7.1: Discrete and Continuous Random Variables

RANDOM VARIABLE
A random variable is a variable whose value is a numerical outcome of a random phenomenon.

DISCRETE RANDOM VARIABLE
A discrete random variable $X$ has a countable number of possible values. The probability distribution of $X$ lists the values and their probabilities.

<table>
<thead>
<tr>
<th>Value of $X$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$\ldots$</th>
<th>$x_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>$p_1$</td>
<td>$p_2$</td>
<td>$p_3$</td>
<td>$\ldots$</td>
<td>$p_k$</td>
</tr>
</tbody>
</table>

The probabilities $p_i$ must satisfy two requirements:

1. Every probability $p_i$ is a number between 0 and 1.
2. $p_1 + p_2 + \ldots + p_k = 1$

Find the probability of any event by adding the probabilities $p_i$ of the particular values $x_i$ that make up the event.

CONTINUOUS RANDOM VARIABLE
A continuous random variable $X$ takes all values in an interval of numbers. The probability distribution of $X$ is described by a density curve. The probability of any event is the area under the density curve and above the values of $X$ that make up the event.

Example Problem:

1. Patients receiving artificial knees often experience pain after surgery. The pain is measured on a subjective scale with possible values of 1 to 5. Assume that $X$ is a random variable representing the pain score for a randomly selected patient. The following table gives part of the probability distribution for $X$.

<table>
<thead>
<tr>
<th>$X$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X)$</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>?</td>
</tr>
</tbody>
</table>

(a) Find $P(X = 5)$.
(b) Find the probability that the pain score is less than 3.
(c) Find the probability that the pain score is greater than 1.
(d) Find the probability that the pain score is either 2 or 4.
### 7.2: Means and Variances of Random Variables

#### Mean of a Discrete Random Variable

Suppose that X is a discrete random variable whose distribution is

<table>
<thead>
<tr>
<th>Value of X</th>
<th>x₁</th>
<th>x₂</th>
<th>x₃</th>
<th>…</th>
<th>xₖ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>p₁</td>
<td>p₂</td>
<td>p₃</td>
<td>…</td>
<td>pₖ</td>
</tr>
</tbody>
</table>

To find the mean of X, multiply each possible value by its probability, then add all the products:

$$\mu_x = x_1 \cdot p_1 + x_2 \cdot p_2 + \ldots + x_k \cdot p_k$$

#### Standard Deviation of a Discrete Random Variable

Suppose that X is a discrete random variable whose distribution is

<table>
<thead>
<tr>
<th>Value of X</th>
<th>x₁</th>
<th>x₂</th>
<th>x₃</th>
<th>…</th>
<th>xₖ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>p₁</td>
<td>p₂</td>
<td>p₃</td>
<td>…</td>
<td>pₖ</td>
</tr>
</tbody>
</table>

and $\mu$ is the mean of X. The variance of X is

$$\sigma_x^2 = (x_1 - \mu_x)^2 \cdot p_1 + (x_2 - \mu_x)^2 \cdot p_2 + \ldots + (x_k - \mu_x)^2 \cdot p_k$$

and the standard deviation is the square root of the previous result.

#### Law of Large Numbers

Draw independent observations at random from any population with finite mean $\mu$. As the number of observations drawn increases, the mean of the observed values eventually approaches the mean $\mu$.

#### Example Problem:

2. Patients receiving artificial knees often experience pain after surgery. The pain is measured on a subjective scale with possible values of 1 to 5. Assume that X is a random variable representing the pain score for a randomly selected patient. The following table gives part of the probability distribution for X.

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X)$</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>.1</td>
</tr>
</tbody>
</table>

(a) Find the mean $\mu$ for this distribution
(b) Find the standard deviation for this distribution.
8.1: The Binomial Distributions

THE BINOMIAL SETTING
1. Each observation falls into one of just two categories, which for convenience we call “success” and “failure.”
2. There is a fixed number \( n \) of observations
3. The \( n \) observations are independent. That is, knowing the result of one observation tells you nothing about the other observations.
4. The probability of success, call it \( p \), is the same for each observation.

BINOMIAL DISTRIBUTION
The distribution of the count \( X \) of successes in the binomial setting is the binomial distribution with parameters \( n \) and \( p \). The parameter \( n \) is the number of observations, and \( p \) is the probability of success on any one observation. The possible values of \( X \) are the whole numbers from 0 to \( n \). As an abbreviation we say that \( X \) is \( B(n, p) \).

BINOMIAL PROBABILITY
If \( X \) has the binomial distribution with \( n \) observations, and probability \( p \) of success on each observation, the possible values of \( X \) are 0, 1, 2, … \( n \). If \( k \) is any one of these values, \( P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \)

MEAN AND STD. DEVIATION OF A BINOMIAL RANDOM VARIABLE
If a count \( X \) has the binomial distribution with \( n \) observations, and probability of success \( p \), the mean (\( \mu \)) and standard deviation (\( \sigma \)) of \( X \) are:

\[ \mu = np \]
\[ \sigma = \sqrt{np(1-p)} \]

Example Problem:
3. Would most wives marry the same man again, if given the chance? According to a 1988 survey conducted by Ladies Home Journal, 80% of women would in fact marry their husband again. To test this claim, you randomly select 6 married women and ask them whether they would marry their husband again. Define \( X \) as the number in the sample that respond yes.

