Review of Probability and Statistics

I. Random Variables and Probability Distributions

A. Some Basic Concepts

1. Sample Space - a set (often denoted by $\Omega$) which includes all possible outcomes of the random experiment of interest.

(a) Example 1: flipping a coin once, $\Omega = \{H, T\}$.

(b) Example 2: rolling a die once, $\Omega = \{1, 2, 3, 4, 5, 6\}$.

(c) Example 3: commuting time from home to work, $\Omega = \{\omega : 0 \leq \omega \leq 3\}$.

2. Event - some subset of the sample space $\Omega$.

(a) Example 1 (cont’): Possible events are $\{H\}$, $\{T\}$, $\{H \text{ or } T\}$. 
(b) Example 2 (con’t): Possible events are:
{1}, {2}, ..., {6}, {2 or 3}, {an odd
(c) Example 3 (con’t): Possible events are: (0, 1/2], [1, 2], etc.

3. Probability (measure) - a set function $P(\cdot)$ defined on subsets of the sample space $\Omega$ (i.e., defined on events), which assigns numerical values called probabilities to these subsets (or events).

Note: If we take a frequentist interpretation, the probability of an outcome of a random experiment is the proportion of times that this outcome occurs when the experiment is repeated a large number of times.

4. We usually requires a probability measure to satisfy three axioms.
   (i) $P(A) \geq 0$ for all events $A \subset \Omega$. 
   (ii) $P(\Omega) = 1$. 
(iii) Let $A_1, A_2, \ldots$ be a possibly infinite sequence of disjoint sets (i.e., $A_j \cap A_k = \emptyset$ for $j \neq k$). Then,

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$$

5. **Remark**: Axiom (iii) allows us to compute the probability of compound events. For example, consider the case of rolling a single die once:

$$P(\text{getting an odd #}) = P(1 \cup 3 \cup 5)$$

$$= P(1) + P(3) + P(5)$$

$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$

$$= \frac{1}{2},$$

where the second equality above follows from axiom (iii).
B. Random Variables

1. Definition - A random variable is a variable whose value is determined at least in part by the elements of chance or by some random process.

2. Two Types of Random Variables
   (a) Discrete Random Variable - a random variable which can take on only a finite, or at most, a countably infinite number of values.

Examples:
   (i) rolling a die
   (ii) flipping a coin
   (iii) decision to buy or not to buy a car
(b) **Continuous Random Variable** - A random variable which can take on an uncountably number of possible values. For example, a continuous random variable may take on any value on the real number line or any interval of the real line.

**Examples:**

(i) height of a human being
(ii) GDP
(iii) stock returns
C. Probability Distributions

1. Discrete Case: Probability distribution of a **discrete** random variable is a listing of the possible values that the random variable can assume along with the probability associated with each value.

(a) **Example 1**: flipping an unfair coin

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Head)</td>
<td>$\frac{2}{3}$</td>
</tr>
<tr>
<td>0 (Tail)</td>
<td>$\frac{1}{3}$</td>
</tr>
</tbody>
</table>

\[1\]
(b) **Example 2**: rolling a (balanced) die

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>5</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>6</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
(c) **Remark:** Note that for a discrete random variable, axiom (iii) allows us to calculate the probability of a compound event in terms of the probability of individual outcomes (or simple events) which make up the compound event.

**Example 2 (con’t):** Suppose we roll a (balanced) die once; then,

\[
P(\text{getting a number greater than 4})
\]

\[
= P(5 \cup 6)
\]

\[
= P(5) + P(6)
\]

\[
= \frac{1}{6} + \frac{1}{6}
\]

\[
= \frac{1}{3}.
\]
(d) A Special Discrete Distribution - Bernoulli Distribution

(i) Definition - A discrete random variable \( X \) is said to have a Bernoulli distribution if its probability distribution is given by

\[
X = \begin{cases} 
1 & \text{with prob. } p \\
0 & \text{with prob. } 1 - p
\end{cases}
\]

Here, \( p \) can either be a known constant or an unknown parameter which is to be estimated from the data.

(ii) Remark: We have already seen that a Bernoulli random variable can be used to model the activity of flipping a coin by letting 1 denote the event of getting a "Head" and 0 denote the event of getting a "Tail." It can also be used to model various decision or choice problems such as a
consumer’s decision to purchase or not to purchase a house.

