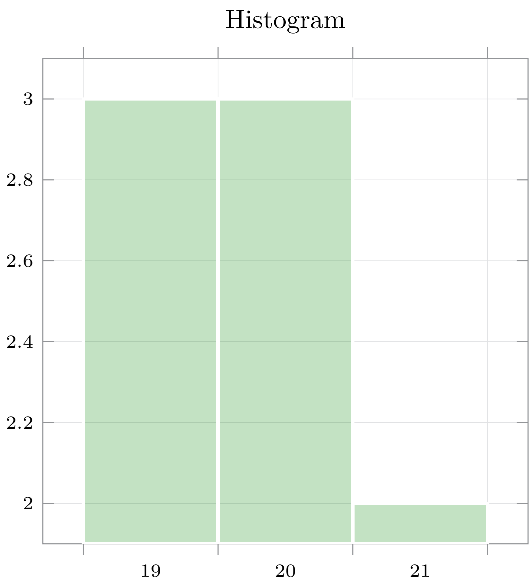
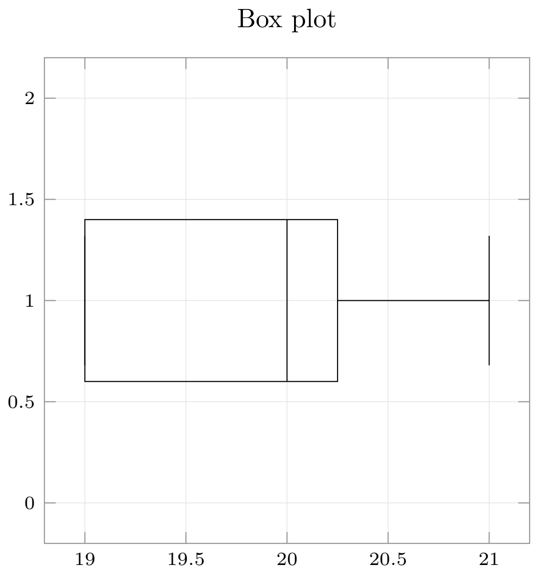
Chapter 24

Data Analysis. Probability Theory

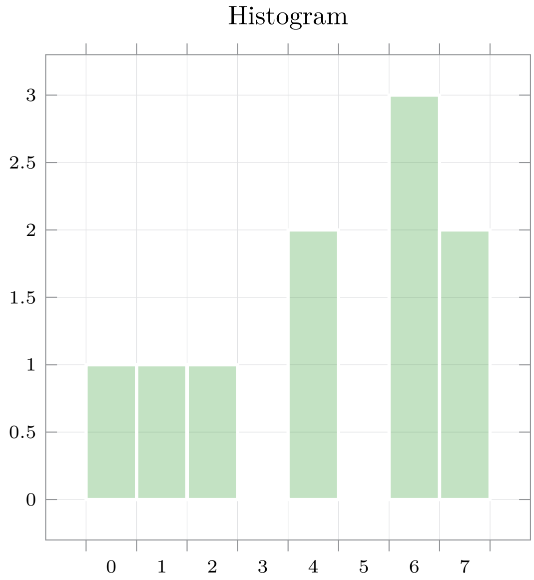
24.1 Data Representation, Average, Spread

