1. Probability

1.1 Sample Space and Events

- Suppose that we are about to perform an experiment whose outcome is not predictable in advance.

- However, while the outcome of the experiment will not be known in advance, let us suppose that the set of all possible outcomes is known.

- This set of all possible outcome of an experiment is known as the *sample space* of the experiment and is denoted by S.

Some examples are the following

1. If the experiment consists of the flipping of a coin, then

$$S = \{H, T\}$$

   where H means that the outcome of the toss is a head and T that it is a tail.

2. If the experiment consists of rolling a die, then the sample space is

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

   where the outcome i means that i appeared on the die, $i = 1, 2, 3, 4, 5, 6$.

3. If the experiments consists of flipping two coins, then sample space consists of the following four points

$$S = \{(H, H), (H, T), (T, H), (T, T)\}.$$

4. If the experiment consists of rolling two dice, then the sample space consists of the following 36 points

$$S = \{(i,j), i,j = 1,2,3,4,5,6\}$$

where i appears on the first die, and j on the second die.

5. If the experiment consists of measuring the lifetime of a car, then the sample space consists of all nonnegative real number. That is

$$S = [0, \infty).$$

- Any subset E of the sample space S is known as an *event*.

  For examples, in Example 1, $E = \{H\}$ or $\{T\}$, Example 2, $E = \{1\}$, or $\{2,4,6\}$, Example 3, $E = \{(H,H),(H,T)\}$, Example 4, $E = (1,6),(2,5),(3,4),(4,3),(5,2),(6,1)$, Example 5, $E = (2,6)$.

- For any two events E and F of a sample space S we define the new event $E \cup F$ to consist of all outcomes that are either in E or in F or in both E and F.

- For any two events E and F , we may also define the new event $E \cap F$, referred to as the intersection of E and F, all of outcomes which are both in E and in F.

- For example, in Example 1, if $E = \{H\}$ and $F = \{T\}$, then

$$E \cup F = \{H, T\}, E \cap F = \phi \; (\mathit{mutually\ exclusive});$$

in Example 2, if $E = \{1, 3, 5\}$ and $F = \{1, 2, 3\}$, then

$$E \cup F = \{1, 2, 3, 5\}, E \cap F = \{1, 3\}.$$

- If $E_1, E_2, \ldots,$ are events, then the union of these events, denoted by $\bigcup_{i=1}^{\infty} E_i$, is defined to be that event which consists of all outcomes that are in E_n for at least one value of $n = 1, 2, \ldots$.

- Similarly, the interaction of events E_n, denoted by $\bigcap_{i=1}^{\infty} E_i$, is defined to be the event consisting of those outcomes that are in all of the events $E_n, n = 1, 2, \ldots$.

- For any event E we define the new event E^c, referred to as the *complement* fo E, to consist of all outcomes in the sample space S that are not in E.

  For example, in Example (4) if $E = \{(1,6),(2,5)(3,4),(4,3),(5,2),(6,1)\}$, then E^c will occur if sum of the dice not equal to seven.

1.2 Probabilities Defined on Events

Consider an experiment whose sample space is S. For each event E of the sample space S, we assume that a number $P(E)$ is defined and satisfies the following three conditions

(i) $0 \leq P(E) \leq 1$.

(ii) $P(S) = 1$.

(iii) For any sequence of events $E_1, E_2, \ldots$ that are mutually exclusive, that is, events for which $E_i \cap E_j = \emptyset$ when $i \neq j$, then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

We refer to $P(E)$ as the Probability of the event E.

- Example 1. In the coin tossing example, if we assume that a head is equally likely to appear as a tail , then we would have

$$P(\{H\}) = P(\{T\}) = \frac{1}{2}$$

On the other hand if we had biased coin and felt that a head was twice as likely to appear as a tail, then we would have

$$P(\{H\}) = \frac{2}{3},\ P(\{T\}) = \frac{1}{3}.$$

- Example 2. In the die tossing example, if we supposed that all six number were equally likely to appear, then we would have

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6},$$

and then

$$P(\{2,4,6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}.$$

- Since E and E^c are always mutually exclusive and $E \cup E^c = S$, then

$$1 = P(S) = P(E \cup F) = P(E) + p(E^c), \text{or } P(E^c) = 1 - P(E).$$

- Furthermore for the events E and F, we have

$$P(E) + P(F) = P(E \cup F) + P(E \cap F)$$

or equivalently

$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

- In fact it can be shown by induction that, for any n events $E_1, E_2, \ldots, E_n$,

$$\begin{aligned}P(E_1 \cup E_2 \cup \cdots \cup E_n) &= \sum_i P(E_i) - \sum_{i<j} P(E_i \cap E_j) \\ &+ \sum_{i<j<k} P(E_i \cap E_j \cap E_k) \\ &- \sum_{i<j<k<l} P(E_i \cap E_j \cap E_k \cap E_l) + \cdots \\ &+(-1)^{n+1} P(E_1 \cap E_2 \cap \cdots \cap E_n).\end{aligned}$$

- Example 3. Tossing two coins, $S = \{(H,H),(H,T),(T,H),(T,T)\}$, and $E = \{(H,H),(H,T)\}, F = \{(H,H),(T,H)\}$, then

$$P(E \cup F) = P(E) + P(F) - P(E \cap F) = \frac{1}{2} + \frac{1}{2} - P(\{H,H\}) = 1 - \frac{1}{4} = \frac{3}{4}$$

1.3 Conditional Probability and Independent Events

If we let E and F denote respectively the event that the sum of the dice is six and the event that the first die is a four, then the probability just obtained is called the *conditional probability* that E occurs given that F has occurred and is denote by

$$P(E|F)$$

- A general formula for $P(E|F)$ which is valid for all events E and F is derived in the same manner as above. Namely if the event F occurs, then in order for E to occur it is necessary for the actual occurrence to be a point in both E and in F, that is, it must be in $E \cap F$. That is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

which is only defined when $P(F) > 0$.