2. Continuous Case: Since a continuous random variable can take on an uncountably infinite number of possible values, it is impossible to list the probabilities for all possible events as we do for simple discrete random variables. Moreover, in this case the probability of a simple event (i.e., the event that the random variable takes on a specific value) must be zero. Hence, we also can no longer do probability calculations for compound events by simply relying on axiom (iii). To compute probabilities in the continuous case, we must introduce a special function $f(x)$, known as the **probability density function** (pdf), such that

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

for real constants $a$ and $b$ such that $a \leq b$. 
(a) **Conditions on** $f(x)$: The pdf $f(x)$ is usually taken to satisfy the following two conditions

(i) $f(x) \geq 0$, for all $x$,

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

(b) **Remark**: Note that the height of the function $f(\cdot)$ does not denote probability, so it is possible that $f(a) > 1$ for some real number $a$. In general,

$$f(a) \neq P(X = a) = 0.$$  
Probability is now the area under the curve $f(x)$.

(c) **Proposition**: If $X$ is a continuous random variable and $a$ and $b$ are real constants such that $a \leq b$, then

$$P(a \leq X \leq b) = P(a < X \leq b)$$

$$= P(a \leq X < b)$$

$$= P(a < X < b)$$
3. **Cumulative Distribution Function (cdf)**

(a) **Discrete Case:** If $X$ is a discrete random variable, then its cdf is given by

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

(b) **Continuous Case:** If $X$ is a continuous random variable, then its cdf is given by

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

(c) **Some Properties of cdf:**

(i) $0 \leq F(x) \leq 1$ for all $x$;

(ii) $F(x)$ is a nondecreasing function as $x$ increases;

(iii) $\lim_{x \to -\infty} F(x) = 0$;

(iv) $\lim_{x \to \infty} F(x) = 1$;

(v) If $X$ is a **discrete** random variable, then $F(x)$ is a step function and the height of the step at $x$ is equal to the probability $P(X = x)$. 
vi If $X$ is a continuous random variable, then $F(x)$ is a 
continuous function.

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

II. Expectations

A. Expected Value or Mean: The expected value $E(X)$ is the “average value" of a 
distribution. It is perhaps the most 
well-known measure of the central 
tendency of a distribution.

1. Discrete Case: Given a discrete 
random variable $X$ which takes on the 
possible values $x_1, \ldots, x_J$; the expected 
value or mean of $X$ is defined as

$$E(X) = \sum_{j=1}^{J} x_j P(X = x_j)$$
(a) **Example 1**: Suppose $X$ has the following discrete distribution

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-2$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>$-1$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>$0$</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>$1$</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>$2$</td>
<td>$\frac{1}{6}$</td>
</tr>
</tbody>
</table>

In this case,

$$E(X) = -2\left(\frac{1}{4}\right) - 1\left(\frac{1}{4}\right) + 0\left(\frac{1}{6}\right) + 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) = -\frac{1}{4}.$$
(b) **Example 2**: Let $X$ have a 
Bernoulli distribution such that 
$P(X = 1) = p$ and 
$P(X = 0) = 1 - p$; then, 
$$E(X) = 1 \cdot p + 0 \cdot (1 - p) = p.$$ 

2. Continuous Case: Let $X$ be a 
continuous random variable with 
probability density function $f(x)$; then, 
the expected value or mean of $X$ is 
given by 
$$E(X) = \int_{-\infty}^{\infty} xf(x) \, dx$$ 

**B. Expectation of Functions of a Random Variable**: Given a random variable $X$, we can create a new random variable 
$$Y = g(X),$$ 
which is a function of the original random variable, and consider its expectation, i.e., 
$$E[Y] = E[g(X)].$$
1. Discrete Case: Let $X$ be a discrete r.v. which takes on the possible values $x_1, \ldots, x_J$; then expected value of $Y = g(X)$ is given by

$$E[g(X)] = \sum_{j=1}^{J} g(x_j)P(X = x_j)$$

2. Continuous Case: Let $X$ be a continuous random variable with probability density function $f(x)$; then, the expected of $Y = g(X)$ is given by

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$
3. **Examples:**

   (a) Let $X$ have a Bernoulli distribution such that
   
   $P(X = 1) = p$ and
   
   $P(X = 0) = 1 - p$, and let
   
   $g(X) = X^2$.

