0 < x < 1$ and $0 < y < x$.
The unconditional (marginal) distribution of Y has p.d.f.
$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_y^1 2\,dx = 2(1-y)$ for $0 < y < 1$ (and $f_Y(y)$ is 0 elsewhere). Then $E[Y] = \int_0^1 y \cdot 2(1-y)\,dy = \frac{1}{3}$, $E[Y^2] = \int_0^1 y^2 \cdot 2(1-y)\,dy = \frac{1}{6}$, and $Var[Y] = E[Y^2] - (E[Y])^2 = \frac{1}{6} - (\frac{1}{3})^2 = \frac{1}{18}$.

(ii) The second approach is to use the relationships $E[Y] = E[\,E[Y|X]\,]$ and $Var[Y] = E[\,Var[Y|X]\,] + Var[\,E[Y|X]\,]$.

From the conditional density $f(y|X=x) = \frac{1}{x}$ for $0 < y < x$, we have $E[Y|X=x] = \int_0^x y \cdot \frac{1}{x}\,dy = \frac{x}{2}$, so that $E[Y|X] = \frac{X}{2}$, and, since $f_X(x) = 2x$,

$$E[\,E[Y|X]\,] = E[\tfrac{X}{2}] = \int_0^1 \tfrac{x}{2} \cdot 2x dx = \tfrac{1}{3} = E[Y] \ .$$

In a similar way, $Var[Y|X=x] = E[Y^2|X=x] - (E[Y|X=x])^2$, where $E[Y^2|X=x] = \int_0^x y^2 \cdot \frac{1}{x}\,dy = \frac{x^2}{3}$, so that $E[Y^2|X] = \frac{X^2}{3}$, and since $E[Y|X] = \frac{X}{2}$, we have $Var[Y|X] = \frac{X^2}{3} - (\frac{X}{2})^2 = \frac{X^2}{12}$.

Then $E[\,Var[Y|X]\,] = E[\frac{X^2}{12}] = \int_0^1 \frac{x^2}{12} \cdot 2x\,dx = \frac{1}{24}$, and

$Var[E[Y|X]] = Var[\frac{X}{2}] = \frac{1}{4}Var[X] = \frac{1}{4} \cdot [E[X^2] - (E[X])^2] = \frac{1}{4} \cdot [\frac{1}{2} - (\frac{2}{3})^2] = \frac{1}{72}$

so that .

$$E[\,Var[Y|X]\,] + Var[\,E[Y|X]\,] = \tfrac{1}{24} + \tfrac{1}{72} = \tfrac{1}{18} = Var[Y] \qquad \square$$

FUNCTIONS AND TRANSFORMATIONS OF RANDOM VARIABLES

Distribution of a function of a continuous random variable X: Suppose that X is a continuous random variable with p.d.f. $f_X(x)$ and c.d.f. $F_X(x)$, and suppose that $u(x)$ is a one-to-one function (usually u is either strictly increasing, such as $u(x) = x^3$, e^x, $\sqrt{x}$ or $ln\, x$, or u is strictly decreasing, such as $u(x) = e^{-x}$). As a one-to-one function, u has an inverse function v, so that $v(u(x)) = x$. Then the random variable $Y = u(X)$ (Y is referred to as a **transformation of X**) has p.d.f. $f_Y(y)$ found as follows: $f_Y(y) = f_X(v(y)) \cdot |v'(y)|$.

$$f_Y(y) = f_X(v(y)) \cdot |v'(y)|$$

If u a strictly increasing function, then

$$F_Y(y) = P[Y \leq y] = P[u(X) \leq x] = P[X \leq v(y)] = F_X(v(y)) \, .$$

Distribution of a function of a discrete random variable X: Suppose that X is a discrete random variable with probability function $f(x)$. If $u(x)$ is a function of x, and Y is a random variable defined by the equation $Y = u(X)$, then Y is a discrete random variable with probability function $g(y) = \sum_{y=u(x)} f(x)$ - given a value of y, find all values of x for which $y = u(x)$ (say $u(x_1) = u(x_2) = \cdots = u(x_t) = y$), and then $g(y)$ is the sum of those $f(x_i)$ probabilities.

If X and Y are independent random variables, and u and v are functions, then the random variables $u(X)$ and $v(Y)$ are independent.

The distribution of a sum of random variables:

(i) If X_1 and X_2 are random variables, and $Y = X_1 + X_2$, then

$$E[Y] = E[X_1] + E[X_2] \text{ and } Var[Y] = Var[X_1] + Var[X_2] + 2Cov[X_1, X_2]$$

(ii) If X_1 and X_2 are discrete non-negative integer valued random variables with joint probability function $f(x_1, x_2)$, then for an integer $k \geq 0$,

$$P[X_1 + X_2 = k] = \sum_{x_1=0}^{k} f(x_1, k - x_1)$$

If X_1 and X_2 are independent with probability functions $f_1(x_1)$ and $f_2(x_2)$, respectively, then

$$P[X_1 + X_2 = k] = \sum_{x_1=0}^{k} f_1(x_1) \cdot f_2(k - x_1)$$

(this is the **convolution method** of finding the distribution of the sum of independent discrete random variables).