Examples:

1. Suppose cards numbered one through ten are placed in a hat, mixed up, and then one of the cards is drawn. If we are told that the number on the drawn card is at least five, then what is the conditional probability that it is ten? (Solution: $1/6$)

2. A family has two children. What is the conditional probability that both are boys given that at least one of them is a boy? Assume that the sample space S is given by $S = \{(b,b),(b,g),(g,b),(g,g)\}$, and all outcomes are equally likely. ((b,g) means, for instance, that the older child is a boy and the younger child a girl.) (Solution: $1/3$)

3. Bev can either take a course in computers or in chemistry. If Bev takes the computer course, then she will receive an A grade with probability $\frac{1}{2}$; if she takes the chemistry course then she will receive an A grade with probability $\frac{1}{3}$. Bev decides to base her decision on the flip of a fair coin. What is the probability that Bev will get an A in chemistry? (Solution: $1/6$)

4. Suppose an urn contains seven black balls and five white balls. We draw two balls from the urn without replacement. Assuming that each ball in the urn is equally likely to be drawn, what is the probability that both drawn balls are black? (Solution: 42/132)

5. Suppose that each of three men at a party throws his hat into the center of the room. The hats are first mixed up and then each man randomly selects a hat. What is the probability that none of the three men selects his own hat? (Solution: 1/3)

- Two events E and F are said to be independent if

$$P(E \cap F) = P(E)P(F).$$

This implies that E and F are independent if

$$P(E|F) = P(E).$$

- That is, E and F are independent if knowledge that F has occurred does not affect the probability that E occurs. That is, the occurrence of E is independent of whether or not F occurs.
- Two events E and F that are not independent are said to be dependent.
- **Example** : Suppose we toss two fair dice. Let E_1 and E_2 denote the event that the sum of the dice is six and Seven, and F denote the event that the first die equals four, are E_1 and F independent, and are E_2 and F independent ? (E_1 and F are not independent, but E_2 and F are independent.)

The definition of independence can be extended to more than two events. The events $E_1, E_2, \ldots, E_n$ are said to be independent if for every subset $E_1, E_2, \ldots, E_r, r \leq n$, of these events

$$P(E_1 \cap E_2 \cap \cdots \cap E_r) = P(E_1)P(E_2)\cdots P(E_r).$$

Intuitively, the events $E_1, E_2, \ldots, E_n$ are independent if knowledge of the occurrence of any of these events has no effect on the probability of any other event.

- Example: (Pairwise Independent Events That Are Not Independent). Let a ball be drawn from an urn containing four balls, numbered 1, 2, 3, 4. Let $E = \{1,2\}, F = \{1,3\}, G = \{1,4\}$. If all four outcomes are assumed equally likely, then

$$\begin{aligned} P(E \cap F) &= P(E)P(F) = P(E \cap G) = P(E)P(G) \\ &= P(F \cap G) = P(F)P(G) = \frac{1}{4} \\ &= P(E \cap F \cap G) \neq P(E)P(F)P(G). \end{aligned}$$

  Hence, even though the events E, F, G are pairwise independent, they are not jointly independent.

- **Example**: There are r players, with player i initially having n_i units, $n_i > 0, i = 1, \ldots, r$. At each stage, two of the players are chosen to play a game, with the winner of the game receiving 1 unit from the loser. Any player whose fortune drops to 0 is eliminated, and this continues until a single player has all $n = \sum_{i=1}^{r} n_i$ units, with that player designated as the victor. Assuming that the results of successive games are independent, and that each game is equally likely to be won by either of its two players, find the probability that player i is the victor. (Solution: $\frac{n_i}{n}$)

Suppose that a sequence of experiments, each of which results in either a "success" or a "failure", is to be performed. Let $E_i, i \geq 1$, denote the event that the ith experiment results in a success. If, for all $i_1, i_2, \ldots, i_n$,

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n}) = \prod_{j=1}^{n} P(E_{i_j})$$

we say that the sequence of experiments consists of independent trials.

1.4 Bayes'Formula

Let E and F be events. We may express E as $E = (E \cap F) \cup (E \cap F^c)$ because in order for a point to be in E, it must either be in both E and F, or it must be in E and not in F . Since $E \cap F$ and $E \cap F^c$ are mutually exclusive, we have that

$$
\begin{aligned}
P(E) &= P(E \cap F) + P(E \cap F^c) \\
&= P(E|F)P(F) + P(E|F^c)P(F^c) \\
&= P(E|F)P(F) + P(E|F^c)(1 - P(F))
\end{aligned}
$$

The equation above states that the probability of the event E is a weighted average of the conditional probability of E given that F has occurred and the conditional probability of E given that F has not occurred, each conditional probability being given as much weight as the event on which it is conditioned has of occurring.