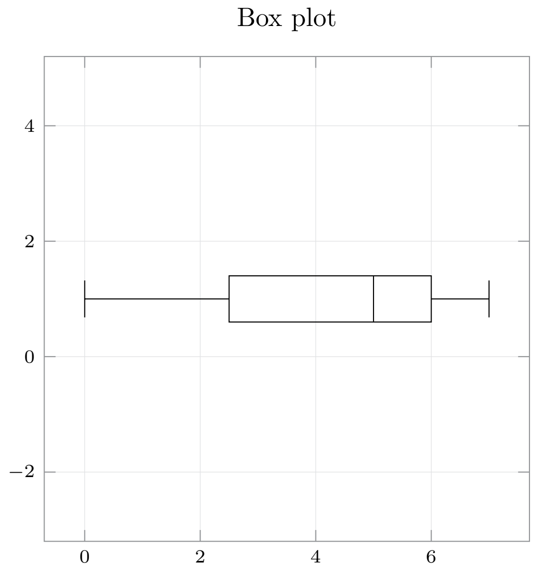
1. Representing the data,



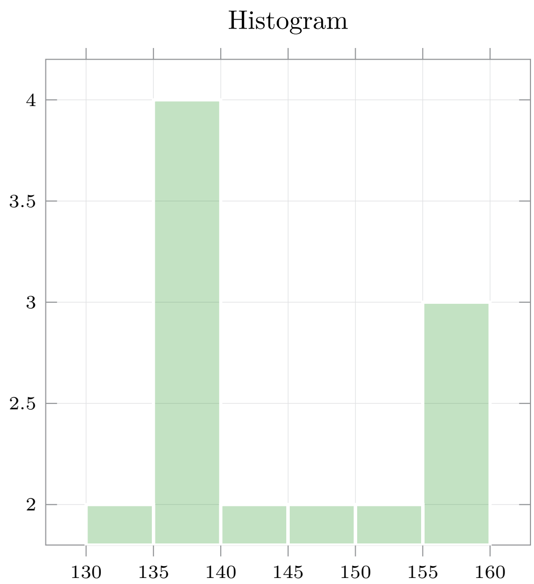


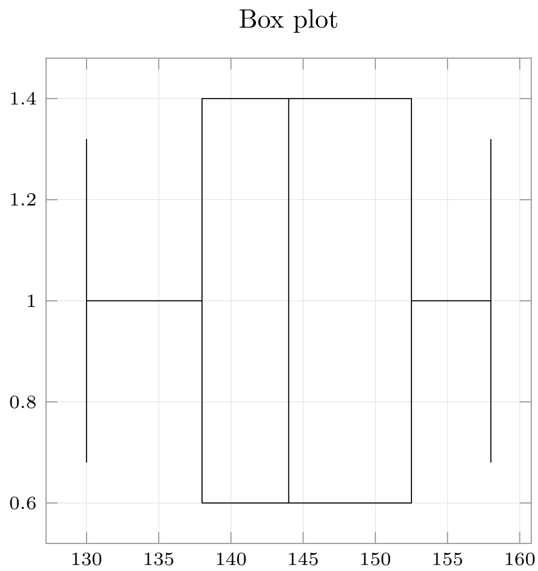
2. Representing the data,



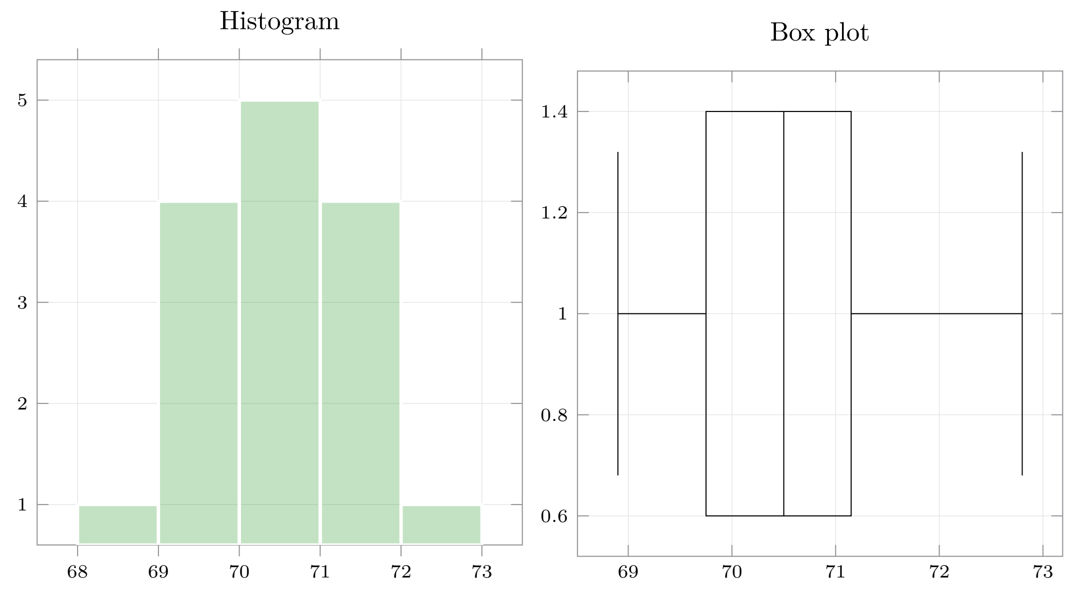


3. Representing the data,

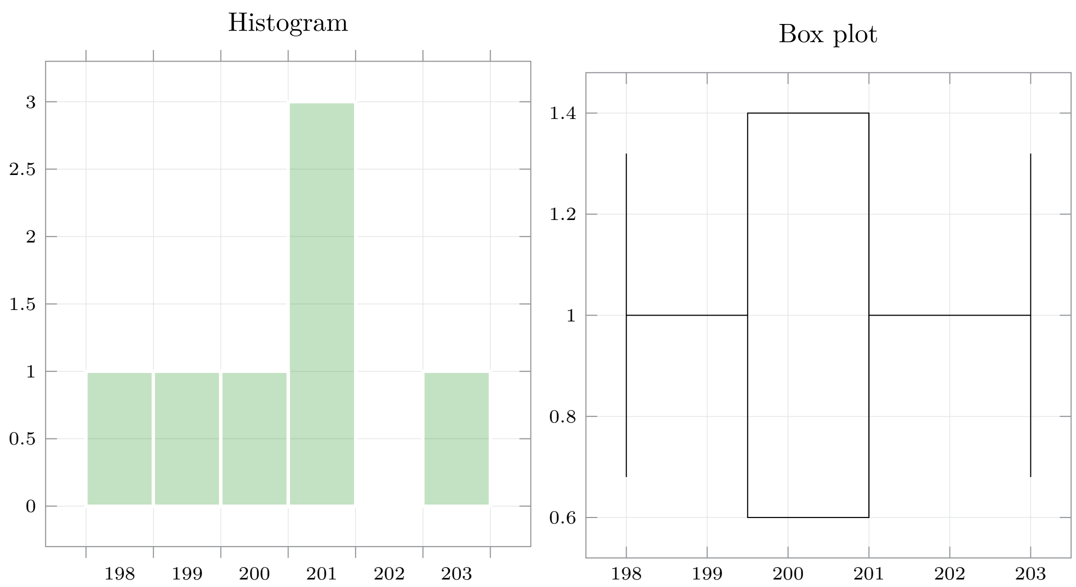




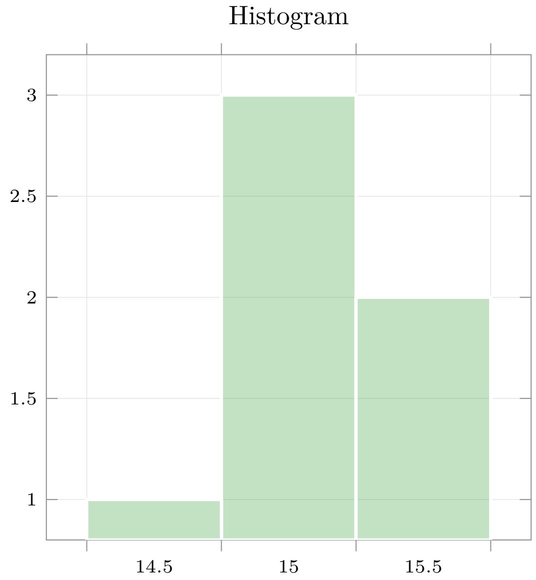
4. Representing the data,

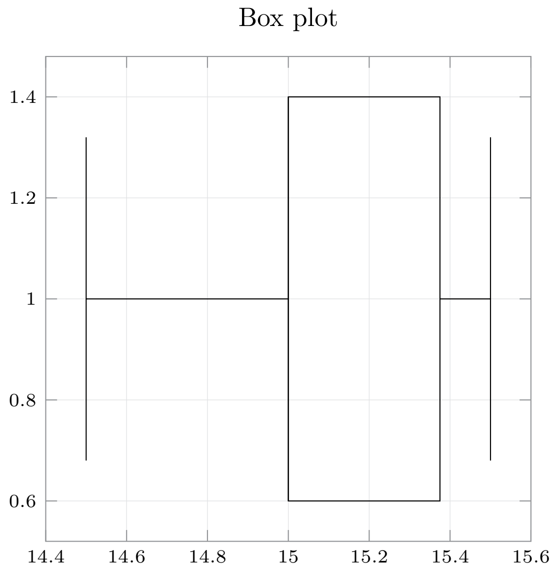


5. Representing the data,

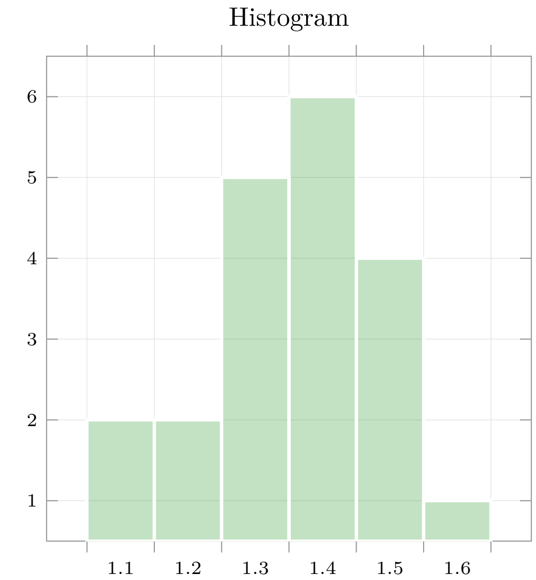


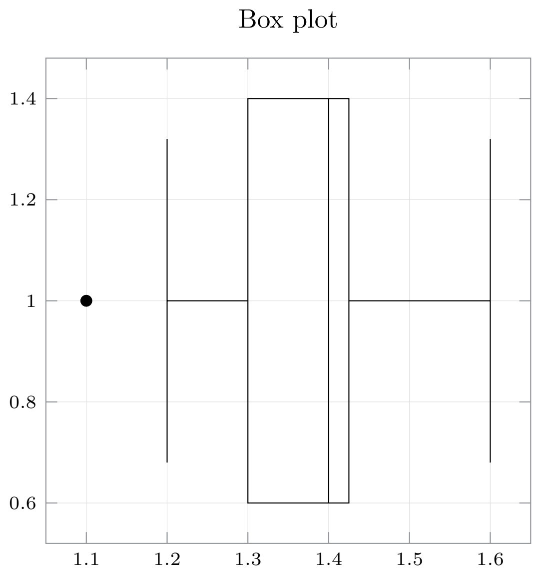
6. Representing the data,



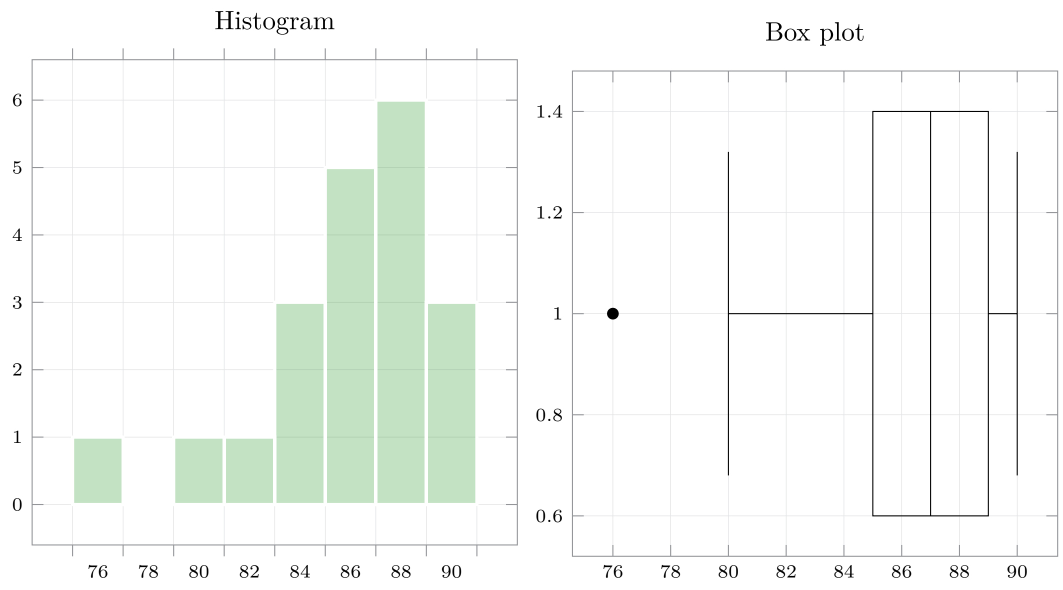


7. Representing the data, where the box plot shows outliers,

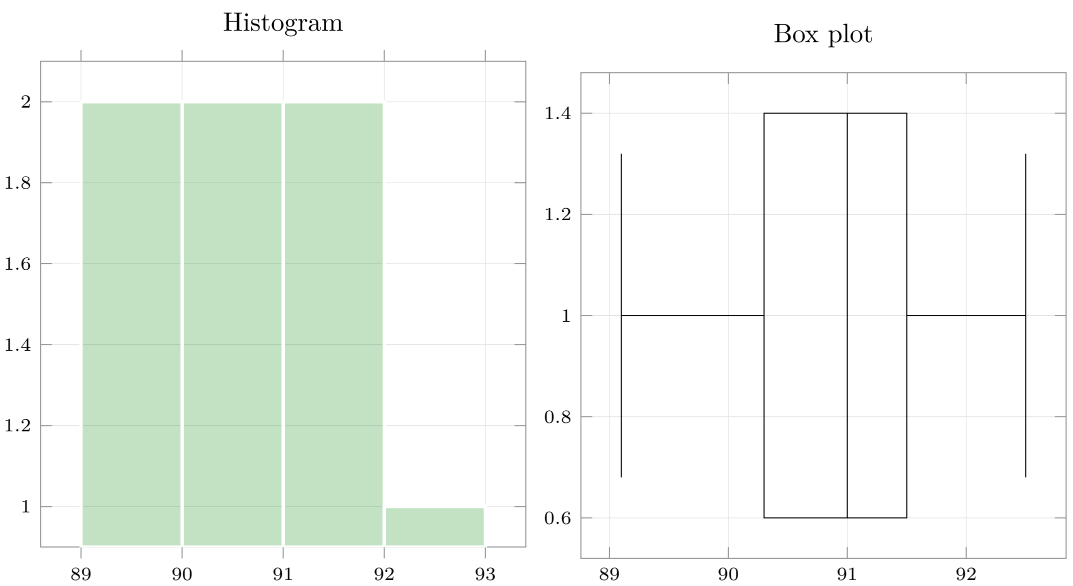




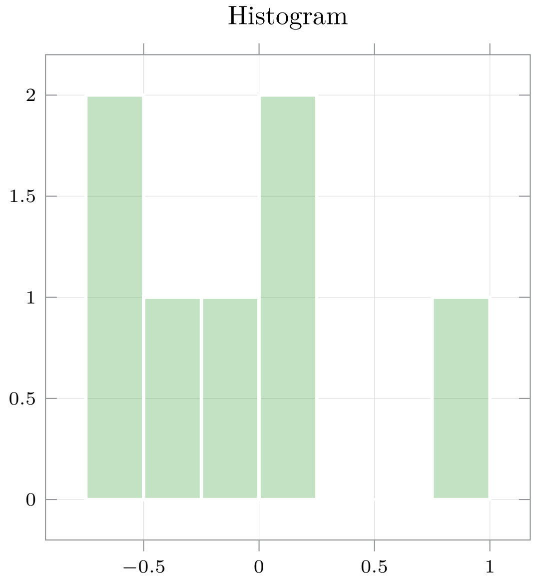
8. Representing the data, where the box plot shows outliers,

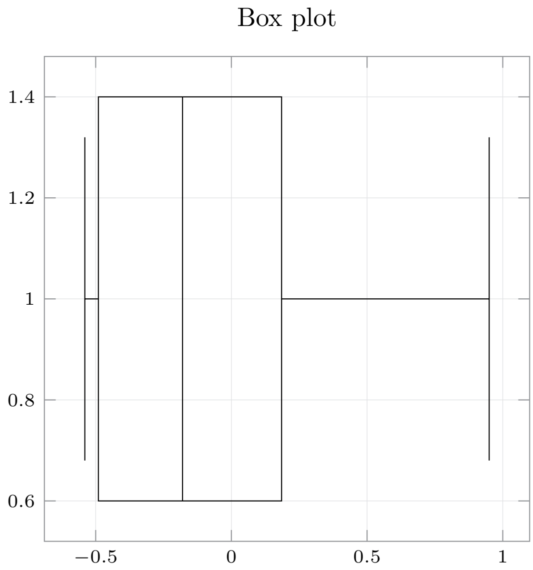


9. Representing the data, where the box plot shows outliers,



10. Representing the data, where the box plot shows outliers,





11. Looking at the statistics of the dataset,

12. Looking at the statistics of the dataset,

Quantity	Value
mean	19.875
std	0.834523
median	20
IQR	1.25

Quantity	Value
mean	4.3
std	2.540779
median	5
IQR	3.5

