**4COSCOO2W MATHEMATICS FOR COMPUTING**

**Week 11 Seminar Tasks**

**Probability Theory & Statistics**

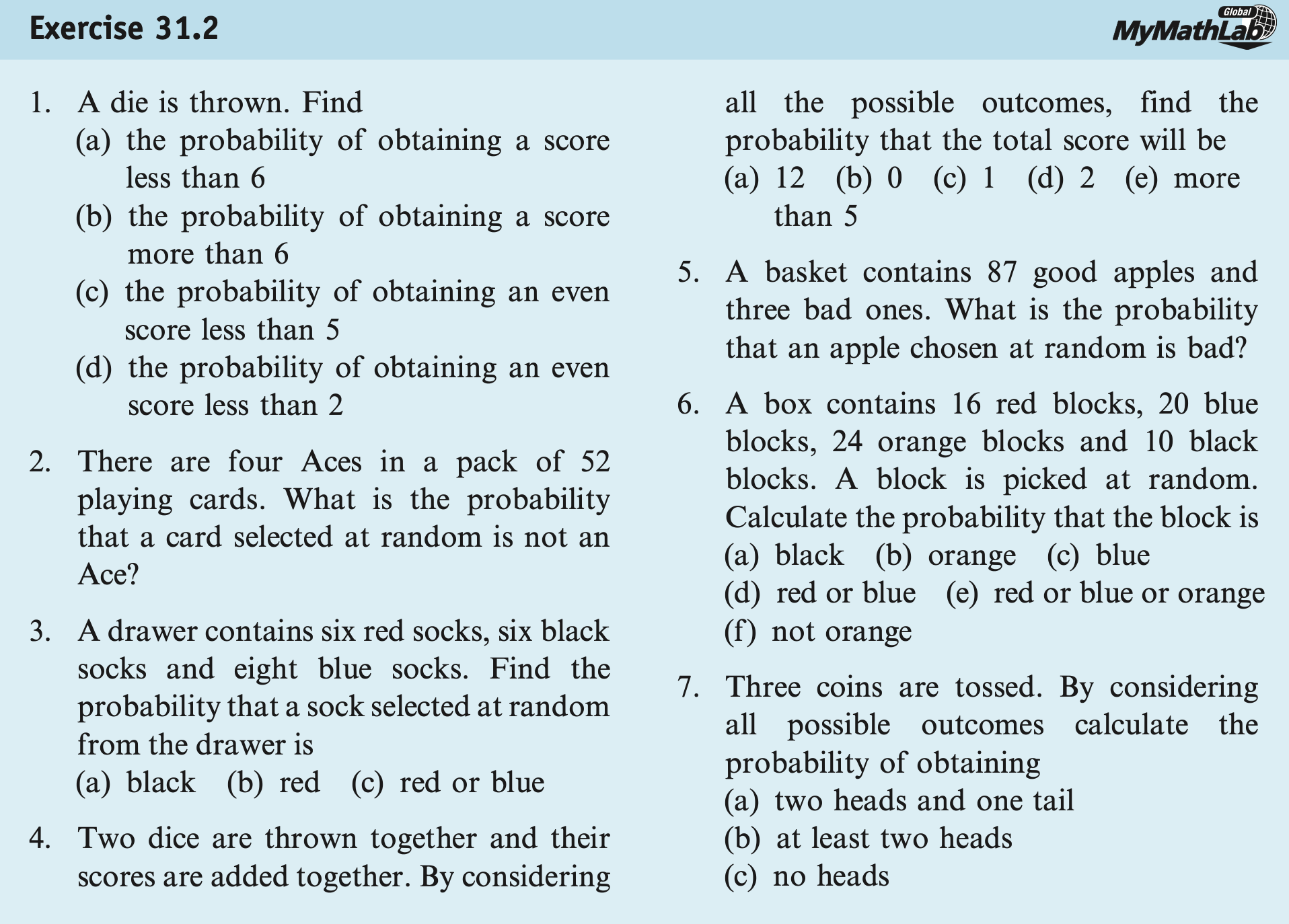
**READING**

Lecture 10-11 Notes (available on Blackboard)