(a) What is the probability that exactly that 4 of the women say yes?
(b) What is the probability that fewer than 3 of the women say yes?
(c) What is the probability that more than 3 of the women say yes?
(d) What is the expected value of \( X \)?
8.2: The Geometric Distributions

THE GEOMETRIC SETTING
1. Each observation falls into one of just two categories, which for convenience we call “success” and “failure.”
2. The probability of success, call it $p$, is the same for each observation.
3. The observations are all independent.
4. The variable of interest is the number of trials required to obtain the first success.

GEOMETRIC PROBABILITY
If $X$ has the geometric distribution with probability $p$ of success and $(1 - p)$ of failure on each observation, the possible values of $X$ are 1, 2, 3, .... If $n$ is any one of these values, the probability that the first success occurs in the $n$th trial is

$$P(X = n) = (1 - p)^{n-1} \cdot p$$

MEAN OF A GEOMETRIC RANDOM VARIABLE
If $X$ is a geometric random variable with probability of success $p$ on each trial, then the mean ($\mu$) of $X$ (also called the expected value of $X$) is:

$$\mu = \frac{1}{p}$$

$P(X>n)$
The probability that it takes more than $n$ trials to see the first success in a geometric setting is,

$$P(X > n) = (1 - p)^n$$

Example Problem:

4. Would most wives marry the same man again, if given the chance? According to a 1988 survey conducted by *Ladies Home Journal*, 80% of women would in fact marry their husband again. To test this claim, you randomly select married women and ask them whether they would marry their husband again. Define $X$ as the number of women you ask until you get a yes answer.

(a) What is the probability that the first yes comes from the 3rd woman you ask?
(b) What is the probability that it takes fewer than 3 women to get a yes answer?
(c) What is the probability that it takes more than 3 of the women to get a yes answer?
(d) What is the expected value of $X$?
1. This problem concerns working with the properties of a discrete probability distribution. The distribution is redisplayed below.

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X)</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>?</td>
</tr>
</tbody>
</table>

(a) Find \( P(X = 5) \).

Since the sum of the probabilities of any distribution must be 1, simply add the other probabilities and subtract from 1. So \( 1 - (.1 + .2 + .3 + .3) = 1 - .9 = .1 \)

(b) Find the probability that the pain score is less than 3.

Less than 3 only includes when \( x=1 \) or \( x=2 \), so we add together those probabilities: \( .1 + .2 = .3 \)

(c) Find the probability that the pain score is greater than 1.

Greater than 1 includes all probabilities for \( x=2 \) to \( x=5 \), so we add those probabilities: \( .2 + .3 + .3 + .1 = .9 \)

(d) Find the probability that the pain score is either 2 or 4.

Adding those probabilities we get: \( .2 + .3 = .5 \)

2. This problem concerns the same discrete probability distribution from #1 (see above).

(a) Find the mean \( \mu \) for this distribution

The formula for the mean of a discrete r.v. is \( \mu_x = x_1 \cdot p_1 + x_2 \cdot p_2 + \ldots + x_k \cdot p_k \)

Applying that formula, we get: \( 1(.1) + 2(.2) + 3(.3) + 4(.3) + 5(.1) = 3.1 \)

(b) Find the standard deviation for this distribution.

The formula for the standard deviation of a discrete random variable is the square root of \( (x_1 - \mu)^2 \cdot p_1 + (x_2 - \mu)^2 \cdot p_2 + \ldots + (x_k - \mu)^2 \cdot p_k \)

Applying that formula, we get: \( (1-3.1)^2(.1) + (2-3.1)^2(.2) + (3-3.1)^2(.3) + (4-3.1)^2(.3) + (5-3.1)^2(.1) = 1.29 \) Lastly, we take the square root: \( \sqrt{1.29} = 1.136 \)
3. We are working with a binomial random variable with $n=6$ and $p=.8$.

(a) What is the probability that exactly that 4 of the women say yes?
$$P(X=4) = \binom{6}{4}(.8)^4(.2)^2 = \text{binompdf}(6,.8,4) = .24576$$

(b) What is the probability that fewer than 3 of the women say yes?
Fewer than 3, means less than 3 so we need $P(X<3)$ which can be found using the binomcdf function as follows: $\text{binomcdf}(6,.8,2) = .01696$

(c) What is the probability that more than 3 of the women say yes?
More than 3, means we need $P(X>3)$ which can be found using the binomcdf function as follows: $1-\text{binomcdf}(6,.8,3) = .90112$

(d) What is the expected value of $X$?
The formula for the expected value or mean of a binomial random variable is $\mu = np$, so the mean of this variable is $\mu = (6)(.8) = 4.8$.

4. We are working with a geometric random variable with $p=.8$.

(a) What is the probability that the first yes comes from the 3rd woman you ask?
$$P(X=3) = (.2)^2(.8)^1 = \text{geometpdf}(.8,3) = .032$$

(b) What is the probability that it takes fewer than 3 women to get a yes answer?
Fewer than 3, means less than 3 so we need $P(X<3)$ which can be found using the geometcdf function as follows: $\text{geometcdf}(.8,2) = .96$

(c) What is the probability that it takes more than 3 of the women to get a yes answer?
More than 3, means we need $P(X>3)$ which can be found using the geometcdf function as follows: $1-\text{geometcdf}(.8,3) = .008$. Alternatively, we can solve the problem using the formula $(.2)^3 = .008$.

(d) What is the expected value of $X$?
The formula for the expected value or mean of a geometric random variable is $\mu = 1/p$, so the mean of this variable is $\mu = 1 / .8 = 1.25$. 