13. Looking at the statistics of the dataset,

14. Looking at the statistics of the dataset,

Quantity	Value
mean	144.67
std	8.973506
median	144
IQR	14.5

Quantity	Value
mean	70.48667
std	1.046673
median	70.5
IQR	1.4

15. Looking at the statistics of the dataset,

16. Looking at the statistics of the dataset,

Quantity	Value
mean	1.355
std	0.135627
median	1.4
IQR	0.125

Quantity	Value
mean	86.3
std	3.628832
median	87
IQR	4

17. The standard deviation is much smaller after removing the outlier.

$$s_1 = 3.54 \qquad s_2 = 1.29$$

24.1.1

18. The standard deviation is much smaller after removing the outlier.

$$s_1 = 3.63 \qquad\qquad s_2 = 2.77 \tag{24.1.2}$$

19. Mean is 100 and median is 0,

$$\mathbf{x} = \begin{bmatrix} 0 & 0 & 300 \end{bmatrix} \tag{24.1.3}$$

The mean can be a very bad characterization of data with severe outliers. The median becomes a much better descriptor in such cases.

20. Let $\bar{x} > x_n$ in a sorted dataset with n elements.

$$n\bar{x} = x_1 + x_2 + \cdots + x_n \qquad\qquad n\bar{x} > x_n + x_n + \cdots + x_n \tag{24.1.4}$$

$$n\bar{x} > x_1 + x_2 + \cdots + x_n \tag{24.1.5}$$

This is a contradiction, which means that the initial assumption was wrong.

