5.2 RANDOM VARIABLES

DEFINITION

A **random variable** is a variable (typically represented by X) that has a single numerical value, determined by chance, for each outcome of a procedure.

DEFINITION

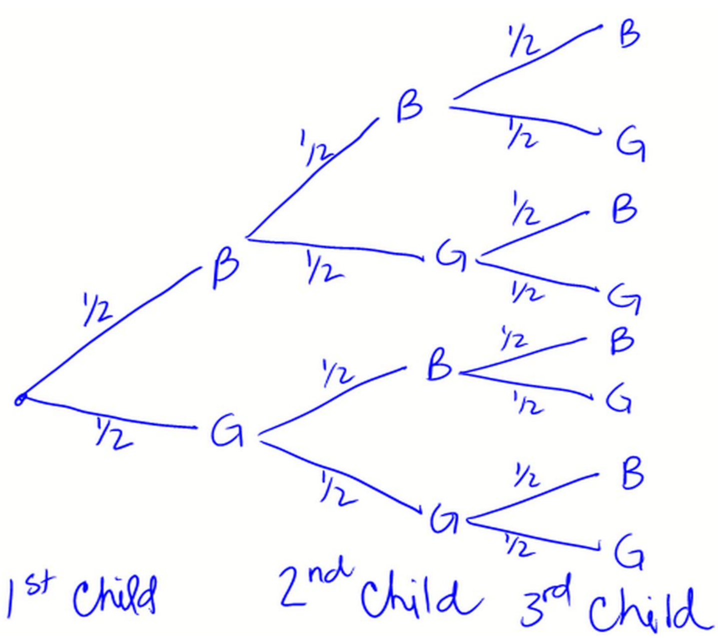
A **probability distribution** is a description that gives the probability for each value of the random variable. It is often expressed in the format of a graph, table, or formula.

NOTE

If a probability value is very small, such as 0.000000123, we can represent it as 0+ in a table, where 0+ indicates that the probability value is a very small positive number. Why not represent this as 0?

Because the event is still possible and a probability of zero means the event is impossible.

Recall the tree diagram we made for a couple having 3 children:



Let X be the # of girls the couple has

X	P(X)
0	$P(X=0)=1/8$
1	$P(X=1)=3/8$
2	$P(X=2)=3/8$
3	$P(X=3)=1/8$

DEFINITION

A **discrete random variable** has either a finite number of values or a countable number of values, where countable refers to the fact that there might be infinitely many values, but they can be associated with a counting process, so that the number of values is 0 or 1 or 2 or 3, etc.

A **continuous random variable** has infinitely many values, and those values can be associated with measurements on a continuous scale without gaps or interruptions.

Example 1: Give two examples of

a. Discrete random variables

(1) The # of wins a team will attain during the regular season of basketball.

(2) The # of crimes that will be committed in Oceanside next year

b. Continuous random variables

(1) The distance (in miles) a person training for a marathon will run before the race.

(2) The time (in seconds) that a song that will be played on the radio will be.

GRAPHS

There are various ways to graph a probability distribution, but we will consider only the probability histogram. A probability histogram is similar to a relative frequency histogram, but the vertical scale shows probabilities instead of relative frequencies based on actual sample events.

REQUIREMENTS FOR A PROBABILITY DISTRIBUTION

1. $\sum P(x) = 1$ where *x* assumes all possible values. The sum of all probabilities must be ___1___, but values such as 0.999 or 1.001 are acceptable because they result from ___rounding___ errors.
2. $0 \le P(x) \le 1$ for every individual value of *x*.

