

University of Kent

ASTRONOMY, SPACE SCIENCE AND ASTROPHYSICS

Probability Distributions

STAGE 1 - PH370 PHYSICS LABS

Monday 18th/26th March 2018

Report Author: Lukasz R Tomaszewski

Lab Partner: Teodora Romanova

Contents

1	Abstract	1
2	Introduction	1
3	Aims & Equipment	1
3.1	Apparatus	1
3.2	Data Collected	2
3.3	Risk Assessment	2
4	Experimental Procedure	2
4.1	Experimental Theory	2
4.2	Experimental Method	3
5	Results & Discussion	4
5.1	Task 1 - Data Taking	4
5.2	Task 2 - Computer Simulation	5
5.3	Task 3 - Rolling Dice	6
6	Analysis	7
7	Conclusion	7

1 Abstract

The main aim of this experiment is to evaluate and prove the relation of statistical data against physically collected data in regards to three types of probability distributions: the Binomial, the Normal (Gaussian) and the Poisson distributions. Within this experiment, three main short physical experiments will take place to prove the hypothesis, and show connections between each type of distribution. The overall analysis of the data collected in this report shows a connection between the Binomial and Normal (Gaussian) distributions as well as a strong correlation between the statistical and physical data, this report will also outline the only error involved was human error as this experiment will mainly be done using computer simulations and programming software.

2 Introduction

This report discusses the experiment regarding probability distributions, specifically three types; the Binomial, the Normal (Gaussian) and the Poisson distributions. The main objective is to test physical data against the statistical data via each distribution. As each of three probability distributions are used in certain circumstances (which are found in section 4.1), then three different experiments will take place using ball bearings, computer simulations and dice. Therefore its imperative to analyze each distribution separately.

The Binomial distribution will be used to calculate the probability of 1 of 500 balls landing in a specific bin (1 bin out of 17 bins), this experiment will be calculated mathematically first then test physically and compared to the statistical data. An analysis of if the statically can be proved physically will be derived. After which the results of the physical experiment will be compiled into a table 1 and plotting via a programming software called Python 3.6, using this program it will plot a histogram from the physical data 1 to calculate the standard error using [3] as reference. This experiment in found under section 5.1 of this report.

To compare Normal or otherwise known as Gaussian distribution statistical data with physical data, the experiment will consist of a IDE (Integrated Development Environment) using the 'Spyder' Python 3.6 software. Modifying the script changing the total number of ball bearings used and the histogram settings will allow multiple graphs to be compiled and compared, through physical differences but most importantly compare how Binomial and the Normal (Gaussian) distributions relate to each other within this experiment. This experiment in found under section 5.2 of this report.

In the third and final experiment is to prove the Poisson and Normal (Gaussian) distributions, this will take the form of numerous dice rolls using a online simulation [1]. Using the Normal (Gaussian) formula which is found is section 4.1, and deriving a formula to describe the probability that a single di will land on one precise face. After which the online simulation will be used to gather a variety of random results of faces shown which will act as the physical data when compared to the statistical formula derived earlier. A final analysis of all three probability distributions will be formed in section 6 in which the experiment will end. This overall experiment is dictated by the lab brief [2].

3 Aims & Equipment

3.1 Apparatus

- x500 Ball bearings
- Python software
- Dice computer simulation

3.2 Data Collected

- x10 Trials of 500 of ball bearings in 17 bins in 1 bin
- Bi-Nominal distribution of 1 ball bearing
- Averages of random rolls of dice

3.3 Risk Assessment

This experiment contains a few hazards that need to be noticed and avoided such as equipment falling off the desk, loose ball bearings and using a the PC for extended periods of time. Such risks include bruising/ injury to lower body if equipment falls of the desk, tripping/ slipping on the ball bearings if they get loose and fall onto the floor, and electrocution and visual stress on the eyes for extended use of the PC. To control these hazards, we will be supervised constantly throughout this lab session, the equipment will be kept away from the edges of the tables at all times, random visual inspections of the board holding the ball bearings to ensure no ball bearings escape and all computers are electrically tested, the two lab partners will swap periodically on using the PC.