By similar logic, the mean cannot be smaller than x_1

24.2 Experiments, Outcomes, Events

1. Each screw can either be left or right-handed.

$$S = \{LLL,\ RLL, LRL, LLR,\ LRR, RLR, RRL,\ RRR\} \tag{24.2.1}$$

2. Each coin can either land heads or tails.

$$S = \{HH,\ HT, TH,\ TT\} \tag{24.2.2}$$

3. Each die can land on the numbers one to six.

$$S = \begin{bmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{bmatrix} \tag{24.2.3}$$

4. Inifnitely many outcomes.

$$S = \{6, 16, 26, 36, 46, 56, 116, 126, 136, 146, 156, \dots, 556, 1116, \dots\} \tag{24.2.4}$$

5. Inifnitely many outcomes.

$$S = \{TH,\ TTH,\ TTTH,\ TTTTH,\ TTTTTH, \dots\} \tag{24.2.5}$$

6. For three nonnegative real numbers T_j,

$$S = \{T_1, T_2, T_3\} \tag{24.2.6}$$

7. For some nonnegative real number as the temperature in Kelvin and pressure in bar,

$$S = \{T, P\} \tag{24.2.7}$$

8. Using combinatorics,

$$S = \{AB, AC, AD, AE, \quad BC, BD, BE, \quad CD, CE, \quad DE\} \tag{24.2.8}$$

9. Representing the normal and defective gaskets with N and D,

$$S = \{D, ND, NND, NNND, NNNND, \dots, NNNNNNNNND\} \tag{24.2.9}$$

At best, the tenth gasket drawn will be the defective one, which means this has 10 outcomes.