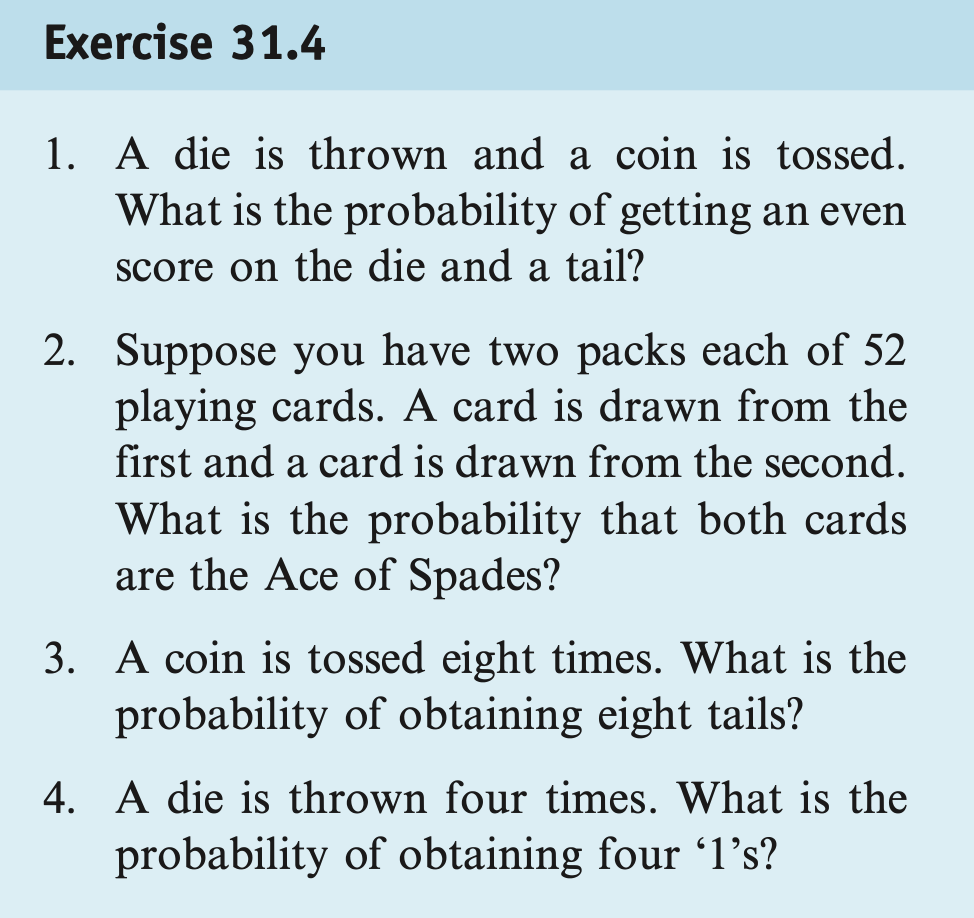
Chapters 30, 31. *Croft, T and Davison R (2016) Foundation maths, 6th ed. Harlow: Pearson.*

**TASK 1.**

*See Lecture 10 Notes*

**Task 1.1.** Exercises 31.2 (1-7) from Chapter 31. Croft, T and Davison R (2016) Foundation maths, 6th ed. Harlow: Pearson.

**Task 1.2.** Exercises 31.4 (1-4) from Chapter 31. Croft, T and Davison R (2016) Foundation maths, 6th ed. Harlow: Pearson.



*NOTE:* You may find the solutions for the above exercises in the textbook.

**TASK 2.**

*See Lecture 10 Notes*

**Task 2.1.** A fair die is rolled twice, and we obtain two numbers: X1= “result of the first roll”, and X2= “result of the second roll”. Let A and B be the events defined as follows: A is "X1<X2"; B is "You observe a 6 at least once".

1. Define the probability space, D, in set builder notation and its cardinality.
2. Find the cardinality of event A represented as a subset of the probability space D. Find the probability of event A
3. Find the cardinality of event B represented as a subset of the probability space D. Find the probability of event B.

**Solution:**

1. Define the probability space, D, in set builder notation and its cardinality.

The probability space D has 36 options – it will combine pairs (X1, X2) where both X1 and X2 are one of 6 possible outputs of each roll. So in set builder notation, the probability space

D = {(X1, X2): X1 ∈{1,2,3,4,5,6}, X2 ∈{1,2,3,4,5,6}}

The cardinality of D, |D| = 36.

1. Find the cardinality of event A represented as a subset of the probability space D. Find the probability of event A

For event A:

A={(1,2),(1,3),(1,4),(1,5),(1,6),(2,3),(2,4),(2,5), (2,6),(3,4),(3,5),(3,6),(4,5),(4,6),(5,6)}.

So the cardinality of A: |A| = 15

1. Find the cardinality of event B represented as a subset of the probability space D. Find the probability of event B.

For event B

B ={(6,1),(6,2),(6,3),(6,4),(6,5),(6,6),(1,6),(2,6),(3,6),(4,6),(5,6)}.

So the cardinality of B: |B| = 11

**Task 2.2.** A fair die is rolled twice, and we get two numbers: X = “result of the first roll” and Y = “result of the second roll”.

