Math 122
Introduction to Statistics
Random Variables

Random variables and probability distributions can be used to describe what will probably happen in an experiment. In particular, they can be used to determine which outcomes of an experiment are unusual.

PURPOSE

- When doing an experiment, ASSUMPTIONS will affect calculation of PROBABILITY DISTRIBUTIONS.
- PROBABILITY DISTRIBUTIONS can be used to calculate the likelihood of certain events.
- If OBSERVED FREQUENCIES from the experiment match the calculated PROBABILITY DISTRIBUTIONS, then great.
- If OBSERVED FREQUENCIES do not match the calculated PROBABILITY DISTRIBUTIONS, then there is a problem with the ASSUMPTIONS.
- RARE EVENT RULE: If the calculated probability of an observed event is very small, then the ASSUMPTIONS may false.

RANDOM VARIABLES

Random Variable: A random variable is a variable which has a single value for each outcome of an experiment. The value of the random variable is determined by chance by the outcome of the experiment. We usually denote a random variable as x.

Examples: These are some examples of random variables:

- The number of cars that pass a certain intersection during the lunch hour.
- The number which appears when a die is rolled.
- The amount of rainfall in a rain-guage after a storm.
- The height of a randomly selected seven year old.
- The number of Heads seen in the outcome of a coin flip.
- The number of female children among three children.

Discrete: Discrete random variables come from sets of numbers that have gaps between the possible values.

Example: The number of female children among three children is a discrete random variable.

Continuous: Continuous random variables come from sets of numbers that cover a range of numbers with no gaps.

Example: The amount of rain in a rain-guage after a storm is a continuous random variable.

PROBABILITY DISTRIBUTION

Probability Distribution: A probability distribution for a random variable x is a description that gives the probability of each possible value of the random variable.

Probability distributions are usually given as graphs or tables or sometimes formulas.

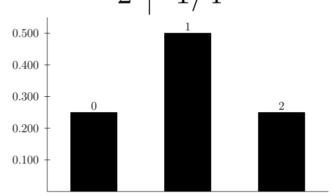
Notation: If x is a random variable and N is a number, we will denote that probability that $x = N$ as $P(x = N)$ or $P(N)$.

Example: Suppose that two coins are flipped and that x is the number of H's that appear. The possible values for x are 0, 1, and 2. This is a discrete random variable. To find the probability distribution for x, we can list all of the outcomes of the experiment along with the associated value for x:

Outcome	x
HH	2
HT	1
TH	1
TT	0

We can then use the classical approach to find the probability of each outcome. We can give this information in a relative frequency table or a histogram:

x	$P(x)$
0	1/4
1	1/2
2	1/4

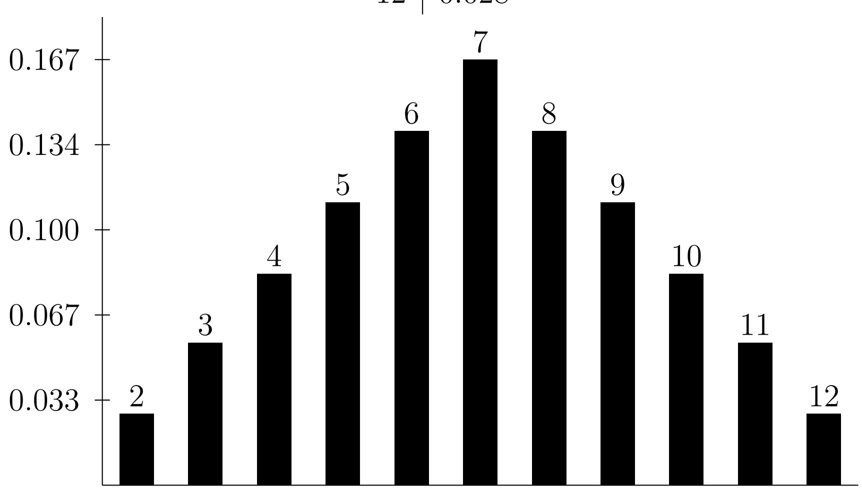


Example: Suppose that two dice are rolled and that x is the sum of the two dice. Then x is a (discrete) random variable. We can find the probability distribution of this random variable. First, we list all of the possible pairs of numbers (in order) that the experiment may result in:

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

We see that the possible values for x are the numbers 2,3,4,5,6,7,8,9,10,11, and 12. If we count the number of outcomes that gives us each sum and divide by 36, we get the following probability distribution:

x	$P(x)$
2	0.028
3	0.056
4	0.083
5	0.111
6	0.139
7	0.167
8	0.139
9	0.111
10	0.083
11	0.056
12	0.028