MEAN, VARIANCE, AND STANDARD DEVIATION

Let X be the # of girls of couple who has 3 children

X	P(x)
0	1/8
1	3/8
2	3/8
3	1/8

1. $\mu = \sum [x \cdot P(x)]$ $\quad \mu = 0(1/8) + 1(3/8) + 2(3/8) + 3(1/8)$

   $\mu = \frac{12}{8} = 1.5$

   On average, a couple who has 3 kids will have 1.5 girls.
2. $\sigma^2 = \sum \left[(x-\mu)^2 \cdot P(x)\right]$
3. $\sigma^2 = \sum \left[x^2 \cdot P(x)\right] - \mu^2$
4. $\sigma = \sqrt{\sum \left[x^2 \cdot P(x)\right] - \mu^2}$

$\sigma^2 = 0^2(1/8) + 1^2(3/8) + 2^2(3/8) + 3^2(1/8) - 1.5^2$

$\sigma^2 = 3/8 + 12/8 + 9/8 - 1.5^2 \rightarrow \sigma^2 = 0.75$ so

$\sigma = \sqrt{.75}$

$\boxed{\sigma \approx 0.9 \text{ girls}}$

ROUND-OFF RULE FOR μ, σ, and σ^2

Round results by carrying one more ___decimal___ place than the number of decimal places used for the ___random___ variable ___X___. If the values of ___X___ are ___integers___, round to one decimal place.

IDENTIFYING UNUSUAL RESULTS WITH THE RANGE RULE OF THUMB

The range rule of thumb may be helpful in ___interpreting___ the value of a ___standard___ ___deviation___. According to the ___range___ ___rule___ of ___thumb___, most values should lie within ___2___ standard deviations of the mean; it is unusual for a value to differ from the mean by more than 2 standard deviations.

Maximum usual value = μ + 2σ

Minimum usual value = μ - 2σ

IDENTIFYING UNUSUAL RESULTS WITH PROBABILITIES

x successes among *n* trials is an unusually high number of successes if the probability of X or more successes is unlikely with a probability of 0.05 or less.

x successes among *n* trials is an unusually low number of successes if the probability of X or fewer successes is unlikely with a probability of 0.05 or less.

RARE EVENT RULE FOR INFERENTIAL STATISTICS

If, under a given assumption, the probability of a particular observed event is extremely small, we conclude that the assumption is probably not correct.

Example 2: Based on information from MRINetwork, some job applicants are required to have several interviews before a decision is made. The number of required interviews and the corresponding probabilities are: 1 (0.09); 2 (0.31); 3 (0.37); 4 (0.12); 5 (0.05); 6 (0.05).

a. Does the given information describe a probability distribution?

X	P(X)
1	0.09
2	0.31
3	0.37
4	0.12
5	0.05
6	0.05

$\sum P(x) \stackrel{?}{=} 1$

$\sum P(x) = 0.09 + 0.31 + 0.37 + 0.12 + 0.05 + 0.05$

$= 0.99$ → close enough to 1

$\boxed{\text{Yes, assuming rounding error.}}$

b. Assuming that a probability distribution is described, find its mean and standard deviation.

X	$P(X)$
1	0.09
2	0.31
3	0.37
4	0.12
5	0.05
6	0.05

$\mu = \sum x \cdot P(x)$, $\mu = 1(0.09)+2(0.31)+3(0.37)+4(0.12)+5(0.05)+6(0.05)$

$\boxed{\mu \approx 2.9 \text{ interviews}}$

$\sigma^2 = \sum x^2 \cdot P(x) - \mu^2$, $\sigma^2 = 1^2(0.09)+2^2(0.31)+3^2(0.37)+4^2(0.12)+5^2(0.05)+6^2(0.05) - 2.9^2$

$\sigma^2 \doteq 1.2 \rightarrow \boxed{\sigma = \sqrt{1.2} = 1.1 \text{ interviews}}$

c. Use the range rule of thumb to identify the range of values for usual numbers of interviews.

$\mu - 2\sigma \le \text{usual \# of interviews} \le \mu + 2\sigma$

$2.9 - 2(1.1) \le \text{usual \# of interviews} \le 2.9 + 2(1.1)$

$\boxed{0.7 \le \text{usual \# of interviews} \le 5.1}$

d. Is it unusual to have a decision after just one interview. Explain.

No since one interview falls within the range of usual # of interviews.