10. Looking at the outcomes making belonging to both events,

$$A \cap B = \{5, 5, 5\} \tag{24.2.10}$$

The events are not mutually exclusive.

11. Looking at the outcomes making belonging to both events,

$$A = \{3, 6, 9, 12\} \qquad\qquad B = \{5, 10\} \tag{24.2.11}$$

The events are mutually exclusive.

12. The subsets are,

$$S = \{a, b, c\} \tag{24.2.12}$$

$$E = \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}, \phi \tag{24.2.13}$$

13. The events are,

$$A = \{(1,1),\ (2,2),\ (3,3),\ (4,4),\ (5,5),\ (6,6)\} \tag{24.2.14}$$

$$B = \{(1,1),\ (1,2),\ (1,3),\ (2,1),\ (2,2),\ (3,1)\} \tag{24.2.15}$$

$$A \cup B = \{(1,1),\ (1,2),\ (1,3),\ (2,1),\ (2,2),\ (3,1),\ (3,3),\ (4,4),\ (5,5),\ (6,6)\} \tag{24.2.16}$$

$$A \cap B = \{(1,1),\ (2,2)\} \tag{24.2.17}$$

$$A^{\complement} = \begin{bmatrix} & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & \end{bmatrix} \tag{24.2.18}$$

$$B^{\complement} = \begin{bmatrix} & & & (1,4) & (1,5) & (1,6) \\ & & (2,3) & (2,4) & (2,5) & (2,6) \\ & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{bmatrix} \tag{24.2.19}$$

14. The outcomes within each event are,

$$A = \{LR,\ RL,\ RR\} \qquad B = \{RL,\ LR,\ LL\} \tag{24.2.20}$$

$$C = \{RR\} \qquad D = \{LL\} \tag{24.2.21}$$

$$\tag{24.2.22}$$

A and B are not mutually exclusive. C and D are.

15. The events in the Venn diagram are,

1: see paris, don't have a good time and don't run out of money.
2: see paris, have a good time and don't run out of money.
3: don't see paris, have a good time and don't run out of money.
4: don't see paris, have a good time and run out of money.
5: don't see paris, don't have a good time and run out of money.
6: see paris, don't have a good time and run out of money.
7: see paris, have a good time and run out of money.

16. Using the definition of the complement,

$$A^{\mathsf{C}} = S - A \qquad (A^{\mathsf{C}})^{\mathsf{C}} = S - (S - A) = A \tag{24.2.23}$$

$$A \cup A^{\mathsf{C}} = A \cup (S - A) \qquad = S \tag{24.2.24}$$

$$A \cap A^{\mathsf{C}} = A \cap (S - A) \qquad = \phi \tag{24.2.25}$$

$$S^{\mathsf{C}} = S - S = \phi \qquad \phi^{\mathsf{C}} = S - \phi = S \tag{24.2.26}$$

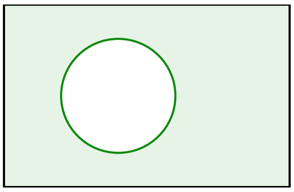
24.2.27

17. A is a subset of B. Then, there are no elements of A that are not in B. Thus, $A \cup B = B$
$A \cup B = B$. Then, there are no elements of A that are not in B. Thus, $A \subseteq B$

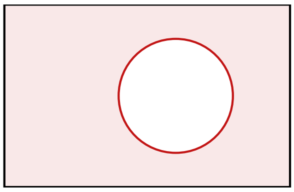
18. A is a subset of B. Then, there are no elements of A that are not in B. Thus, $A \cap B = A$
$A \cap B = A$. Then, there are no elements of A that are not in B. Thus, $A \subseteq B$