Example. Consider two urns. The first contains two white and seven black balls, and the second contains five white and six black balls. We flip a fair coin and then draw a ball from the first urn or the second urn depending on whether the outcome was heads or tails. What is the conditional probability that the outcome of the toss was heads given that a white ball was selected? (Solution: 22/67)

Example. In answering a question on a multiple-choice test a student either knows the answer or guesses. Let p be the probability that she knows the answer and $1 - p$ the probability that she guesses. Assume that a student who guesses at the answer will be correct with probability $1/m$, where m is the number of multiple-choice alternatives. What is the conditional probability that a student knew the answer to a question given that she answered it correctly? (Solution: $\frac{mp}{1+(m-1)p}$)

Example. A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive result" for 1 percent of the healthy persons tested. (That is, if a healthy person is tested, then, with probability 0.01, the test result will imply he has the disease.) If 0.5 percent of the population actually has the disease, what is the probability a person has the disease given that his test result is positive? (Solution: 0.323)

- Suppose that $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that $\bigcup_{i=1}^{n} F_i = S$. In other words, $E = \bigcup_{i=1}^{n}(E \cap F_i)$ and using the fact that the events $E \cap F_i, i = 1, \ldots, n$ are mutually exclusive,

$$P(E) = \sum_{i=1}^{n} P(E \cap F_i) = \sum_{i=1}^{n} P(E|F_i)P(F_i).$$

- Thus, for given events $F_1, F_2, \ldots, F_n$ of which one and only one must occur, $P(E)$ is equal to a weighted average of $P(E|F_i)$, each term being weighted by the probability of the event on which it is conditioned.

- Suppose now that E has occurred and we are interested in determining which one of the F_j also occurred. By the above equation we have that

$$P(F_j|E) = \frac{P(E \cap F_j)}{P(E)} = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

This equation is known as *Bayes' formula*.

Example. You know that a certain letter is equally likely to be in any one of three different folders. Let α_i be the probability that you will find your letter upon making a quick examination of folder i if the letter is, in fact, in folder $i, i = 1, 2, 3$. (We may have $\alpha_i < 1$.) Suppose you look in folder 1 and do not find the letter. What is the probability that the letter is in folder 1 ? (Solution: $\frac{1-\alpha_1}{3-\alpha_1}$)

1.5 Random Variables

- It frequently occurs that in performing an experiment we are mainly interested in some functions of the outcome as opposed to the outcome itself.

- For instance, in tossing dice we are often interested in the sum of the two dice and are not really concerned about the actual outcome. That is, we may be interested in knowing that the sum is seven and not be concerned over whether the actual outcome was $(1, 6)$ or $(2, 5)$ or $(3, 4)$ or $(4, 3)$ or $(5, 2)$ or $(6, 1)$.

- These quantities of interest, or more formally, these real-valued functions defined on the sample space, are known as *random variables*.

- **Example**. Letting X denote the random variable that is defined as the sum of two fair dice; then X could be $2, 3, \ldots, 12$ or

$$P(X = 2) = P(\{1, 1\}) = \frac{1}{36},$$

$$P(X = 3) = P(\{1, 2\}, \{2, 1\}) = \frac{2}{36},$$

$$\ldots\ldots$$

$$P(X = 12) = P(\{6, 6\}) = \frac{1}{36}$$

and

$$1 = P\left\{\bigcup_{n=2}^{12}\{X = n\}\right\} = \sum_{n=2}^{12} P(X = n).$$

- **Example**. For a second example, suppose that our experiment consists of tossing two fair coins. Letting Y denote the number of heads appearing, then Y is a random variable taking on one of the values 0, 1, 2 with respective probabilities $\frac{1}{4}, \frac{2}{4}, \frac{1}{4}$. Of course,

$$P(Y = 0) + P(Y = 1) + P(Y = 2) = 1.$$

Example. Suppose that we toss a coin having a probability p of coming up heads, until the first head appears. Letting N denote the number of flips required, then assuming that the outcome of successive flips are independent, N is a random variable taking on one of the values $1, 2, 3, \ldots,$ with respective probabilities

$$P(N = n) = P\{(\underbrace{T, T, \ldots, T}_{n-1}, H\} = (1-p)^{n-1}p,\ n \geq 1$$

and

$$P\left(\bigcup_{n=1}^{\infty}\{N = n\}\right) = \sum_{n=1}^{\infty} P(N = n) = p\sum_{n=1}^{\infty}(1-p)^{n-1} = \frac{p}{1-(1-p)} = 1.$$

Example. Suppose that our experiment consists of seeing how long a battery can operate before wearing down. Suppose also that we are not primarily interested in the actual lifetime of the battery but are concerned only about whether or not the battery lasts at least two years. In this case, we may define the random variable I by

- $I = 1$, if the lifetime of battery is two or more years
- $I = 0$, otherwise

If E denotes the event that the battery lasts two or more years, then the random variable I is known as the indicator random variable for event E. (Note that I equals 1 or 0 depending on whether or not E occurs.

Example. Suppose that independent trials, each of which results in any of m possible outcomes with respective probabilities $p_1, \ldots, p_m, \sum_{i=1}^m p_i = 1$,are continually performed. Let X denote the number of trials needed until each outcome has occurred at least once. (Solution:

$$
\begin{aligned}
P(X = n) \quad &= \quad \sum_{i=1}^{m} p_i(1-p_i)^{n-1} - \sum_{i<j}(p_i + p_j)(1 - p_i - p_j)^{n-1} \\
&\quad \sum_{i<j<k}(p_i + p_j + p_k)(1 - p_i - p_j - p_k)^{n-1} - \cdots
\end{aligned}
$$

- *Discrete random variables*: a finite or a countable number of possible variables.