DEFINITION

The **expected value** of a discrete random variable is denoted by $E(x)$, and it represents the mean value of the outcomes. It is obtained by finding the value of $\sum[x \cdot P(x)]$.

$$E = \sum[x \cdot P(x)]$$

Example 3: There is a 0.9968 probability that a randomly selected 50-year old female lives through the year (based on data from the U.S. Department of Health and Human Services). A Fidelity life insurance company charges $226 for insuring that the female will live through the year. If she does not survive the year, the policy pays out $50,000 as a death benefit.

a. From the perspective of the 50-year-old female, what are the values corresponding to the two events of surviving the year and not surviving?

If she lives, she pays $226 → $-226

If she dies, the family gets 50000 → 50000 − 226 = $49774

b. If a 50-year-old female purchases the policy, what is her expected value?

X	P(X)
-226	0.9968
49774	0.0032

$E(x) = -226(0.9968) + 49774(0.0032)$

$E(x) = \$-66.$

c. Can the insurance company expect to make a profit from many such policies? Why?

Yes. The insurance company can expect to make an average of $66 per policy of this type per year.

5.3 BINOMIAL PROBABILITY DISTRIBUTIONS

DEFINITION

A **binomial probability distribution** results from a procedure that meets all of the following requirements:

1. The procedure has a fixed number of trials.
2. The trials must be independent.
3. Each trial must have all outcomes classified into 2 categories (commonly referred to as success and failure).
4. The probability of a success remains the same in all trials.

NOTATION FOR BINOMIAL PROBABILITY DISTRIBUTIONS

S and F (success and failure) denote the two possible categories of outcomes

$P(S) = p$

$P(F) = 1 - p = q$

$n \rightarrow$ fixed # of trials

$x \rightarrow$ specific # of successes in n trials, $0 \le x \le n$

$p \rightarrow$ probability of success in 1 trial

$q \rightarrow$ probability of failure in 1 trial

$P(x) \rightarrow$ probability of exactly x successes among the n trials.

Example 1: A psychology test consists of multiple-choice questions, each having four possible answers (a, b, c, and d), one of which is correct. Assume that you guess the answers to six questions.

a. Use the multiplication rule to find the probability that the first two guesses are wrong and the last four guesses are correct. $P(W) = 3/4$, $P(C) = 1/4$

$P(WWCCCC) = \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \approx \boxed{0.00220}$

b. Beginning with WWCCCC, make a complete list of the different possible arrangements of 2 wrong answers and 4 correct answers, then find the probability for each entry in the list.

WWCCCC	CWWCCC	CCWWCC	CCCWWC	CCCCWW
WCWCCC	CWCWCC	CCWCWC	CCCWCW	
WCCWCC	CWCCWC	CCWCCW		
WCCCWC	CWCCCW			
WCCCCW				

15 arrangements

Prob. of each entry ≈ 0.00220

c. Based on the preceding results, what is the probability of getting exactly 4 correct answers when 6 guesses are made?

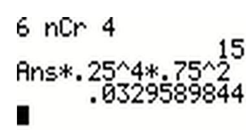
P(exactly 4 C out of 6) = 15(0.00220) ≈ 0.0330

d. Now use the Binomial Probability Formula to find probability of getting exactly 4 correct answers when 6 guesses are made.

$n = 6$ $\quad p = \frac{1}{4}$

$x = 4$ $\quad q = \frac{3}{4}$

$$P(X=4) = {}_6C_4\left(\tfrac{1}{4}\right)^4\left(\tfrac{3}{4}\right)^{6-4} = {}_6C_4\left(\tfrac{1}{4}\right)^4\left(\tfrac{3}{4}\right)^2 \approx 0.0330$$



BINOMIAL PROBABILITY FORMULA

$$P(x) = \frac{n!}{(n-x)!x!} \cdot p^x \cdot q^{n-x} \qquad \text{for } x = 0, 1, 2, \ldots, n$$

$${}_nC_x \, p^x \, q^{n-x}$$

Example 2: Assuming the probability of a pea having a green pod is 0.75, use the binomial probability formula to find the probability of getting exactly 2 peas with green pods when 5 offspring peas are generated.