1. What is the probability that X = 4

**Answer:** This is 1 out of 6

1. What is the probability that Y = 4

**Answer:** This is 1 out of 6

1. What is the probability that both X= 4 and Y = 4

**Answer:**

**Task 2.3.** A fair die is rolled twice, and we obtain two numbers: X1= “result of the first roll”, and X2= “result of the second roll”.Let A be the event that “X =4 or Y =4” and B be the event that “X + Y=7”.

1. What is P(B)?
2. What is P(A ∩ B)?



1. Given that the sum of the rolls of two dice (X and Y) is 7 (this is event B), what is the probability that either X or Y (or both) is 4 (this is event A)? In other words, if we already know that the sum of the two dice is 7, how likely is it that one of the dice shows a 4?

Find the Conditional Probability P(A|B).

**Solution**

Let’s form sets describing events A and B:

A={(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),(1,4),(2,4),(3,4),(4,4),(5,4),(6,4)},

B={(6,1),(5,2),(4,3),(3,4),(2,5),(1,6)},

A ∩ B = {(3,4), (4,3)}, so



a)

b)

c)

P(A|B) = = = =

**TASK 3. STATS BASICS**

*See Lecture 11 Notes*

As part of a performance evaluation, a series of stress tests were conducted on a new software application. The dataset below records the time (in milliseconds) it took for the application to complete different tasks during these tests.

Dataset: Task Completion Times (in milliseconds)

|  |  |
| --- | --- |
| Task | Time (ms) |
| T1 | 120 |
| T2 | 115 |
| T3 | 130 |
| T4 | 110 |
| T5 | 125 |
| T6 | 105 |
| T7 | 135 |
| T8 | 95 |

You are required to manually calculate the following statistical measures to analyse the software's performance:

* **Mean:** Calculate the average task completion time.
* **Range:** Determine the range of task completion times.
* **Median:** Find the median task completion time, which represents the central tendency of the dataset.
* **Mode:** Identify the most frequent task completion time. If all times are unique, note that the dataset has no mode.
* **Standard Deviation:** Compute the standard deviation to gauge the consistency of the task completion times.
* **Variance:** Calculate the variance to measure the degree of spread in the task completion times.
* **Interquartile Range (IQR):** Determine the IQR to evaluate the spread of the middle 50% of the task completion times.

**Solution**

Process for Manual Calculation.

* Sort the task completion times in ascending order.
* Apply the statistical formulas as discussed in class to compute each measure.
* For calculating the IQR, precisely identify the 25th and 75th percentiles and compute the

Present a detailed computation for each statistical measure and provide an interpretation of what these measures imply about the software's performance.

**Mean (Average Time)**

Result: Mean time = 116.875 ms

**Range**

Calculation:

Result: Range = 40 ms

**Median**

Sorted Data: [95, 105, 110, 115, 120, 125, 130, 135]

Calculation: The median is the average of the 4th and 5th terms,

Result: Median time = 117.5 ms

**Mode**

Result: Since all times are unique, there is no mode.

**Standard Deviation**

To calculate the standard deviation in a detailed manner, we follow these steps:

1. Find the Mean: Calculate the average (mean) of the data points.
2. Calculate the Differences: Subtract the mean from each data point to find the differences.
3. Square the Differences: Square each of the differences.
4. Sum the Squares: Add up all the squared differences.
5. Divide by N-1: Divide this sum by the number of data points minus one (this is called the "sample standard deviation" formula).
6. Square Root: Take the square root of the result from step 5 to get the standard deviation.

Squared Differences from the Mean:

The Sum of Squares = =

Variance (using sample variance formula):

Variance = =

Standard Deviation:

Standard Deviation ( ≈ ms

The standard deviation is approximately 13.346 ms, which measures the amount of variation or dispersion of the task completion times from the mean (average time).

**Variance**

Calculation: Square of the standard deviation.

Result: Variance ≈ 178.125 (ms)^2

**Interquartile Range (IQR)**

The position of a given quartile in a sorted dataset can be determined by the following formula:

Where:

is the quartile percentage (25 for Q1, 50 for the median, 75 for Q3).

is the number of observations in the dataset.

If the quartile position calculated is not a whole number, we use interpolation between the two closest data points. The formula for interpolation is:

Where:

is the value at the floor of the position (i.e., the value at the lower position).

is the calculated position (which might be a fractional number).

is the value at the next higher position in the dataset.

First Quartile (Q1) Calculation.

Position of Q1 (25th percentile): .

Interpolate between the 2nd and 3rd values in the sorted data (105 and 110):

Exact Q1: ms.

Third Quartile (Q3) Calculation.

Position of Q3 (75th percentile): .

Interpolate between the 6th and 7th values in the sorted data (125 and 130):

Exact Q3: ms.

Interquartile Range (IQR):

= 22.5 ms.

The Interquartile Range (IQR) for the dataset is 22.5 ms, which represents the range between the first and third quartiles and provides insight into the variability of the middle 50% of the data, excluding outliers.