- *Continuous random variables*: a continuum of possible variables. One example is the random variable denoting the lifetime of a car, when the car?s lifetime is assumed to take on any value in some interval (a, b).

- *Cumulative distribution function*(cdf) $F(\cdot)$ of the random variable X is defined for any real number b, $-\infty < b < \infty$ by

$$F(b) = P(X \leq b).$$

- Some properties of the cdf F are

(i) $F(b)$ is a nondiscreaing function of b,
(ii) $\lim_{b \to \infty} F(b) = F(\infty) = 1$,
(iii) $\lim_{b \to -\infty} F(b) = F(-\infty) = 0$,

-

$$F(a < X \leq b) = F(b) - F(a)$$

$$P(X < b) = \lim_{h \to 0^+} P(X \leq b - h) = \lim_{h \to 0^+} F(b - h)$$

1.6 Discrete Random Variables

As was previously mentioned, a random variable that can take on at most a countable number of possible values is said to be discrete.

- For a discrete random variable X, we define the probability mass function $p(a)$ of X by $p(a) = P\{X = a\}$

- The probability mass function $p(a)$ is positive for at most a countable number of values of a. That is, if X must assume one of the values $x_1, x_2, \ldots,$ then

$$p(x_i) > 0, i = 1, 2, \ldots$$

$$p(x) = 0, \text{ all other values of } x$$

- Since X must take on one of the values x_i , for the cumulative function F, we have

$$\sum_{i=1}^{\infty} p(x_i) = 1 \text{ and } F(a) = \sum_{\text{all } x_i \leq a} p(x_i)$$

- For instance, suppose X has a probability mass function given by $p(1) = \frac{1}{2},\ p(2) = \frac{1}{3},\ p(3) = \frac{1}{6}$.

- *The Bernolli Random Variable*

$$p(0) = P(X = 0) = 1 - p, \quad p(1) = P(X = 1) = p$$

- *The Binomial Random Variable*

$$p(X = i) = \left(\begin{array}{c} n \\ i \end{array} \right) p^i (1-p)^{n-i}, i = 0, 1, \ldots, n.$$

- **Example.** Suppose that an airplane engine will fail, when in flight, with probability $1 - p$ independently from engine to engine; suppose that the airplane will make a successful flight if at least 50 percent of its engines remain operative. For what values of p is a four-engine plane preferable to a two-engine plane? (Solution: the four-engine plane is safer when the engine success probability is at least as large as $\frac{2}{3}$, whereas the two-engine plane is safer when this probability falls below $\frac{2}{3}$)

- *The Geometric Random Variable*: Suppose that independent trials, each having probability p of being a success, are performed until a success occurs. If we let X be the number of trials required until the first success, then X is said to be a geometric random variable with parameter p. Its probability mass function is given by

$$p(n) = P(X = n) = (1 - p)^{n-1} p, n = 1, 2, \ldots.$$

- *The Poisson Random Variable*: A random variable X, taking on one of the values $0, 1, 2, \ldots,$ is said to be a Poisson random variable with parameter λ, if for some $\lambda > 0$,

$$p(i) = P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}, i = 0, 1, 2, \ldots$$

- **Example**. Suppose that the number of typographical errors on a single page of this book has a Poisson distribution with parameter $\lambda = 1$. Calculate the probability that there is at least one error on this page. (Solution: 0.632)

- **Example**. If the number of accidents occurring on a highway each day is a Poisson random variable with parameter $\lambda = 3$, what is the probability that no accidents occur today? (Solution: 0.05)

1.7 Continuous Random Variables

Let X be such a random variable. We say that X is a continuous random variable if there exists a nonnegative function $f(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P(X \in B) = \int_B f(x)dx.$$

The function $f(x)$ is called the probability density function of the random variable X.

- All probability statements about X can be answered in terms of $f(x)$. For instance, letting $B = [a, b]$, we obtain that

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

- The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P(X \in (-\infty, a]) = \int_{-\infty}^a f(x)dx, \quad \frac{d}{da}F(a) = f(a).$$

- *The Uniform Random Variable*: In general, we say that X is a uniform random variable on the interval (α, β) if its probability density function is given by

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

- **Example**. Calculate the cumulative distribution function of a random variable uniformly distributed over (α, β). (Solution: $0, a < \alpha; \frac{a-\alpha}{\beta-\alpha}, \alpha < a < \beta; 1, a \geq \beta$)

- **Example**. If X is uniformly distributed over $(0, 10)$, calculate the probability that (a) $X < 3$, (b) $X > 7$, (c) $1 < X < 6$. (Solution: $\frac{3}{10}, \frac{3}{10}, \frac{1}{2}$)

- *Exponential Random Variables*: A continuous random variable whose probability density function is given, for some $\lambda > 0$, by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if} x < 0 \end{cases}$$

$$F(a) = \int_0^a \lambda e^{-\lambda x} dx = 1 - e^{-\lambda a},\ a \geq 0.$$

- *Gamma Random Variables*: A continuous random variable whose density is given by
$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & \text{if } x \geq 0 \\ 0, & \text{if} x < 0 \end{cases}$$
where $\lambda > 0, \alpha > 0$, and

$$\Gamma(\alpha) = \int_0^\infty e^x x^{\alpha-1} dx, \quad \Gamma(n) = (n-1)!$$

- *Normal Random Variables*: a normal random variable (or simply that X is normally distributed) with parameters μ and σ^2 if the density of X is given by
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty.$$

- An important fact about normal random variables is that if X is normally distributed with parameters μ and σ^2 then $Y = \alpha X + \beta$ is normally distributed with parameters $\alpha\mu + \beta$ and $\alpha^2\sigma^2$.