Example: Suppose that Bob buys a $100,000 life insurance policy for $250 for a year. Let x be the amount of money that Bob (or his family) "profits" from the policy. Suppose that the probability that Bob lives through the year is .9995. We can find the probability distribution for x. First, there are two outcomes: either Bob lives through the year (and loses his $250) or he dies (and "wins" $100,000). In the first case (which happens with probability .9995) his profit is $-\$250$. In the second case, his profit is $\$100,000 - \$250 = \$99750$ (which happens with probability 1-.9995=.0005). Thus, the probability distribution for x is

x	$P(x)$
$-\$250$	.9995
$\$99750$	.0005

Example: A certain lottery game costs \$1 to play. In the game, a player selects a sequence of three digits (0,1,2,3,4,5,6,7,8, or 9). If he selects the winning sequence, then the player wins \$5. Let x be the profit by someone playing this game. We can find the probability distribution for x. There are 1000 possible sequences and one winning sequence, so the probability of winning is $\frac{1}{1000}$ and the probability of losing is $\frac{999}{1000}$. If the player wins, his profit is $\$5 - \$1 = \$4$. If he loses, his profit is $-\$1$. The probability distribution of x is

x	$P(x)$
$\$4$	.001
$-\$1$	.999

Requirements of a Probability Distribution: All values of a probability distribution should be between 0 and 1 (inclusive) and should add to 1 (up to roundoff error).

SUMMARY STATISTICS

Summary Statistics for a Random Variable x:

- Mean or Expected Value: $\mu = \sum x \cdot P(x)$
- Variance: $\sigma^2 = \sum (x-\mu)^2 \cdot P(x)$
- Standard Deviation: $\sigma = \sqrt{\sum (x-\mu)^2 \cdot P(x)}$

Note: We may calculate the mean by hand for very small examples. We will calculate all other summary statistics with a calculator.

Interpretation of Mean: If an experiment is repeated many times and the values of x are averaged, then the average should be close to μ.

Example: The expected value for the two-flips example above is

$$\mu = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1.$$

If we flipped the two coins a thousand times and averaged the number of heads which appeared on each flip, the average should be close to 1.

Example: The expected value of Bob's life insurance policy above is

$$\mu = (-250) \cdot .9995 + (99750) \cdot .0005 = -200.$$

Bob's expected profit is $-\$200$. What does this mean? From the insurance company's point of view, they stand to make an average of \$200 per year on each customer similar to Bob.

Example: A calculator says that the mean sum which appears when two dice are rolled is $\mu = 7$ with a standard deviation of $\sigma = 2.42$. This means that if you rolled two dice 1000 times and averaged the sums of the dice, you should get an average near 7.

USUAL VALUES – RANGE RULE OF THUMB

We can use the summary statistics from a random variable to define the range of usual values following the Range Rule of Thumb.

Maximum and Minumum Usual Values: For a random variable x with mean μ and standard deviation σ,

- The minimum usual value for x is $\mu - 2\sigma$.
- The maximum usual value for x is $\mu + 2\sigma$.

Any value of x less than $\mu - 2\sigma$ is considered unusually low. Any value above $\mu + 2\sigma$ is unusually high.

Example: For the sum of two dice example,

- The minimum usual value is $\mu - 2\sigma = 7 - 2 \cdot 2.42 = 2.16$
- The maximym usual value is $\mu + 2\sigma = 7 + 2 \cdot 2.42 = 11.84$

This means the usual values for this random variable are 2,3,4,5,6,7,8,9,10, and 11. The unusual values are 2 and 12.

USUAL VALUES 5% RULE

Probabilities of Ranges: We can calculate probabilities such as $P(x \leq a)$ or $P(a \leq x \leq b)$ by adding up the probabilities of the values of x which satisfy the condition in parenthesis.

Example: Consider this probability distribution:

x	$P(x)$
0	0.01
1	0.02
2	0.03
3	0.04
4	0.16
5	0.28
6	0.27
7	0.08
8	0.08
9	0.02
10	0.01

To find $P(1.5 \leq x \leq 4)$, we identify which values of x satisfy $1.5 \leq x \leq 4$. These are 2, 3, 4, so

$$P(1.5 \leq x \leq 4) = P(2) + P(3) + P(4) = .03 + .04 + .16 = .33.$$

5% **Rule:** Suppose that N is a value that a random variable x can take.

- N is an unusually low value for x if $P(x \leq N) \leq 0.05$.
- N is an unusually high value for x if $P(x \geq N) \leq 0.05$.

Example: In the previous probability distribution, 0 and 1 are unusually low values of x while 9 and 10 are unusually high values of x.

RRT vs. 5% **Rule:** The Range Rule of Thumb and the 5% Rule often give different results. When identifying unusual values of a random variable, be sure to be aware of which rule you are using.