6 Probability Distributions

6.1 Summarizing Possible Outcomes and Their Probabilities

Random Variable

A **random variable** is a numerical measurement of the outcome of a random phenomenon. Often, the randomness results from the use of random sampling or a randomized experiment to gather the data.

Ex.1: Number of touchdowns for Red Raiders, Duration of a call

Probability Distribution

The **probability distribution** of a random variable specifies its possible values and their probabilities.

Ex.2: Probability distribution of number of home runs in a game for Boston Red Sox. The table lists the possible values for the number of home runs and the corresponding probabilities.

Number of Home Runs	**Probability**
0	0.23
1	0.38
2	0.22
3	0.13
4	0.03
5	0.01
6 or more	0.00

Probability distributions can be defined for both discrete random variables and continuous random variables. We will first look at discrete random variables.

Probability Distribution of a Discrete Random Variable

A **discrete** random variable X takes a set of seperate values (such as 0,1,2,....). Its **probability distribution** assigns a probability $P(x)$ to each possible value of x.

- For each x, the probability $P(x)$ falls between 0 and 1.
- The sum of the probabilities for all the possible x values equals 1.

Ex.3: Refer to **Ex.2**.

- for each x, the probability falls between 0 and 1. (all the values 0.23, 0.38, 0.22, 0.13, 0.03, and 0.01 are between 0 and 1)
- sum of the probabilities for all the possible x values equals 1. (0.23 + 0.38 + 0.22 + 0.13 + 0.03 + 0.01 = 1)

Therefore, it is a probability distribution.

Probability of at least 3 homeruns is
$P(X \geq 3) = P(3) + P(4) + P(5) + P(6) = 0.13 + 0.03 + 0.01 + 0.00 = 0.17$

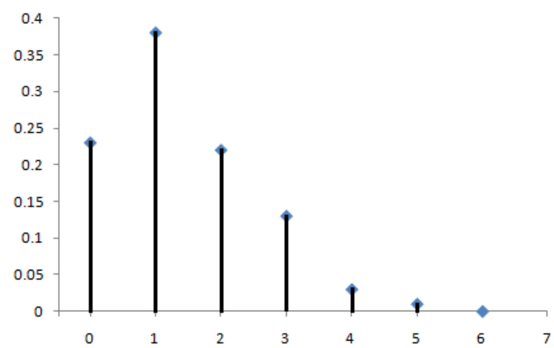
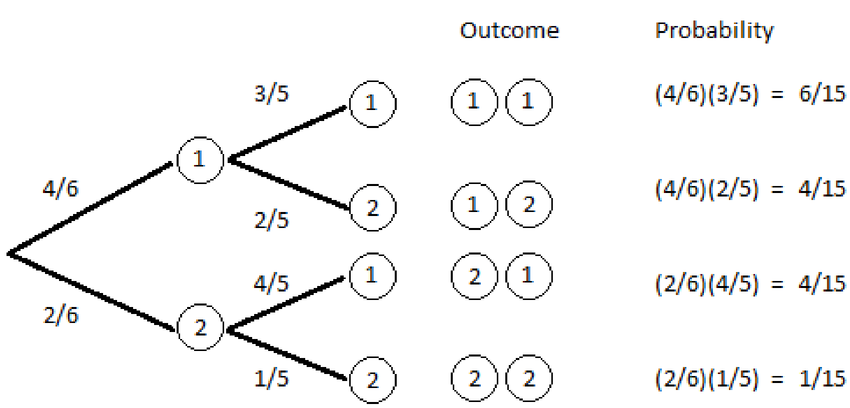


Figure 1: Probability Distribution for Ex.2

Ex.4: From six marbles numbered as 1,1,1,1,2,2, two marbles will be drawn at random without replacement. Let X denote the sume of the numbers on the selected marbles. List the possible values of X and determine the probability distribution.



X - sum of the two numbers

The probability distribution for this experiment is given in the following table.

X	**Probability**
2	$\frac{6}{15}$
3	$\frac{8}{15}$
4	$\frac{1}{15}$

The Mean of a Probability Distribution (μ)

The **mean of a probability distribution** for a discrete random variable is

$$\mu = \sum xP(x)$$

where the sum is taken over all possible values of x. This is also called the expected value for X.

