Chapter 4: Random Variables and Probability Distributions

Section 1: Two Types of Random Variables

Most experiments have sample points that correspond to values of some numerical variable.

Ex.: when you toss a coin 10 times, you are interested in the number of heads/tails not the exact sequence
Ex.: when you randomly choose a person, you are not interested in who it is, but rather his height/weight/etc.

A random variable is a variable that assumes numerical values associated with the random outcomes of an experiment, where one (and only one) numerical values is assigned to each sample point.

Random variables that can assume a countable number (finite or infinite) of values are called discrete. If we can list the values of a random variable \( x \), even though the list is never ending, we call the list countable.

Random variables that can assume values corresponding to any of the points contained in one or more intervals (i.e., values that are infinite and uncountable) are called continuous. For any two values of a continuous random variable, there are an infinite number of other possible values in between.

Section 2: Probability Distributions for Discrete Random Variables

A complete description of a discrete random variable requires that we specify the possible values the random variable can assume and the probability associated with each value.

The probability distribution of a discrete random variable is a graph, table, or formula that specifies the probability associated with each possible value the random variable can assume.

Requirements for the Probability Distribution of a Discrete Random Variable, \( x \)

1. \( p(x) \geq 0 \) for all values of \( x \)
2. \( \Sigma p(x) = 1 \) where the summation of \( p(x) \) is over all possible values of \( x \).

Because probability distributions are analogous to the relative frequency distribution of Chapter 2, it should be no surprise that the mean and standard deviation are useful descriptive measures.

If a discrete random variable \( x \) were observed a very large number of times, and the data generated were arranged in a relative frequency distribution, the relative frequency distribution would be indistinguishable from the probability distribution for the random variable. Thus, the probability distribution for a random variable is a theoretical model for the relative frequency distribution of a population.

Give examples about grades on exams and how computing the mean is equivalent to taking expected value
The mean, or expected value, of a discrete random variable \( x \) is \( \mu = E(x) = \Sigma xp(x) \).

Do not think about everyday meaning of expected: a random variable may never equal its expected value. You can think of \( \mu \) as the mean value of \( x \) in a very large (actually infinite) number of repetition of the experiment.

The variance of a discrete random variable \( x \) is \( \sigma^2 = E[(x - \mu)^2] = \Sigma(x - \mu)^2p(x) \).

The standard deviation of a discrete random variable is equal to the square root of the variance, i.e., \( \sigma = \sqrt{\sigma^2} = \sqrt{\Sigma(x - \mu)^2p(x)} \).

Properties: For any constants \( a \) and \( b \) and any random variable \( X \), we have:

\[
E(aX+b) = aE(X)+b \\
Var(aX+b) = a^2Var(X) \\
SD(aX+b) = \left|a\right|SD(X)
\]
Section 3: The Binomial Distribution

Combinations Rule
A sample of \( n \) elements is to be drawn from a set of \( N \) elements. Then, the number of different samples possible is denoted by \( \binom{N}{n} \) and is equal to \( \frac{N!}{n!(N-n)!} \) where the factorial symbol means that \( n! = n(n-1)(n-2)\cdots(3)(2)(1) \).

For example, \( 5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \). [Note: The quantity 0! is defined to be equal to 1.]

(Also briefly mention permutations ?

Many experiments result in dichotomous responses – that is, responses for which there exist two possible alternatives. Give examples of such experiments and their responses

Characteristics of a Binomial Experiment
1. The experiment consists of \( n \) identical trials.
2. There are only two possible outcomes on each trial. We will denote one outcome by \( S \) (for success) and the other by \( F \) (for failure).
3. The probability of \( S \) remains the same from trial to trial. This probability is denoted by \( p \), and the probability of \( F \) is denoted by \( q \). Note that \( q = 1 - p \).
4. The trials are independent.
5. The binomial random variable \( x \) is the number of \( S \)s in \( n \) trials.