4 Experimental Procedure

4.1 Experimental Theory

The Binomial Distribution:

$$B(N, r) = \left(\frac{N!}{r!(N-r)!} \right) \times x^r (1-x)^{N-r} \quad (1)$$

Where;

$N \Rightarrow$ Number of trials

$r \Rightarrow$ Number of successes

$x \Rightarrow$ Success in one trial

The Binomial distribution is used specifically where the outcome is uncertain but only two outcomes are possible. For example tossing a coin, the only results can be heads or tails.

The Normal (Gaussian) Distribution:

$$G(x) = \frac{1}{\sigma \times \sqrt{2\pi}} \times \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (2)$$

Standard Error:

$$\mu = Nx \quad (3)$$

Mean/ Average:

$$\sigma^2 = Nx(1-x) \quad (4)$$

Where;

$N \Rightarrow$ Number of trials

$\sigma \Rightarrow$ Standard Deviation/ Error

$\mu \Rightarrow$ Mean/ Average

$x \Rightarrow$ Success in one trial

This distribution is the most important of its kind as it's continuous, the variable can assume any listed value, it relies on two parameters; The mean (Eq.4) and the standard deviation (Eq.3).

The Poisson Distribution:

$$p(N) = \frac{e^{-\bar{N}} \bar{N}^N}{N!} \quad (5)$$

Where;

$N \Rightarrow$ Number of successful outcomes

$\bar{N} \Rightarrow$ Mean number of outcomes

The Poisson distribution is much like the two previous distributions but this distribution has a specific purpose in where it expresses the probability of successful outcomes. By a set number of outcomes, with the mean number of outcomes.

4.2 Experimental Method

In the first task of this experiment, is to use a symmetrical board filled with a number of round pins, and also filled with 500 ball bearings and 17 bins on each side of the board for the ball bearings to be collected in to. With all the ball bearings on one side of the board, the board is turned vertically so the ball bearings fall down into a small gap which separates the two sides of the board, then flow through the gap hitting the pins on their way down. When all 500 ball bearings have filled the bins then the number of ball bearings in each individual bin must be counted. After this the process is repeated by rotating the board, this was repeated 10 times. Using "Spyder" Python 3.6 the values of this experiment were inputted into a histogram in which the standard deviation (standard error) was retrieved and calculated.

Within the second task using "Spyder" Python 3.6, the pre-written script displayed a scenario similar to task 1, where the a histogram is plotted and the binomial distribution was calculated within the script. By editing a few lines in the script; the number of balls used, the number of bins and pins within the board, once changed the script outputted a identical graph similar to the pre-written script. Using the data and graphs, compare and "show qualitatively that the binomial distribution may be increasingly well approximated by the normal distribution" [2].

In the third and final part of this experiment, utilizing an online simulation for rolling dice [1]. The first of many questions will be answered by predicting the calculating the peak of the Normal (Gaussian) distribution for a single di rolled over and over, this data assists in why the calculated value is unhelpful in predicting a dice roll. Plotting a chart in the form of a table (Table 2) with all the possible combinations two dice will shown if rolled, from this the distribution can be deduced and the common value can help predict future rolls. The second part of this experiment consisted of a having two distinguishable di, much like before, plotting a chart in the form of a table (Table 3) where the plot of the distribution will be compared to the first part of this experiment. The final part of this experiment is that of data collection where a random dice rolling [1] will be used to collect multiple values under

different circumstances; number of faces on the di and total number of di where a predication algorithm can be formed in the shape of a distribution.