Ex.5: The mean of the number of homeruns for in a game for Red Sox is,

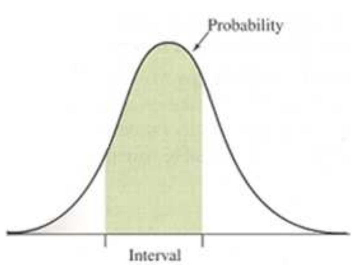
$\mu = 0(0.23) + 1(0.38) + 2(0.22) + 3(0.13) + 4(0.03) + 5(0.01) = 1.38$

This also means that you can expect the Red Sox to have an average of 1.38 homeruns per game for that season.

Probability Distribution of a Continuous Random Variable

A **continuous** random variable has possible values that form an interval. Its **probability distribution** is specified by a curve that determines the probability that the random variable falls in any particular interval of values.

- Each interval has probability between 0 and 1. This is the area under the curve, above that interval.
- The interval containing all possible values has probability equal to 1, so the total area under the curve equals 1.



Smooth Curve Approximation

6.2 Finding Probabilities of Bell-Shaped Distributions

Normal Distributions

The **normal distribution** is symmetric, bell-shaped, and characterized by its mean μ and standard deviation σ. The probability within any particular number of standard deviations of μ is the same for all normal distributions. This probability equals 0.68 within 1 standarad deviation, 0.95 within 2 standard deviations, and 0.997 within 3 standard deviations.

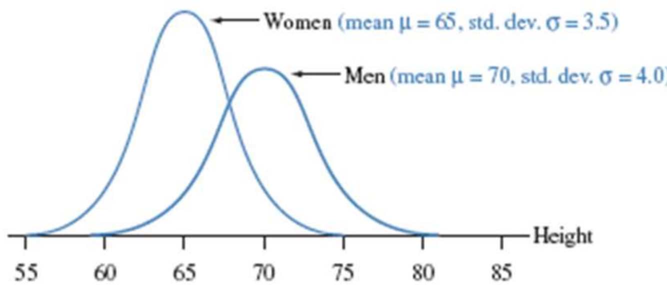


Figure 2: Normal Distributions for Women's Height and Men's Height

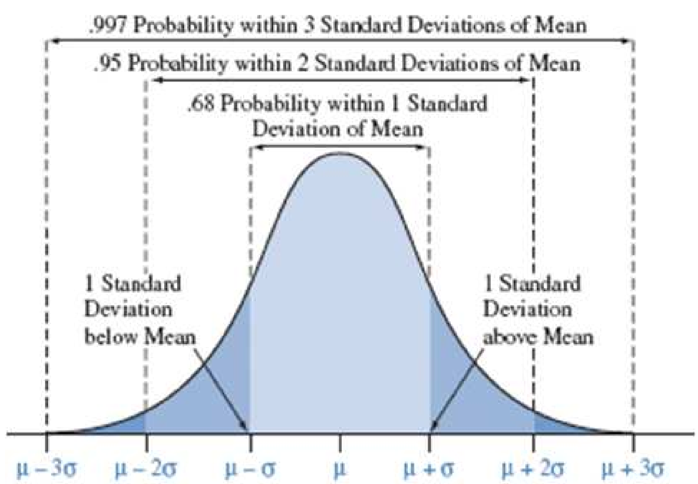


Figure 3: Empirical Rule for Normal Distribution

Z-Scores and the Standard Normal Distribution

- The z-score for a value x of a random variable is the number of standard deviations that x falls from the mean. It is calculated by

$$z = \frac{x - \mu}{\sigma}$$

- A negative (positive) z-score indicates that the value is below (above) the mean.

- z-scores can be used to calculate the probabilities of a normal random variable using the normal tables.

Using Z-Scores to Find Normal Probabilities or Random Variable x Values

- Given the x value, finding a probability
  - Convert the x value to the corresponding z-score using $z = \frac{x-\mu}{\sigma}$
  - Use the normal table to find the corresponding **cumulative** probability
  - Convert the cumulative probability to the probability of interest
- Given the probability, finding the value of x
  - Convert the given probability to the corresponding **cumulative** probability
  - Use the normal table to find the corresponding z-score
  - Convert the z-score to the corresponding x value using $z = \frac{x-\mu}{\sigma}$

Reading the Standard Normal Cumulative Probability Table

The table of standard normal cumulative probabilities tabulates the normal cumulative probability falling **below** a z-score. For a given z-score, the cumulative probability can be read using information.

- Round the z-score to two decimal places. Then the first column gives the first two digits (to the first decimal place) of the z value. The first row gives the third digit (the second decimal place) of the z value.
- The corresponding probability found in the body of the table gives the probability of falling **below** the z-score
- To find the probability of falling above that z-score, do (1 - cumulative probability).