1.8 Expectation of a Random Variable

If X is a discrete random variable having a probability mass function $p(x)$, then the expected value of X is defined by

$$\mathrm{E}(X) = \sum_{x:p(x)>0} xp(x)$$

- **Example**. Find $\mathrm{E}(X)$ where X is the outcome when we roll a fair die. (Solution: $\frac{7}{2}$)

- **Example**. Calculate $\mathrm{E}[X]$ when X is (1) a Bernoulli random variable with parameter p,(2) a binomial random variable with parameters n and p, (3) a geometric random variable having parameter p, (4) a Poisson random variable with parameter λ. (Solution: (1) p, (2) np, (3) $\frac{1}{p}$, (4) λ)

We may also define the expected value of a continuous random variable. This is done as follows. If X is a continuous random variable having a probability density function $f(x)$, then the expected value of X is defined by

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

- **Example.** (i) Expectation of an Exponential Random Variable with parameter λ. (ii) Expectation of a Normal Random Variable with parameter μ and σ^2. (Solution: (1) $\frac{1}{\lambda}$,(ii) μ)
- Expectation of a Function of a Random Variable $\mathrm{E}(g(X))$,
  (1) for a discrete random variable X with mass function $p(x)$

$$\mathrm{E}(g(X)) = \sum_{x:p(x)>0} g(x)p(x)$$

  (2) for a continuous random variable with density function $f(x)$

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

Example. Let X be normally distributed with parameters μ and σ^2. Find Var(X). (Solution: σ^2)

Example. Calculate Var(X) when X is binomially distributed with parameters n and p. (Solution: $np(1-p)$)

1.9 Jointly Distribution of Random Variables

We are often interested in probability statements concern- ing two or more random variables. To deal with such probabilities, we define, for any two random variables X and Y , the joint cumulative probability distribution function of X and Y by

$$F(a,b) = P(X \leq a, Y \leq b), -\infty < a, b < \infty$$

- *Marginal distribution* of X and Y:

$$F_x(a) = P(X \leq a) = P(X \leq a, Y < \infty) = F(a, \infty)$$

$$F_y(b) = P(Y \leq b) = P(X \leq \infty, Y < b) = F(\infty, b)$$

- In the case where X and Y are both discrete random variables, it is convenient to define the *joint or marginal probability mass function* of X and Y by

$$p(x,y) = P(X = x, Y = y),$$

$$p_X(x) = \sum_{y:p(x,y)>0} p(x,y),\ \ p_Y(y) = \sum_{x:p(x,y)>0} p(x,y)$$

- X and Y are jointly continuous if there exists a function $f(x, y)$, defined for all real x and y, having the property that for all sets A and B of real numbers
$$P(X \in A, Y \in B) = \int_B \int_A f(x, y) dx, dy)$$
The function $f(x, y)$ is called the joint probability density function of X and Y .

- The probability density of X can be obtained from a knowledge of f (x , y) by the following reasoning:
$$P(X \in A) = P(X \in A, Y \in (-\infty, \infty)) = \int_{-\infty}^{\infty} \int_A f(x, y) dx dy = \int_A f_X(x) dx$$
where $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$

- Because
$$F(a, b) = P(X \leq a, Y \leq b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x, y) dy dx$$
differentiation yields
$$\frac{d^2}{da db} F(a, b) = f(a, b).$$

If X and Y are random variables and g is a function of two variables, then

- In the discrete case

$$\mathrm{E}(g(X,Y)) = \sum_y \sum_x g(x,y)p(x,y)$$

In the continuous case

$$\mathrm{E}(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$$

- If $g(X,Y) = aX + bY$, then

$$\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$$

If $X_1, X_2, \ldots, X_n$ are n random variables, then for any n constants $a_1, a_2, \ldots, a_n$,

$$\mathrm{E}(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1\mathrm{E}(X_1) + a_2\mathrm{E}(X_2) + \cdots + a_n\mathrm{E}(X_n)$$

Example. At a party N men throw their hats into the center of a room. The hats are mixed up and each man randomly selects one. Find the expected number of men who select their own hats. (Solution 1)

Example. Suppose there are 25 different types of coupons and suppose that each time one obtains a coupon, it is equally likely to be any one of the 25 types. Compute the expected number of different types that are contained in a set of 10 coupons. (Solution: $25(1-(\frac{24}{25})^{10})$)

Example. Let $R_1, \cdots, R_{n+m}$ be a random permutation of $1, \cdots, n+m$. (That is, $R_1, \cdots, R_{n+m}$ is equally likely to be any of the $(n+m)!$ permutations of $1, \cdots, n+m$.) For a given $i \leq n$, let X be the ith smallest of the values $R_1, ..., R_n$. Find $\mathrm{E}[X]$. (Solution: $i + m\frac{i}{n+1}$)

Independent Random Variables

The random variables X and Y are said to be independent if, for all a, b,

$$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b) \quad \text{or} \quad F(a, b) = F_X(a)F_y(b)$$

- When X and Y are discrete, the condition of independence reduces to

$$p(x, y) = p_X(x)p_Y(y)$$

- while if X and Y are jointly continuous, independence reduces to

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

- **Proposition**. If X and Y are independent, then for any functions h and g

$$\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)]$$

Covariance and Variance of Random Variables

The covariance of any two random variables X and Y , denoted by Cov(X, Y), is defined by

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

- If X and Y are independent, then Cov$(X, Y) = 0$.
- For any random variable X, Y, Z and constant c

$$\text{Cov}(X, X) = \text{Var}(X), \text{Cov}(X, Y) = \text{Cov}(Y, X),$$

$$\text{Cov}(cX, Y) = c\text{Cov}(X, Y), \text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$$

Definition: If $X_1, \dots, X_n$ are independent and identically distributed, then the n random variable $\bar{X} = \sum_{i=1}^{n} X_i/n$ is called the sample mean.