(iii) If X_1 and X_2 are continuous random variables with joint density function $f(x_1, x_2)$ then the density function of $Y = X_1 + X_2$ is

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

If X_1 and X_2 are independent continuous random variables with density functions $f_1(x_1)$ and $f_2(x_2)$, then the density function of $Y = X_1 + X_2$ is

$$f_Y(y) = \int_{-\infty}^{\infty} f_1(x_1) \cdot f_2(y - x_1)\, dx_1$$

(iv) If $X_1, X_2, ..., X_n$ are random variables, and the random variable Y is defined to be

$$Y = \sum_{i=1}^{n} X_i \text{ , then } E[Y] = \sum_{i=1}^{n} E[X_i] \text{ and}$$

$$Var[Y] = \sum_{i=1}^{n} Var[X_i] + 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} Cov[X_i, X_j] \ .$$

If $X_1, X_2, ..., X_n$ are mutually independent random variables, then

$$Var[Y] = \sum_{i=1}^{n} Var[X_i] \text{ and } M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdots M_{X_n}(t)$$

(v) If $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_m$ are random variables and $a_1, a_2, ..., a_n$, b, $c_1, c_2, ..., c_m$ and d are constants, then

$$Cov[\sum_{i=1}^{n} a_i X_i + b, \sum_{j=1}^{m} c_j Y_j + d] = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i c_j Cov[X_i, Y_j]$$

(vi) **The Central Limit Theorem:** Suppose that X is a random variable with mean μ and standard deviation σ and suppose that $X_1, X_2, ..., X_n$ are n independent random variables with the same distribution as X. Let $Y_n = X_1 + X_2 + \cdots + X_n$. Then $E[Y_n] = n\mu$ and $Var[Y_n] = n\sigma^2$, and as n increases, the distribution of Y_n approaches a normal distribution $N(n\mu, n\sigma^2)$. This is a justification for using the normal distribution as an approximation to the distribution of a sum of random variables.

(vii) **Sums of certain distributions:** Suppose that $X_1, X_2, ..., X_k$ are independent random variables and

$$Y = \sum_{i=1}^{k} X_i$$

distribution of X_i	distribution of Y
Bernoulli $B(1, p)$	binomial $B(k, p)$
binomial $B(n_i, p)$	binomial $B(\sum n_i, p)$
geometric p	negative binomial k, p
negative binomial n_i, p	negative binomial $\sum n_i, p$
Poisson λ_i	Poisson $\sum \lambda_i$
$N(\mu_i, \sigma_i^2)$	$N(\sum \mu_i, \sum \sigma_i^2)$

Example 128: The random variable X has an exponential distribution with a mean of 1. The random variable Y is defined to be $Y = 2\,ln\,X$. Find $f_Y(y)$, the p.d.f. of Y.

Solution: $F_Y(y) = P[Y \leq y] = P[2\,ln\,X \leq y] = P[X \leq e^{y/2}] = 1 - e^{-e^{y/2}}$

$$\rightarrow f_Y(y) = F_Y'(y) = \frac{d}{dy}\,(1 - e^{-e^{y/2}}) = \frac{1}{2}e^{y/2} \cdot e^{-e^{y/2}}.$$

Alternatively, since $Y = 2\,ln\,X$ ($y = u(x) = 2\,ln\,x$, and ln is a strictly increasing function with inverse $x = v(y) = e^{y/2}$), and $X = e^{Y/2}$, it follows that

$$f_Y(y) = f_X(e^{y/2}) \cdot \left|\frac{d}{dy}\,e^{y/2}\right| = \frac{1}{2}e^{y/2} \cdot e^{-e^{y/2}}.$$ □

Example 129: Suppose that X and Y are independent discrete integer-valued random variables with X uniformly distributed on the integers 1 to 5, and Y having the following probability function

$$f_Y(0) = .3 \quad f_Y(1) = .5 \quad f_Y(3) = .2\ .$$

Let $Z = X + Y$. Find $P[Z = 5]$.

Solution: Using the fact that $f_X(x) = .2$ for $x = 1, 2, 3, 4, 5$, and the convolution method for independent discrete random variables, we have

$$\begin{aligned} f_Z(5) &= \sum_{i=0}^{5} f_X(i) \cdot f_Y(5-i) \\ &= (0)(0) + (.2)(0) + (.2)(.2) + (.2)(0) + (.2)(.5) + (.2)(.2) = .20\ . \end{aligned}$$ □

Example 130: X_1 and X_2 are independent exponential random variables each with a mean of 1. Find $P[X_1 + X_2 < 1]$.