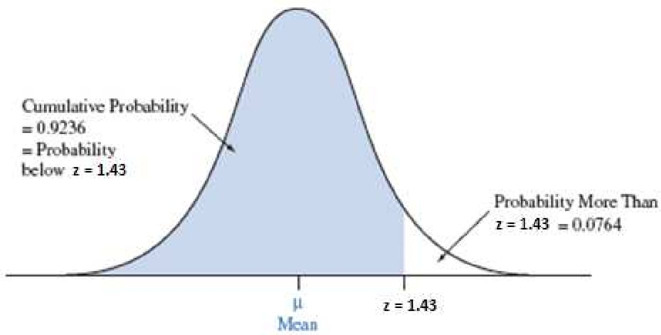
Ex.6: Find the probability that a normal random variable takes a value less than what corresponds to z-score 1.43. i.e. $P(z < 1.43)$.

Look up the cumulative probability that corresponds to $z = 1.43$. This probability is 0.9236. Therefore the probability that we wanted to find = 0.9236.

To find the probability of the normal random variable taking a value larger than what corresponds to z-score 1.43, i.e. $P(z > 1.43)$, do $1 - 0.9236$, which is equal to 0.0764.

In the case that the z-score is not given, but the x value and the mean and the standard deviation is given, the x value has to first be converted to the corresponding

Second Decimal Place of z										
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
...										
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9139	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441



z-score using $z = \frac{x-\mu}{\sigma}$. Then the same procedure explained above is used.

Ex.7: One year, the scores on the exam have approximately a normal distribution with mean 83 and standard deviation 5. About what proportion of the students earned a 'A'?

Knowing that a 'A' is given for scores above 90, the problem is to find the probability of a student scoring above 90. i.e. $P(X > 90)$.

The z-score that corresponds to $x = 90$ is calculated as $z = \dfrac{90-83}{5} = 1.40$. Now the problem is equivalent to finding the probability of z-score being larger than 1.40. By looking up the table for $z = 1.40$, we get the cumulative probability (the probability below $z = 1.40$) to be 0.9192. Therefore, the probability of z-score being larger than 1.40 is $1 - 0.9192 = 0.0808$. Therefore there is 8.08% chance of someone earning a 'A'.

If the problem is to find the probability of the random variable falling within a given interval, the problem can be solved using the difference of two cumulative probabilities. This is explained in the following example.

Ex.8: Refer to Ex.7. About what proportion of the students earned a 'B'?

Knowing that a 'B' is given for scores between 80-90, the problem is to find the probability of a student scoring between 80 and 90. As can be seen in the figure below, this probability of interest can be viewed as a difference of two cumulative probabilities.

$P(80 < X < 90) = P(X < 90) - P(X < 80)$

Lets find $P(X < 90)$ by first converting it to the corresponding z-score and then looking up the table.

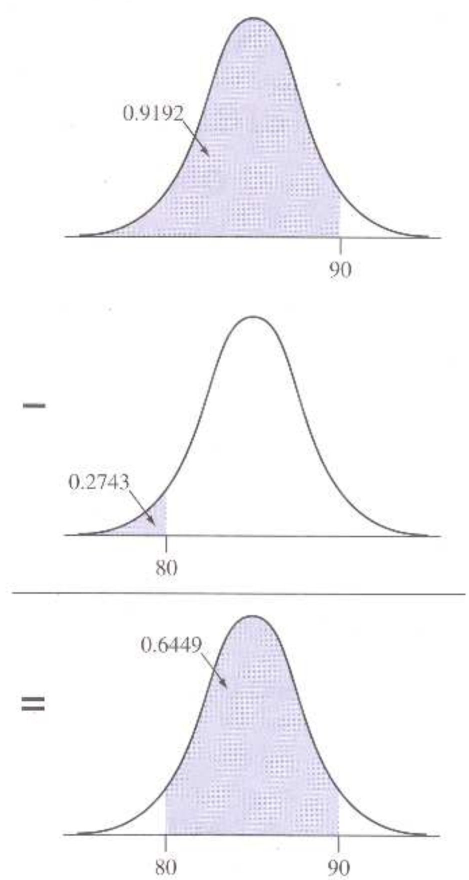
$$z_1 = \frac{90 - 83}{5} = 1.40$$
$$P(X < 90) = P(z < 1.40) = 0.9192$$

We can find $P(X < 80)$ in the same manner.

$$z_2 = \frac{80 - 83}{5} = -0.60$$
$$P(X < 80) = P(z < -0.60) = 0.2743$$

Now the probability of interest can be found by

$$P(80 < X < 90) = P(X < 90) - P(X < 80) = 0.9192 - 0.2743 = 0.6449$$



Finding the Z-Score for a Given Probability

To find the z-score for a given probability, first, this given probability has to be converted to a cumulative probability. Then the closest value to this probability should be found from the normal table. (Search for it in the middle part of the table). Then read the corresponding z-score.