19. Checking De-Morgan's laws,

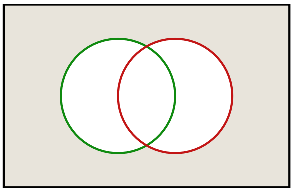
A^{C}



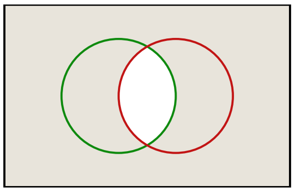
B^{C}



$(A \cup B)^{\mathsf{C}}$

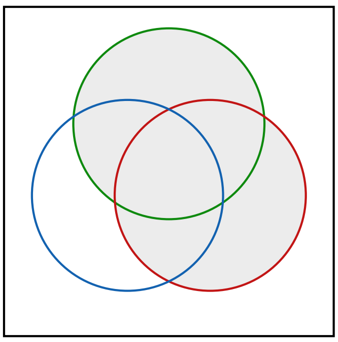


$(A \cap B)^{\mathsf{C}}$

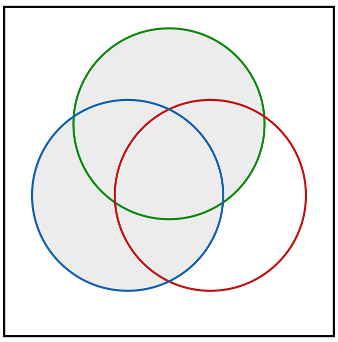


20. Checking the relations using Venn diagrams,

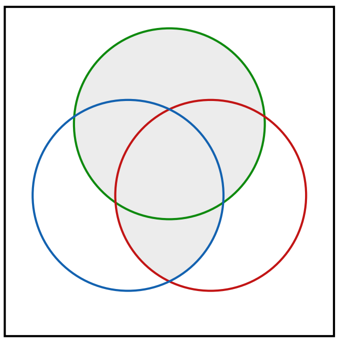
$\mathbf{A} \cup \mathbf{B}$

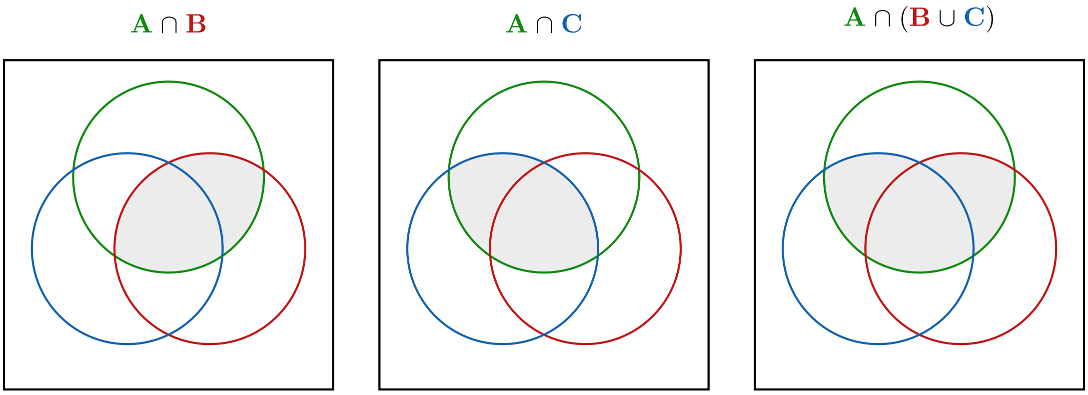


$\mathbf{A} \cup \mathbf{C}$



$\mathbf{A} \cup (\mathbf{B} \cap \mathbf{C})$





24.3 Probability

1. Three fair dice, each of which rolls a number from 1 to 6. The unwanted outcomes are,

$$A = \{(6,6,6), (6,6,5), (5,6,6), (6,5,6)\} \tag{24.3.1}$$

$$P(A^{\complement}) = 1 - P(A) = 1 - \frac{4}{6^3} = \frac{53}{54} \tag{24.3.2}$$

2. Two fair dice, each of which rolls a number from 1 to 6. The required outcomes are,

$$A = \{(2,2), (3,1), (1,3), (2,3), (3,2), (4,1), (1,4), (1,5), (2,4), (3,3), (4,2), (5,1)\} \tag{24.3.3}$$

$$P(A^{\complement}) = \frac{12}{6^2} = \frac{1}{3} \tag{24.3.4}$$

3. Drawing with replacement, $P(D) = 0.1$

$$P(A) = (0.9)^3 = 0.729 \tag{24.3.5}$$

Drawing without replacement results in a slightly smaller probability,

$$P(B) = \frac{90}{100} \cdot \frac{89}{99} \cdot \frac{88}{98} = 0.7265 \tag{24.3.6}$$

4. **(a)** Drawing with replacement, $P(D) = 0.1$ the direct method is,

$$P(E_1) = 3 \cdot 0.1 \cdot 0.9^2 + 3 \cdot (0.1)^2 \cdot 0.9 + 0.1^3 = 0.271 \tag{24.3.7}$$

whereas the complement method gives,

$$P(E_2) = 1 - (0.9)^3 = 0.271 \tag{24.3.8}$$

(b) Drawing without replacement, $P(D) = 0.1$ the direct method is,

$$P(E_1) = \frac{10}{100}\cdot\frac{90}{99}\cdot\frac{89}{98} + \frac{90}{100}\cdot\frac{10}{99}\cdot\frac{89}{98} + \frac{90}{100}\cdot\frac{89}{99}\cdot\frac{10}{98} \tag{24.3.9}$$

$$+ \frac{10}{100}\cdot\frac{9}{99}\cdot\frac{90}{98} + \frac{10}{100}\cdot\frac{90}{99}\cdot\frac{9}{98} + \frac{90}{100}\cdot\frac{10}{99}\cdot\frac{9}{98} \tag{24.3.10}$$

$$+ \frac{10}{100}\cdot\frac{9}{99}\cdot\frac{8}{98} = \frac{67}{245} \tag{24.3.11}$$

whereas the complement method gives,

$$P(E_2) = 1 - \frac{90}{100}\cdot\frac{89}{99}\cdot\frac{88}{98} = \frac{67}{245} \tag{24.3.12}$$

5. $P(L) = 1/3$ and $P(R) = 2/3$

$$P(E) = 1 - \frac{1}{3}\cdot\frac{1}{3} = \frac{8}{9} \tag{24.3.13}$$

6. $P(L) = 1/3$ and $P(R) = 2/3$ when sampling without replacement,

$$P(E) = 1 - \frac{10}{30}\cdot\frac{9}{29} = \frac{26}{29} \tag{24.3.14}$$

The probability of the complement goes down, so the probability of the event should increase.

7. When sampling from a large population and when each kind of item that can be drawn is present in much larger numbers than the number of draws.

8. $P(F) = 0.1$ and $P(S) = 0.9$

$$P(E) = 0.9^4 = 0.6561 \tag{24.3.15}$$

9. 2 out of 500 sheets contain spots. In order to draw 5 clean sheets,

$$P(E) = \frac{498}{500}\cdot\frac{497}{499}\cdot\frac{496}{498}\cdot\frac{495}{497}\cdot\frac{494}{496} = 98.008\ \% \tag{24.3.16}$$

The slight increase above 98 % comes from the non-replacement.

10. $P(F) = P(M) = 0.5$, when drawing with replacement,

$$E = \{FF, MFF, FMF, MMFF, FMMF, MMFF\} \tag{24.3.17}$$

$$= \frac{1}{4} + \frac{1}{4} + \frac{3}{16} = \frac{11}{16} \tag{24.3.18}$$

11. $N_U = N_O = 50$ and $N_D = 100$

(a) Two rods of desired length, drawing without replacement

$$P(A) = \frac{100}{200} \cdot \frac{99}{199} = \frac{99}{398} \tag{24.3.19}$$

(b) Exactly one of desired length, drawing without replacement

$$P(B) = 2 \cdot \frac{100}{200} \cdot \frac{100}{199} = \frac{100}{199} \tag{24.3.20}$$

(c) Zero of desired length, drawing without replacement

$$P(C) = \frac{100}{200} \cdot \frac{99}{199} = \frac{99}{398} \tag{24.3.21}$$

12. Let the probability of failure per switch be p,

$$P(A) = 0.99 = (1-p)^4 \qquad p = 0.251\ \% \tag{24.3.22}$$

13. Let the probability of failure per switch be p,

$$P(E) = (1-p)^4 \qquad = 1 - (1-0.04)^3 = 11.53\ \% \tag{24.3.23}$$

14. For k defective out of two drawn,

$$P(E_0) = (0.98)^2 = 0.9604 \qquad P(E_1) = 2 \cdot 0.02 \cdot 0.98 = 0.0392 \tag{24.3.24}$$

$$P(E_2) = (0.02)^2 = 0.0004 \qquad \sum_{j=0}^{2} P(E_j) = 0.9604 + 0.0392 + 0.0004 = 1 \tag{24.3.25}$$

Since these events together cover the entire sample space of drawing two plugs, their probabilities sum to unity.