The Binomial Probability Distribution
\[
p(x) = \binom{n}{x} p^x q^{n-x} \quad (x = 0, 1, 2, \ldots, n)
\]
where
\[
p = \text{Probability of a success on a single trial}
\]
\[
q = 1 - p \quad \text{Probability of a failure on a single trial}
\]
\[
n = \text{Number of trials}
\]
\[
x = \text{Number of successes in } n \text{ trials}
\]
\[
n - x = \text{Number of failures in } n \text{ trials}
\]
\[
\binom{n}{x} = \frac{n!}{x!(n-x)!}
\]

Mean, Variance, and Standard Deviation for a Binomial Random Variable
Mean: \( \mu = np \)
Variance: \( \sigma^2 = npq \)
Standard deviation: \( \sigma = \sqrt{npq} \)

Suppose there is a large dichotomous population and a sample is drawn from it, and we look at \( X \) the number of success in the sample. If the sample is drawn without replacement then clearly \( X \) is not binomial.
However, if the sample size is small relative to the population size, binomial probabilities provide a good approximation. In this case, in practice \( X \) is modeled as a binomial.

* Mention the fact that the sum of two binomial random variables may or may not be binomial through the examples of (coin + die) and (coin + coin)

Section 4: Other Discrete Distributions: Poisson and Hypergeometric

Characteristics of a Hypergeometric Random Variable
1. The experiment consists of randomly drawing \( n \) elements without replacement from a set of \( N \) of elements, \( r \) of which are \( S \)’s (for success) and \( (N-r) \) of which are \( F \)’s (for failure).
2. The hypergeometric random variable \( x \) is the number of \( S \)’s in the draw of \( n \) elements.
Probability Distribution, Mean, and Variance of the Hypergeometric Random Variable

\[ p(x) = \frac{(r)(N-r)}{(s)(N-s)} \]

\[ \mu = \frac{nr}{N} \]

\[ N = \text{Total number of elements} \]
\[ r = \text{Number of S's in the n elements} \]
\[ n = \text{Number of elements drawn} \]
\[ x = \text{Number of S's drawn in the n elements} \]

\[ \sigma^2 = \frac{r(N-r)n(N-n)}{N^2(N-1)} \]

Briefly mention the following distributions: Geometric, Poisson.

Section 5: Probability Distributions for Continuous Random Variables

A random variable \( X \) is continuous if \( P(X = x) = 0 \) for all \( x \).

The probability distribution for a continuous random variable, \( x \), can be represented by a smooth curve – a function of \( x \), denoted \( f(x) \). The curve is called a density function or frequency function. The probability that \( x \) falls between two values, \( a \) and \( b \), i.e., \( P(a < x < b) \), is the area under the curve between \( a \) and \( b \).

A density curve is a mathematical model of a distribution.
The total area under the curve, by definition, is equal to 1, or 100%.
The area under the curve for a range of values is the probability of all observations for that range.

Section 6: The Normal Distribution

Describes many random processes or continuous phenomena
Can be used to approximate discrete probability distributions (Example: Binomials when \( n \) is large)
Basis for classical statistical inference

"Bell-shaped" and symmetrical; Mean, median are equal

Probability Distribution for a Normal Random Variable \( x \)

Probability density function: 
\[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

\[ \mu = \text{Mean of the normal random variable} \ x \]
\[ \sigma = \text{Standard deviation} \]
\[ \pi = 3.1415... \]
\[ e = 2.71828... \]

The standard normal distribution is a normal distribution with \( \mu = 0 \) and \( \sigma = 1 \). A random variable with a standard normal distribution, denoted by the symbol \( z \), is called a standard normal random variable.

Converting a Normal Distribution to a Standard Normal Distribution

If \( X \) is a normal random variable with mean \( \mu \) and standard deviation \( \sigma \), then the random variable \( Z \), defined by the formula 
\[ Z = \frac{X-\mu}{\sigma} \]
has a standard normal distribution.
The value \( Z \) is the number of standard deviations between \( X \) and \( \mu \).

Briefly mention the following distributions: Uniform, Exponential.