Ex.9: Find the z-score having area 0.9 to its right under the standard normal curve.

First we have to convert this area to cumulative area/probability by finding the area that is to the left of the z-score of interest. This can be done by $1 - 0.9 = 0.1$. Then by looking at the table, we can see that the closest value we can find to 0.1 from the table is 0.1003 which corresponds to z-score -1.28. (Remember that you are searching for the value 0.1 in the middle of the table. i.e. in probabilities. Not in the first column and the first row.) Therefore the answer is -1.28.

Z-Scores and the Standard Normal Distribution

- A **standard normal distribution** is a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.
- When a random variable has a normal distribution and its values are converted to z-scores by subtracting the mean and dividing by the standard deviation, the z-scores have the standard normal distribution.

6.3 Finding Probabilities When Each Observation Has Two Possible Outcomes

There are many scenarios that have only two possible outcomes.

Ex.10:

- Passing or Failing an exam
- Accept, or decline an offer from a bank for a credit card
- Have, or do not have, health insurance
- Win or loose a gamble

In general, even if the experiment has more than two outcomes, **if we are interested in only one event**, we can reduce the problem into having only two outcomes. Simply 'getting the outcome we are interested in' and 'not getting the outcome we are interested in'.

Ex.11: Consider the case of rolling a die. Assume we are interested in getting '5'. Although the number of outcomes for this scenario is 6, we can reduce the problem into having only two outcomes. i.e. 'getting 5' and 'not getting 5'.

Under certain conditions, a random variable X that counts the number of observations of a particular type has a probability distribution called the binomial distribution. The following gives the conditions under which we can model something as a binomial distribution.

Conditions for the Binomial Distribution

- Each of n trials has two possible outcomes: The outcome of interest is call 'success' and the other outcome is called 'failure'
- Each trial has the same probability of success, denoted by p
- The n trials are independent
- The binomial random variable X is the number of successes in the n trials

Probabilities for a Binomial Distribution

Denote the probability of success on a trial by p. For n independent trials, the probability of x successes equals,

$$P(x) = \frac{n!}{x!(n-x)!}\, p^x \, (1-p)^{n-x}, x = 0, 1, 2, \dots, n$$

Note: $n!$, or the **factorial of n** means the multiplication of numbers 1 through n. $n! = n \times (n-1) \times \ldots 3 \times 2 \times 1$

Ex.12: An archer is able to hid the bull's-eye 55% of the time. If she shoots 8 arrows, what is the probability that she gets exactly 5 bull's-eyes? Assume each shot is independent of the others.

This scenario has only two outcomes. Hitting the bull's-eye or not hitting the bull's-eye. Each shot is independent of each other. Therefore, we can model this using binomial distribution. In the example, we are interested in counting successful hits. Therefore let's define 'success' as 'hitting the bull's-eye'. Then $p = 0.55$. n the number of trials is equal to 8. Therefore the probability of hitting exactly 4 bull's-eyes is given by

$$\begin{aligned} P(X=4) &= \frac{8!}{5!(8-5)!}\, 0.55^5\, (1-0.55)^{8-5} \\ &= 56\,(0.55)^5\,(0.45)^3 = 0.2568 \end{aligned}$$

Binomial Mean and Standard Deviation

The binomial probability distribution for n trial with probability p of success on each trial has mean μ (this is also the expected number of successes in n trials) and standard deviation σ given by

$$\mu = np,\ \ \sigma = \sqrt{np(1-p)}$$

Ex.13: Refer to Ex.12. Find the expected number of successful hits to the bull's-eye out of the 8 hits (find the mean) and the standard deviation.

Mean $\mu = 8(0.55) = 4.4$
Standard deviation $\sigma = \sqrt{8(0.55)(1-0.55)} = 1.407$

Approximation of Binomial Distribution by Normal Distribution

Note: A binomial distribution can be well approximated by a normal distribution when n is large enough such that np and $np(1-p)$ are both at least 15.