   Then,
   
   $E[g(X)] = E[X^2]$
   
   $= 1^2 \cdot p + 0^2 \cdot (1 - p)$
   
   $= p$. 

(b) Let $X$ be the same Bernoulli r.v. but let
\[ g(X) = \ln(X + 1). \]
Then,
\[
E[g(X)] = E[\ln(X + 1)]
\]
\[
= \ln(2) \cdot p + \ln(1) \cdot (1 - p)
\]
\[
= \ln(2) \cdot p + 0 \cdot (1 - p)
\]
\[
= p \ln 2.
\]
4. Some Facts about Expectations:

(a) \( E(c) = c \) for any constant \( c \).

(i) **Example:** \( E(2) = 2 \).

(ii) **Note:** This implies that \( E[E(X)] = E(X) \). Hence, to take the Bernoulli example again letting \( p = 0.5 \), we see that

\[
\begin{align*}
E[E(X)] & \\
= E[p] & \\
= E[0.5] & \\
= 0.5 & \\
= E(X)
\end{align*}
\]
(b) For any constants $a$ and $b$, let
\[ g(X) = a + bX; \]
then,
\[ E[g(X)] = E[a + bX] = a + bE(X). \]
This is true regardless of whether $X$ is discrete or continuous.

(c) **Remark:** Note that the result in part (a) is a special case of the result in part (b) where $a = c$ and $b = 0$, so we only give a proof of part (b) below.

(d) **Proof of part (b) in discrete case:**
\[ E[a + bX] \]
\[ = \sum_{j=1}^{J} (a + bx_j)P(X = x_j) \]
\[ = \sum_{j=1}^{J} aP(X = x_j) + \sum_{j=1}^{J} bx_jP(X = x_j) \]
\[ = a \sum_{j=1}^{J} P(X = x_j) + b \sum_{j=1}^{J} x_jP(X = x_j) \]
\[ = a + bE[X]. \]

5. **Variance and Standard Deviation**
Variance and standard deviation measure the dispersion, variability, or the “spread” of a probability distribution.

(a) **Definition of Variance:**
\[ \sigma_X^2 = Var(X) = E \{(X - E[X])^2\} \]

(b) **Remark:** Note that variance has units which are the units of the square of \(X\), so it might be difficult to interpret.

(c) **Definition of Standard Deviation:**
\[ \sigma_X = \sqrt{\text{Var}(X)} = \sqrt{E\{(X - E[X])^2\}} \]
(d) **Example:** Consider the following discrete distribution

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1$</td>
<td>$\frac{1}{8}$</td>
</tr>
<tr>
<td>$0$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>$2$</td>
<td>$\frac{3}{8}$</td>
</tr>
<tr>
<td>$1$</td>
<td></td>
</tr>
</tbody>
</table>

In this case,

$$E[X] = (-1) \cdot \left(\frac{1}{8}\right) + 0 \cdot \left(\frac{1}{2}\right) + (2) \cdot \left(\frac{3}{8}\right)$$

$$= \frac{5}{8}$$

and, thus,
\[
\sigma^2_X = Var(X) \\
= (-1 - 5/8)^2 \cdot (1/8) + (0 - 5/8)^2 \cdot (1/2) \\
+ (2 - 5/8)^2 \cdot (3/8) \\
= 79/64
\]

\[
\sigma_X = \sqrt{Var(X)} \\
= \sqrt{79}/8
\]

(e) Useful Representation:

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

Proof:

\[
Var(X) = E \{(X - E[X])^2\} \\
= E\{X^2 - 2XE[X] + (E[X])^2\} \\
= E[X^2] - 2E[X]E[X] + E\{(E[X])^2\} \\
= E[X^2] - 2(E[X])^2 + (E[X])^2 \\
= E[X^2] - (E[X])^2
\]
(f) **An Application:** Let $X$ have a Bernoulli distribution such that $P(X = 1) = p$ and $P(X = 0) = 1 - p$. (For example, suppose that the r.v. $X$ is used to model whether you will get a virus when you click on an e-mail.)