Proposition. Suppose that $X_1, \ldots, X_n$ are independent and identically distributed with expected value μ and variance σ^2. Then, (a) $\mathrm{E}(\bar{X}) = \mu$, (b) Var$(\bar{X}) = \sigma^2/n$, (c) Cov$(\bar{X}, X_i - \bar{X}) = 0$.

Example. Compute the variance of a binomial random variable X with parameters n and p. (Solution: $np(1-p)$)

Example. (Sums of Independent Poisson Random Variables). Let X and Y be independent Poisson random variables with respective means λ_1 and λ_2. Calculate the distribution of $X + Y$. (Solution: a Poisson distribution with mean $\lambda_1 + \lambda_2$)

Joint Probability Distribution of Functions of Random Variables

Let X_1 and X_2 be jointly continuous random variables with joint probability density function $f(x_1, x_2)$. suppose that $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ for some functions g_1 and g_2. Assume that the functions g_1 and g_2 satisfy the following conditions:

1. The equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2 with solutions given by, say, $x_1 = h_1(y_1, y_2), x_2 = h_2(y_1, y_2)$.
2. The functions g_1 and g_2 have continuous partial derivatives at all points (x_1, x_2) and are such that the following 2×2 determinant

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} = \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1} \neq 0$$

at all points (x_1, x_2).

Under these two conditions it can be shown that the random variables Y_1 and Y_2 are jointly continuous with joint density function given by

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)|J(x_1, x_2)|^{-1}$$

where $x_1 = h_1(y_1, y_2), x_2 = h_2(y_1, y_2)$.

Example. If X and Y are independent gamma random variables with parameters (α, λ) and (β, λ), respectively, compute the joint density of $U = X + Y$ and $V = X/(X+Y)$. (Solution: $f(u,v) = \frac{\lambda e^{-\lambda u}(\lambda u)^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)} \frac{v^{\alpha-1}(1-v)^{\beta-1}\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$, U with a gamma distribution of parameter $(\alpha+\beta, \lambda)$ and V with the beta density with parameter (α, β) are independent.)

- The extension to the joint density function of the n random variables $X_1, X_2, \cdots, X_n$ which is given and want to compute the joint density function of $Y_1, Y_2, \ldots, Y_n$, where $Y_1 = g_1(X_1, \ldots, X_n), \ldots, Y_n = g_n(X_1, \ldots, X_n)$. Furthermore it has a unique solution $x_1 = h_1(y_1, \ldots, y_n), \ldots, x_n = h_n(y_1, \ldots, y_n)$.

  Under these assumptions the joint density function of the random variables Y_i is given by

  $$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = f_{X_1,\ldots,X_n}|J(x_1, \ldots, x_n)|^{-1}$$

  where $x_i = h_i(y_1, \ldots, y_n), i = 1, 2, \ldots, n$, and the Jacobian determinant $J(x_1, \ldots, x_n) \neq 0$.

1.10 Moment Generating Functions

The *moment generating function* $\phi(x)$ of the random variable X is defined for all values t by

$$\phi(t) = \mathrm{E}[e^{tX}] = \begin{cases} \sum_x e^{tX} p(x), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{tX} f(x) dx, & \text{if } X \text{ is continuous} \end{cases}$$

- We call $\phi(t)$ the moment generating function because all of the moments of X can be obtained by successively differentiating $\phi(t)$. For example,

$$\phi'(t) = \frac{d}{dt}\mathrm{E}[e^{tX}] = \mathrm{E}\left[\frac{d}{dt}e^{tX}\right] = \mathrm{E}[Xe^{tX}]$$

and hence

$$\phi'(0) = \mathrm{E}[X].$$

Similarly,

$$\phi''(0) = \mathrm{E}[X^2], \ldots, \phi^n(0) = \mathrm{E}[X^n], n \geq 1.$$

Table 2.1

Discrete probability distribution	**Probability mass function, $p(x)$**	**Moment generating function, $\phi(t)$**	**Mean**	**Variance**
Binomial with parameters n, p, $0 \leq p \leq 1$	$\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$	$(pe^t + (1-p))^n$	np	$np(1-p)$
Poisson with parameter $\lambda > 0$	$e^{-\lambda} \frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$	$\exp\{\lambda(e^t - 1)\}$	λ	λ
Geometric with parameter $0 \leq p \leq 1$	$p(1-p)^{x-1}$, $x = 1, 2, \ldots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

Table 2.2

Continuous probability distribution	**Probability density function, $f(x)$**	**Moment generating function, $\phi(t)$**	**Mean**	**Variance**
Uniform over (a, b)	$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential with parameter $\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$	$\frac{\lambda}{\lambda - t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma with parameters (n, λ), $\lambda > 0$	$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!}, & x \geq 0 \\ 0, & x < 0 \end{cases}$	$\left(\frac{\lambda}{\lambda - t}\right)^n$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$
Normal with parameters (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \times \exp\{-(x-\mu)^2/2\sigma^2\}$, $-\infty < x < \infty$	$\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$	μ	σ^2

- An important property of moment generating functions is that the moment generating function of the sum of independent random variables is just the product of the individual moment generating functions

- To see this, suppose that X and Y are independent and have moment generating functions $\phi_X(t)$ and $\phi_Y(t)$, respectively. Then

$$\phi_{X+Y}(t) = \mathrm{E}e^{t(X+Y)} = \mathrm{E}e^{tX}e^{tY} = \mathrm{E}e^{tX}\mathrm{E}e^{tY} = \phi_X(t)\phi_Y(t)$$

- **Example**.(Sums of Independent Poisson Random Variables). Calculate the distribution of $X+Y$ when X and Y are independent Poisson random variables with means λ_1 and λ_2, respectively (Solution: $X+Y$ is Poisson distributed with mean $\lambda_1+\lambda_2$.