Solution: Using the convolution method, the density function of $Y = X_1 + X_2$ is

$$f_Y(y) = \int_0^y f_{X_1}(t) \cdot f_{X_2}(y-t)\,dt = \int_0^y e^{-t} \cdot e^{-(y-t)}\,dt = ye^{-y},$$

so that

$$P[X_1 + X_2 < 1] = P[Y < 1] = \int_0^1 ye^{-y}dy = [-ye^{-y} - e^{-y}]\Big|_{y=0}^{y=1} = 1 - 2e^{-1}$$

(the last integral required integration by parts). □

Example 131: Given n independent random variables $X_1, X_2, ..., X_n$ each having the same variance of σ^2, and defining

$$U = 2X_1 + X_2 + \cdots + X_{n-1} \quad \text{and} \quad V = X_2 + X_3 + \cdots + 2X_n\ ,$$

find the coefficient of correlation between U and V.

Solution: $\rho_{UV} = \frac{Cov[U,V]}{\sigma_U \sigma_V}$; $\sigma_U^2 = (4 + 1 + 1 + \cdots + 1)\sigma^2 = (n+2)\sigma^2 = \sigma_V^2$.

Since the X 's are independent, if $i \neq j$ then $Cov[X_i, X_j] = 0$. Then, noting that $Cov[W,W] = Var[W]$, we have

$$\begin{aligned} Cov[U,V] &= Cov[2X_1, X_2] + Cov[2X_1, X_3] + \cdots + Cov[X_{n-1}, 2X_n] \\ &= Var[X_2] + Var[X_3] + \cdots + Var[X_{n-1}] = (n-2)\sigma^2\ . \end{aligned}$$

Then, $\rho_{UV} = \frac{(n-2)\sigma^2}{(n+2)\sigma^2} = \frac{n-2}{n+2}$. □

Example 132: Independent random variables $X,\ Y$ and Z are identically distributed. Let $W = X + Y$. The moment generating function of W is $M_W(t) = (.7 + .3e^t)^6$.
Find the moment generating function of $V = X + Y + Z$.

Solution: For independent random variables,
$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = (.7 + .3e^t)^6$. Since X and Y have identical distributions, they have the same moment generating function. Thus,

$$M_X(t) = (.7 + .3e^t)^3 \text{ , and then } M_V(t) = M_X(t) \cdot M_Y(t) \cdot M_Z(t) = (.7 + .3e^t)^9 \text{ .}$$

Alternatively, note that the moment generating function of the binomial $B(n,p)$ is $(1 - p + pe^t)^n$. Thus, $X + Y$ has a $B(6, .3)$ distribution, and each of X, Y and Z has a $B(3, .3)$ distribution, so that the sum of these independent binomial distributions is $B(9, .3)$, with m.g.f. $(.7 + .3e^t)^9$. □

Example 133: The birth weight of males is normally distributed with mean 6 pounds, 10 ounces, standard deviation 1 pound. For females, the mean weight is 7 pounds, 2 ounces with standard deviation 12 ounces. Given two independent male/female births, find the probability that the baby boy outweighs the baby girl.

Solution: Let random variables X and Y denote the boy's weight and girl's weight, respectively. Then, $W = X - Y$ has a normal distribution with mean

$$6\tfrac{10}{16} - 7\tfrac{2}{16} = -\tfrac{1}{2} \text{ lb. and variance } \sigma_X^2 + \sigma_Y^2 = 1 + \tfrac{9}{16} = \tfrac{25}{16} \text{ .}$$

Then, $P[X > Y] = P[X - Y > 0] = P\left[\frac{W-(-\frac{1}{2})}{\sqrt{25/16}} > \frac{-(-\frac{1}{2})}{\sqrt{25/16}}\right] = P[Z > .4]$,

where Z has standard normal distribution (W was standardized). Referring to the standard normal table, this probability is $.34$. □

Example 134: If the number of typographical errors per page type by a certain typist follows a Poisson distribution with a mean of λ, find the probability that the total number of errors in 10 randomly selected pages is 10.

Solution: The 10 randomly selected pages have independent distributions of errors per page. The sum of m independent Poisson random variables with parameters $\lambda_1, \lambda_2 ...., \lambda_m$ has a Poisson distribution with parameter $\sum \lambda_i$. Thus, the total number of errors in the 10 randomly selected pages has a Poisson distribution with parameter 10λ.
The probability of 10 errors in the 10 pages is $\frac{e^{-10\lambda}(10\lambda)^{10}}{10!}$. □