Previously, we have shown that

$$E[X] = 1 \cdot p + 0 \cdot (1 - p) = p,$$

$$E[X^2] = 1^2 \cdot p + 0^2 \cdot (1 - p) = p$$

It follows that

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

$$= p - p^2$$

$$= p(1 - p).$$
Remark: \( Var(X) \geq 0 \) and \( Var(X) = 0 \) if and only if there is constant \( c \) such that

\[
P(X = c) = 1,
\]
in which case

\[
E[X] = c.
\]

What this says is that the variance of any constant is zero and if a r.v. has zero variance, then it is essentially a constant.

Example: \( Var(4) = 0, \)
\( Var(E[X]) = 0 \) (why?)
Useful Result: For constants $a$ and $b,$

$$Var(a + bX) = b^2 Var(X).$$

Hence, adding a constant to a r.v. does not change the variance, but multiplying a r.v. by a constant increases (decreases) the variance by a factor equal to the square of the constant if $|b| > 1$ (if $|b| < 1$).

Proof:

$$Var(a + bX)$$

$$= E \left\{ (a + bX - E[a + bX])^2 \right\}$$

$$= E \left\{ (a + bX - a - bE[X])^2 \right\}$$

$$= E \left\{ (b[X - E[X]])^2 \right\}$$

$$= b^2 E \left\{ (X - E[X])^2 \right\}$$

$$= b^2 Var(X).$$

Examples: $Var(2X) = 4Var(X),$ $Var(6 + X) = Var(X).$
6. **Raw Moments and Centered Moments**

(a) The $r^{th}$ (raw) moment of a r.v. $X$ is defined as $E[X^r]$.

(b) **Examples:**

(i) $E[X]$ - 1st moment or mean of $X$;

(ii) $E[X^2]$ - 2nd (raw) moment of $X$ or 2nd moment of $X$ about the origin.

(c) The $r^{th}$ central moment of a r.v. $X$ is defined as $E\{(X - E[X])^r\}$, i.e., it is the $r^{th}$ moment taken about the mean.

(d) **Example:** $Var(X) = E\{(X - E[X])^2\}$ - 2nd central moment or 2nd moment taken about the mean.
(e) The moments measures certain properties of a distribution:

(i) As discussed before, the 1st moment is a measure of central tendency.

(ii) The 2nd central moment is a measure of dispersion.

(iii) The 3rd central moment tells us about skewness, i.e., how much asymmetry there is in the distribution.

(iv) The fourth central moment tells us about kurtosis, i.e., how peaked or flat is a distribution.

(f) **Remark:** It is possible for a moment “not to exist” if the probability of getting extreme values is too high, i.e., if the tails of the probability distribution is too thick.
III. Joint, Marginal, and Conditional Probability Distributions

So far, we have only discussed univariate probability distributions. However, in economics, we are typically interested in the occurrence of events involving more than one r.v. For example, how does the income distribution of women compare with that of men? Do college graduates earn more than high school dropouts? To answer these questions, we would have to have a mathematical model for the distribution of more than one r.v.’s. To keep the discussion simple, we will focus on the case with two r.v.’s or the bivariate case.
Example of Bivariate Distribution: (Discrete Case)
Suppose that a family is planning to have three children. Let

\[ X = \text{number of girls among the three children}, \]

\[ Y = \begin{cases} 
1 & \text{if 1st child is a girl} \\
0 & \text{otherwise}
\end{cases} \]

Sample Space:
<table>
<thead>
<tr>
<th>Outcome</th>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBB</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BBG</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>BGB</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>GBB</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BGG</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>GBG</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>GGB</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>GGG</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
Joint (Bivariate) Probability Distribution

<table>
<thead>
<tr>
<th>((x, y))</th>
<th>(p(x, y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0)</td>
<td>1/8</td>
</tr>
<tr>
<td>(1, 0)</td>
<td>1/8</td>
</tr>
<tr>
<td>(1, 1)</td>
<td>1/8</td>
</tr>
<tr>
<td>(2, 0)</td>
<td>1/8</td>
</tr>
<tr>
<td>(2, 1)</td>
<td>1/8</td>
</tr>
<tr>
<td>(3, 1)</td>
<td>1/8</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Remark: \(p(x, y)\) is sometimes called the joint probability function. Note that, evaluated at the ordered pair \((x, y)\), \(p(x, y)\) is simply the probability of the event

\[ \{X = x \cap Y = y\}, \]

i.e.,

\[ p(x, y) = P(X = x \cap Y = y). \]