- **Example**.(Sums of Independent Normal Random Variables). Show that if X and Y are independent normal random variables with parameters (μ_1, σ_1^2) and (μ_2, σ_2^2), respectively, then $X+Y$ is normal with mean $\mu_1+\mu_2$ and variance $\sigma_1^2+\sigma_2^2$.

1.11 Limit Theorems

Theorem (Strong Law of Large Number) Let $X_1, X_2, \ldots$ be a sequence of independent random variables having a common distribution, and let $\mathrm{E}[X_i] = \mu$. Then, with probability 1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu, \quad \text{as} \quad n \to \infty.$$

Theorem (Central Limit Theorem). Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables, each with mean μ and variance σ^2 .Then the distribution of

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is as $n \to \infty$

$$P\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx.$$

Example. (Normal Approximation to the Binomial). Let X be the number of times that a fair coin, flipped 40 times, lands heads. Find the probability that $X = 20$. Use the normal approximation and then compare it to the exact solution. (Solution: approximated 0.1272, exact, 0.1254)

Example. Let $X_i, i = 1, 2, \ldots, 10$ be independent random variables, eachbeing uniformly distributed over $(0, 1)$. Estimate $P(\sum_{i=1}^{10} X_i > 7)$. (Solution: 0.0143)

Example. The lifetime of a special type of battery is a random variable with mean 40 hours and standard deviation 20 hours. A battery is used until it fails, at which point it is replaced by a new one. Assuming a stockpile of 25 such batteries, the lifetimes of which are independent, approximate the probability that over 1100 hours of use can be obtained. (Solution: 0.1587)

1.12 Conditional Probability and Conditional Expectation

- One of the most useful concepts in probability theory is that of conditional probability and conditional expectation. The reason is twofold.

- First, in practice, we are often interested in calculating probabilities and expectations when some partial information is available; hence, the desired probabilities and expectations are conditional ones.

- Secondly, in calculating a desired probability or expectation it is often extremely useful to first "condition" on some appropriate random variable.

The Discrete Case

Recall that for any two events E and F , the conditional probability of E given F is defined, as long as $P(F) > 0$, by

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

- Hence, if X and Y are discrete random variables, then it is natural to define the conditional probability mass function of X given that $Y = y$, by

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P_Y(y)},$$

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \sum_{a \leq x} P_{X|Y}(a|y).$$

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

for all values of y such that $P(Y = y) > 0$.

The Continuous Case

- If X and Y have a joint probability density function $f(x,y)$, then the *conditional probability density function* of X, given that $Y=y$, is defined for all values of y such that $f_Y(y)>0$, by

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{f(x,y)}{\int_{-\infty}^{\infty}f(x,y)dx}$$

- The *conditional expectation* of X, given $Y=y$, is defined for all values of y such that $f_Y(y)>0$, by

$$\mathrm{E}[X|Y=y]=\int_{-\infty}^{\infty}xf_{X|Y}(x|y)dx.$$

- If X is independent of Y, then

$$p_{X|Y}(x|y)=P(X=x|Y=y)=P(X=x) \quad \text{or} \quad f_{X|Y}(x|y)=f_X(x)$$

Example. If X_1 and X_2 are independent binomial random variables with respective parameters (n_1, p) and (n_2, p), calculate the conditional probability mass function of X_1 given that $X_1 + X_2 = m$. (Solution: $C_{n_1}^k C_{n_2}^{m-k} / C_{n_1+n_2}^m$)

Example. If X and Y are independent Poisson random variables with respective means λ_1 and λ_2, calculate the conditional expected value of X given that $X + Y = n$. (Solution: $n\frac{\lambda_1}{\lambda_1+\lambda_2}$)

Example. Suppose that the joint density of X and Y is given by

$$f(x, y) = \begin{cases} 6xy(2 - x - y), & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the conditional expectation of X given that $Y = y$, where $0 < y < 1$. (Solution: $\frac{5-4y}{8-6y}$)

Example. Let X_1 and X_2 be independent exponential random variables with rates μ_1 and μ_2. Find the conditional density of X_1 given that $X_1 + X_2 = t$. (Solution: $f_{X_1|X_1+X_2}(x|t) = \frac{(\mu_1-\mu_2)e^{-(\mu_1-\mu_2)x}}{1-e^{-(\mu_1-\mu_2)t}}$, if $\mu_1 \neq \mu_2$. If $\mu_1 = \mu_2$, it is uniformly distributed on $(0, t)$)

Computing Expectations and Probabilities by Conditioning

- Let us denote by $\mathrm{E}[X|Y]$ that function of the random variable Y whose value at $Y = y$ is $\mathrm{E}[X|Y = y]$. Note that $\mathrm{E}[X|Y]$ is itself a random variable. An extremely important property of conditional expectation is that for all random variables X and Y