5 Results & Discussion

5.1 Task 1 - Data Taking

Probability of an individual ball falling into bin No.15;

$$p(x) = {}^{16}c_2 \times 0.5^2 \times 0.5^{14} = 1.83 \times 10^{-3} \quad (6)$$

Probability of an individual ball falling into bin No.12;

$$p(x) = {}^{16}c_5 \times 0.5^5 \times 0.5^{11} = 0.0667 \quad (7)$$

Probability of an individual ball falling into bin No.9;

$$p(x) = {}^{16}c_8 \times 0.5^8 \times 0.5^8 = 0.196 \quad (8)$$

Table 1: Results of Task 1 - Data Taking

Trials Vs Bins	1	2	3	4	5	6	7	8	9	10	Average
1	0	2	1	0	0	0	0	0	0	0	0.3
2	2	1	0	1	1	1	0	0	2	1	0.9
3	1	6	0	6	4	3	1	3	5	3	3.2
4	18	18	9	19	20	9	2	17	6	8	12.6
5	11	27	13	32	25	28	8	27	26	23	22.0
6	23	41	25	53	45	33	19	31	46	37	35.3
7	56	67	76	76	67	59	60	72	74	47	65.4
8	76	79	69	61	80	78	60	66	69	81	71.9
9	93	69	82	63	79	77	86	83	80	72	78.4
10	75	72	75	77	81	91	97	93	74	92	82.7
11	66	55	77	57	51	60	89	53	55	69	63.2
12	37	30	36	34	20	40	41	34	33	38	34.2
13	29	19	25	15	21	14	25	15	22	20	20.5
14	12	7	12	6	5	6	5	5	2	10	7.0
15	4	6	8	4	3	0	6	2	3	0	3.6
16	1	1	1	0	3	0	1	0	2	0	0.9
17	0	0	1	0	0	0	0	0	0	0	0.1
Total	504	500	510	504	505	499	500	501	499	501	502.3

The output values above are expected as the environment in which the balls have been placed into, the balls fall onto a pin within the board which allows the individual ball to fall either right or left, in terms of values 0.5 either way. The only interaction that could alter this is for human interference of defects in the pins or board, thus causing an unfair experiment.

5.2 Task 2 - Computer Simulation

Comparing Table 1 and Figure 1, it shows a strong correlation between the theoretical data in Figure 1 which is formed via mathematical methods of the Normal (Gaussian) distribution and the Table 1 of physically collected data. Thus implying that the binomial distribution is connected with the Normal (Gaussian) distribution.