15. Looking at the probabilites using the complement method,

(a) Firing one shot,

$$P(A) = 1 - \frac{1}{2} = 0.5 \tag{24.3.26}$$

(b) Firing two shot,

$$P(A) = 1 - \frac{3^2}{4^2} = 0.4375 \tag{24.3.27}$$

(c) Firing four shot,

$$P(A) = 1 - \frac{7^4}{8^4} = 0.4138 \tag{24.3.28}$$

16. Looking at the probabilites for the events,

$$P(A) = \frac{1}{2} \qquad P(B) = \frac{1}{2} \tag{24.3.29}$$

$$P(C) = \frac{1}{2} \tag{24.3.30}$$

Looking at the four relations,

$$P(A \cup B) = \frac{1}{4} \qquad P(B \cup C) = \frac{1}{4} \tag{24.3.31}$$

$$P(A \cup C) = \frac{1}{4} \qquad P(A \cup B \cup C) = 0 \tag{24.3.32}$$

the first three hold, but the fourth does not.

17. B is a subset of A,

$$A = B + (A - B) \qquad P(A) = P(B) + P(A - B) \tag{24.3.33}$$

$$P(A) \geq P(B) \tag{24.3.34}$$

where the equality only holds if the two sets are identical.

18. Let $A \cap B = D$,

$$P(D \cap C) = P(D|C) \cdot P(C) = P(D) \cdot P(C|D) \tag{24.3.35}$$

$$= P(C|A \cap B) \cdot P(A \cap B) \tag{24.3.36}$$

$$= P(C|A \cap B) \cdot P(B|A) \cdot P(A) \tag{24.3.37}$$

19. Divisibility of a number by 2, 3, 5 is the same as the corresponding digit on the coin being 1.

$$S = \{7, 15, 10, 6\} \tag{24.3.38}$$

Clearly, no number in this set is divisible by all three of the divisors, even though there exist 2 out of 4 numbers that are divisible by two of them.

24.4 Permutations and Combinations

1. Assigning 10 drivers to 10 buses,

$$A = 10 \cdot 9 \cdot 8 \cdot 1 = 10! \tag{24.4.1}$$

Assigning 10 buses to 10 drivers,

$$B = 10 \cdot 9 \cdot 8 \cdot 1 = 10! \tag{24.4.2}$$

2. 5 letters taken 2 at a time, first without then with repetition,

$$P_1 = 5 \cdot 4 = 20 \qquad P_2 = 5^2 = 25 \tag{24.4.3}$$

$$C_1 = \binom{5}{2} = \frac{5!}{2!\ 3!} = 10 \qquad C_2 = 5^2 = \binom{5+2-1}{2} = 15 \tag{24.4.4}$$

$$\tag{24.4.5}$$

3. For drawing the plastic then the rubber gaskets,

$$N = 6! = 15 \qquad P(E_1) = \frac{2!\ 4!}{N} = \frac{1}{15} \tag{24.4.6}$$

$$P(E_1) = \frac{2}{6} \cdot \frac{1}{5} = \frac{1}{15} \tag{24.4.7}$$

For drawing the rubber then the plastic gaskets,

$$N = 6! = 15 \qquad P(E_1) = \frac{4!\ 2!}{N} = \frac{1}{15} \tag{24.4.8}$$

$$P(E_1) = \frac{4}{6} \cdot \frac{3}{5} \cdot \frac{2}{4}\frac{1}{3} = \frac{1}{15} \tag{24.4.9}$$

4. Box contains $2G$, $3Y$ and $5R$ balls

$$P(G, Y, B) = \frac{2!\ 3!\ 5!}{10!} = \frac{1}{2520} \tag{24.4.10}$$

5. The total number of combinations is,

$$N = \binom{10}{3} \cdot \binom{5}{2} \cdot \binom{6}{2} = 18000 \tag{24.4.11}$$

6. The total number of combinations is,

$$N = \binom{50}{4} = \frac{47 \cdot 48 \cdot 49 \cdot 50}{24} = 230300 \tag{24.4.12}$$

7. The total number of combinations is,

$$N = \binom{10}{4} = 210 \qquad N_0 = \binom{8}{4} = 70 \tag{24.4.13}$$

$$N_1 = \binom{8}{3} \cdot \binom{2}{1} = 112 \qquad N_2 = \binom{8}{2} \cdot \binom{2}{2} = 28 \tag{24.4.14}$$

Note that $N_0 + N_1 + N_2 = N$, as in the theory of probabilities.

8. The total number of combinations is,

$$N = \binom{52}{13} = 635\ 013\ 559\ 600 \tag{24.4.15}$$

9. The total number of permutations is,

$$N = \frac{6!}{6} = 120 \tag{24.4.16}$$

A round table means that the permutations `ABCDEF` and `FABCDE` are equivalent, unlike a straight line that has a definite start and end point.

10. The total number of combinations is,

$$N = \binom{100}{10} \qquad N_1 = \binom{97}{7} \cdot \binom{3}{3} \tag{24.4.17}$$

$$P = \frac{N_1}{N} = \frac{2}{2695} \tag{24.4.18}$$

11. All three digits are different, which means there are 9 different choices each for the second and 8 leftover choices for the third digit.

$$N = 9 \cdot 8 = 72 \tag{24.4.19}$$

12. The total number of combinations is, (innocent first, guilty second).

$$N = \binom{9}{3} \tag{24.4.20}$$

$$N_1 = \binom{6}{0} \cdot \binom{3}{3} = 1 \qquad P_1 = \frac{N_1}{N} = \frac{1}{84} \tag{24.4.21}$$