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

- If Y is a discrete random variable, then the equation above states that

$$\mathrm{E}[X] = \sum_y \mathrm{E}[X|Y = y]P(Y = y)$$

- While if Y is continuous with density $f_Y(y)$, then the above equation says that

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} \mathrm{E}[X|Y = y]f_Y(y)dy$$

Let E denote an arbitrary event and define the indicator random variable X by

$$X = \begin{cases} 1, & \text{if } E \text{ occurs} \\ 0, & \text{if } E \text{ does not occur} \end{cases}$$

It follows from the definition of X that

- $$\mathrm{E}[X] = P(E), \mathrm{E}[X|Y = y] = P(E|Y = y), \text{ for any random variable } Y$$

- $$P(E) = \begin{cases} \sum_y P(E|Y = y)P(Y = y), & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} P(E|Y = y) f_Y(y) dy, & \text{if } Y \text{ is continuous} \end{cases}$$

Conditional expectations can also be used to compute the variance of a random variable. Specifically, we can use

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

- The conditional variance of X given that $Y = y$ is defined by

$$\mathsf{Var}(X|Y = y) = \mathrm{E}[(X - \mathrm{E}[X|Y = y])^2|Y = y] = \mathrm{E}[X^2|Y = y] - (\mathrm{E}[X|Y = y])^2$$

- Hence

$$\mathsf{Var}(X) = \mathrm{E}[\mathsf{Var}(X|Y)] + \mathsf{Var}(E[X|Y])$$

Example. Sam will read either one chapter of his probability book or one chapter of his history book. If the number of misprints in a chapter of his probability book is Poisson distributed with mean 2 and if the number of misprints in his history chapter is Poisson distributed with mean 5, then assuming Sam is equally likely to choose either book, what is the expected number of misprints that Sam will come across? (Solution: 7/2)

Example. (The Expectation of the Sum of a Random Number of Random Variables). Suppose that the expected number of accidents per week at an industrial plant is four. Suppose also that the numbers of workers injured in each accident are independent random variables with a common mean of 2. Assume also that the number of workers injured in each accident is independent of the number of accidents that occur. What is the expected number of injuries during a week? (Solution: 8)

Example. (The Variance of a Compound Random Variable). Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with distribution F having mean μ and variance σ^2, and assume that they are independent of the nonnegative integer valued random variable N, and its expected value was determined, the random variable $S = \sum_{i=1}^{N} X_i$ is called a compound random variable. Find its variance. (Solution: $\sigma^2 \mathrm{E}[N] + \mu^2 \mathrm{Var}(N)$)

Example. Suppose that X and Y are independent continuous random variables having densities f_X and f_Y , respectively. Compute $P(X < Y)$.

Example. An insurance company supposes that the number of accidents that each of its policyholders will have in a year is Poisson distributed, with the mean of the Poisson depending on the policyholder. If the Poisson mean of a randomly chosen policyholder has a gamma distribution with density function

$$g(\lambda) = \lambda e^{-\lambda}, \lambda \geq 0$$

what is the probability that a randomly chosen policyholder has exactly n accidents next year? (Solution: $P(X = n) = (n+1)/2^{n+2}$)

Example. (The Ballot Problem). In an election, candidate A receives n votes, and candidate B receives m votes where $n > m$. Assuming that all orderings are equally likely, show that the probability that A is always ahead in the count of votes is $(n-m)/(n+m)$.

Case Studies: Gambling Problems

We consider an amount $\$S$ of S dollars which is to be shared between two players A and B. At each round, player A may earn $\$1$ with probability $p \in (0, 1)$, and in this case player B loses $\$1$. Conversely, player A may lose $\$1$ with probability $q = 1 - p$, and in this case player B gains $\$1$. We let X_n represent the wealth of player A at time $n \in N$, while $S - X_n$ represents the wealth of player B at time $n \in N$.

The initial wealth X_0 of player A could be negative,1 but for simplicity we will assume that it is comprised between 0 and S. Assuming that the value of $X_n, n \geq 0$, belongs to $\{1, 2, \dots, S - 1\}$ at the time step n, at the next step $n + 1$ we will have

$$\begin{cases} X_{n+1} = X_n + 1 & \text{if player A wins round } n+1, \\ X_{n+1} = X_n - 1, & \text{if player B wins round } n+1. \end{cases}$$

As soon as X_n hits one of the boundary points $\{0, S\}$, the process remains frozen at that state over time. In other words, the game ends whenever the fortune of any of the two players reaches 0, in which case the other player's account contains S.

Among the main issues of interest are:

- the probability that player A (or B) gets eventually ruined (Ruin Probabilities)

- the mean duration of the game.

According to the above problem description, for all $n \in N$ we have

$$P(X_{n+1} = k+1 | X_n = k) = p \text{ and } P(X_{n+1} = k-1 | X_n = k) = q,$$

$1 \leq k \leq S-1$, and in this case the chain is said to be time homogeneous since the transition probabilities do not depend on the time index n. Since we will not focus on the behaviour of the chain after it hits states 0 or N, the law of X_{n+1} given $\{X_n = 0\}$ or $\{X_n = S\}$ can be left unspecified.

The probability space Ω corresponding to this experiment could be taken as

$$\Omega = \{-1, +1\}^N,$$

with any element $\omega \in \Omega$ represented by a countable sequence of $+1$ or -1, depending whether the process goes up or down at each time step. However here we will not focus on this particular representation.