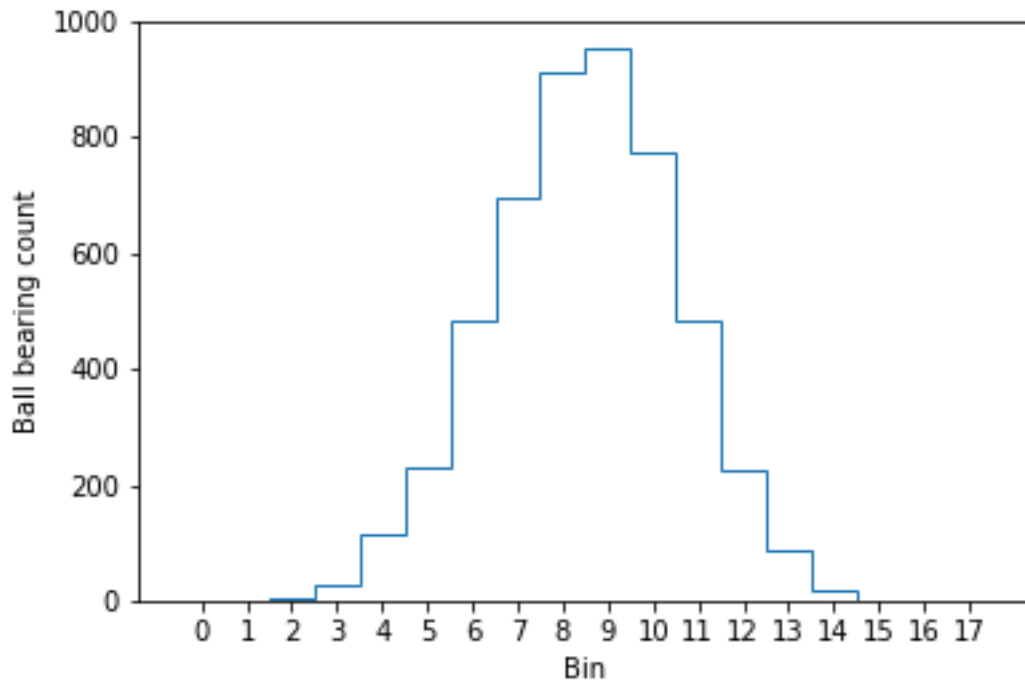


Figure 1: Simulation of Task 1 - Data Taking

Python 3.6 Script

```
# -*- coding: utf-8 -*-
"""
Created on Wed Apr 4 16:19:36 2018

@author: lrgtomaszewski
"""

#Binomial experiment script

import numpy as np
from matplotlib import pyplot as plt
import scipy.special

num_ball_bearings = 5000

num_levels = 17
```

```

locations_5000 = np.zeros(num_ball_bearings)

for level in range(num_levels):
    shifts = np.random.randint(0,2,size=num_ball_bearings)
    locations_5000 += shifts

plt.hist(locations_5000,bins=np.arange(-0.5,1+num_levels,1),histtype='step',label='5000')

plt.xticks(np.arange(0,1+num_levels,1,dtypes=int))

theory_values = np.zeros(num_levels+1)

n = num_levels

for k in range(0,n+1):
    theory_values[k] = num_ball_bearings * scipy.special.binom(n, k) * 0.5**k *
    ↪ 0.5**(n - k) #the binomial distribution formula, with p set to 0.5 (equal
    ↪ probability of falling left or right)

plt.bar(range(0,num_levels+1),theory_values,1,color='',edgecolor='red',label='Theory')

plt.xlabel('Bin')
plt.ylabel('Ball bearing count')
plt.savefig('T2 Graph')
plt.legend()
plt.show()

```

5.3 Task 3 - Rolling Dice

Probability of a di landing on a face;

$$p() = \frac{1}{N} \rightarrow \frac{1}{6} \quad (9)$$

Average value of roll for a six sided di;

$$p() = \frac{(1 + 2 + 3 + 4 + 5 + 6)}{6} \rightarrow \frac{7}{2} = 3.5 \quad (10)$$

If a di is rolled N times, is the data suitable for prediction?;

- 1) The value is a half, it's impossible to roll a half.
- 2) Maths does not take into account any effect of nature.

Table 2: Results for Task 3 - Rolling Dice for two identical di

x	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$$Average = \frac{252}{36} = 7(For2dice) \quad (11)$$

The most common value is 7 and the total number of combinations it has is 6. Thus using 2 dice allows for double the value of a single di (6 cahnges to 12) thus doubling the mean.

Table 3: Results for Task 3 - Rolling Dice for two distinguishable di

x	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

6 Analysis

Analyzing the results which were concluded in Section 5 proves a strong correlation between statistical data and physical data, in each three experiments, the statistical data matched the physically collected data. Even though this experiment was conducted fairly and mainly using computer simulations and the programming software "Spyder" Python 3.6 so there was no error to be calculated for such operations. The only error found and derived was from human error where the users had an impact on the ball bearings falling down the board in Task 1 - Data Taking, this was to be expected as this short experiment relies completely on human interference but it is unclear and cannot be calculated the damage it has done in terms of the fairness of this experiment overall.

7 Conclusion

In conclusion, with respect to the purpose of this report outlined in Section 2, the experiment proved that there is a strong connection between each of the 3 probability distributions statistical data and physical data. In each short experiment, it was proved that the statistics match the physical data collected from the experiment, this is highlighted throughout Section 5. The experiment was kept as fair and unbiased to promote true variation in results to fully verify the implied hypothesis of this experiment, the only error obtained within this experiment is truly human error which cannot be calculated but can be taken into account.

References

- [1] Wartoft M. <http://www.roll-dice-online.com/>.
- [2] G. Roch. Llr.4 probability distributions. University of Kent Moodle 2017, June 2017. LLR.4 Lab script.
- [3] J. R. Taylor. Introduction to error analysis. University Science Books, 1997. ISBN - 978-0-935702-75-0.