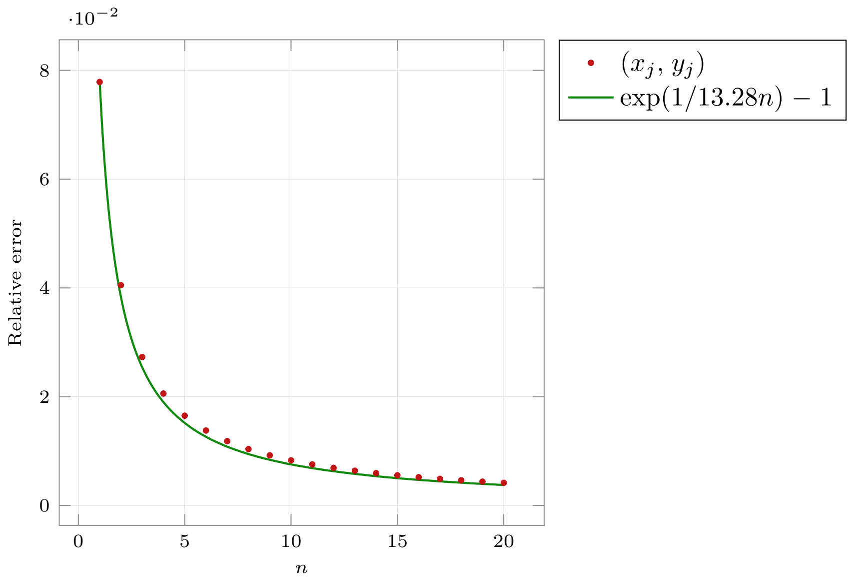
$$N_2 = \binom{6}{3} \cdot \binom{3}{0} = 1 \qquad P_2 = \frac{N_2}{N} = \frac{5}{21} \tag{24.4.22}$$

13. Stirling formula,

(a) Tabulating the errors in the Stirling approximation,

n	Error %	n	Error %
1	7.79	11	0.75
2	4.05	12	0.69
3	2.73	13	0.64
4	2.06	14	0.59
5	1.65	15	0.55
6	1.38	16	0.52
7	1.18	17	0.49
8	1.04	18	0.46
9	0.92	19	0.44
10	0.83	20	0.42

(b) An empirical formula using the exponential function is,



(c) The actual error is smaller than the upper bound on the error

$$\exp\left(\frac{1}{13.28n}\right) < \exp\left(\frac{1}{12n}\right) \tag{24.4.23}$$

(d) TBC.

14. Permutations and combinations

(a) Permutations of n different things taken k at a time gives,

- n choices for the first item
- $n-1$ choices for the second item, since one has already been taken
- $n-2$ choices for the third item, since two have already been taken
- $n-(k-1)$ choices for the k^{th} item, since $(k-1)$ items have already been taken.

By this logic, the number of permutations is,

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots (n-+1) = \frac{n!}{(n-k)!} \tag{24.4.24}$$

With repetition, the number of permutations is simply the number of choices for each draw (n) raised to the power of the number of draws (k).

(b) Consider the relation for $k=1$,

$$C_1 = n = \binom{n+1-1}{1} \tag{24.4.25}$$

$$C_2 = n + \binom{n}{2} = \frac{n(n+1)}{2} = \binom{n+2-1}{2} \tag{24.4.26}$$

$$C_3 = n + 2 \cdot \binom{n}{2} + \binom{n}{3} = \frac{n(n+1)(n+2)}{6} = \binom{n+3-1}{3} \tag{24.4.27}$$

Consider the arrangement of k identical objects separated by $(n-1)$ dividers, resulting in each of the objects belonging to one of n classes.
This is the same problem as the one being asked for, and hence has the same number of possible combinations.

$$O \;|\; |\; O\; O \;|\; O \;|\; O \tag{24.4.28}$$

is an example of 5 objects and 5 classes, represented by the 4 class dividers. The new divider can be placed in one of $(k+n)$ possible positions, for example

$$\begin{gathered} \bullet\; O \;|\; |\; O\; O \;|\; O \;|\; O \\ |\; O\; \bullet \;|\; O\; O \;|\; O \;|\; O \end{gathered} \tag{24.4.29}$$

Although the red and green new dividers result in the same final combination, they started off as different initial combinations.

This implies there is a $(k+1)$ fold overcounting.

$$C_{k+1} = C_k \cdot \frac{(n+k)}{(k+1)} = \frac{(n+k)!}{(n-1)!\ (k+1)!} = \binom{n+k}{k+1} \tag{24.4.30}$$

By induction, the relation is true for all k.

(c) Deriving the relation, for some non-negative integer k,

$$\binom{a}{k} + \binom{a}{k+1} = \frac{a!}{(a-k)!\ k!} + \frac{a!}{(a-k-1)!\ (k+1)!} \tag{24.4.31}$$

$$= \frac{a!}{(a-k)!\ k!}\left[1 + \frac{(a-k)}{(k+1)}\right] \tag{24.4.32}$$

$$= \frac{a!}{(a-k)!\ k!}\left[\frac{a+1}{(k+1)}\right] \tag{24.4.33}$$

$$= \frac{(a+1)!}{(a-k)!\ (k+1)!} = \binom{a+1}{k+1} \tag{24.4.34}$$

(d) Out of the n factors $(a+b)$, choose k factors from which to take a and the then take b from the remaining $(n-k)$ factors. The result is,

$$T_k = \binom{n}{k} a^k\ b^{n-k} \tag{24.4.35}$$

This is why the coefficient of the term $a^k b^{n-k}$ is the number of combinations of n objects taken k at a time.

(e) Proving the relation,

$$(a+b)^{p+q} = (a+b)^p\ (a+b)^q \tag{24.4.36}$$

$$\left[\sum_{r=0}^{p+q} \binom{p+q}{r} a^r\ b^{p+q-r}\right] = \left[\sum_{m=0}^{p} \binom{p}{m} a^m\ b^{p-m}\right] \left[\sum_{n=0}^{q} \binom{q}{n} a^n\ b^{q-n}\right] \tag{24.4.37}$$

Matching coefficients of a^r,

$$\binom{p+q}{r} = \binom{p}{0}\binom{q}{r} + \binom{p}{1}\binom{q}{r-1} + \dots + \binom{p}{r}\binom{q}{0} \tag{24.4.38}$$

$$= \sum_{k=0}^{r} \binom{p}{k}\ \binom{q}{r-k} \tag{24.4.39}$$

(f) TBC

15. Birthday problem, with 20 people. The probability of all their birthdays being distinct is,

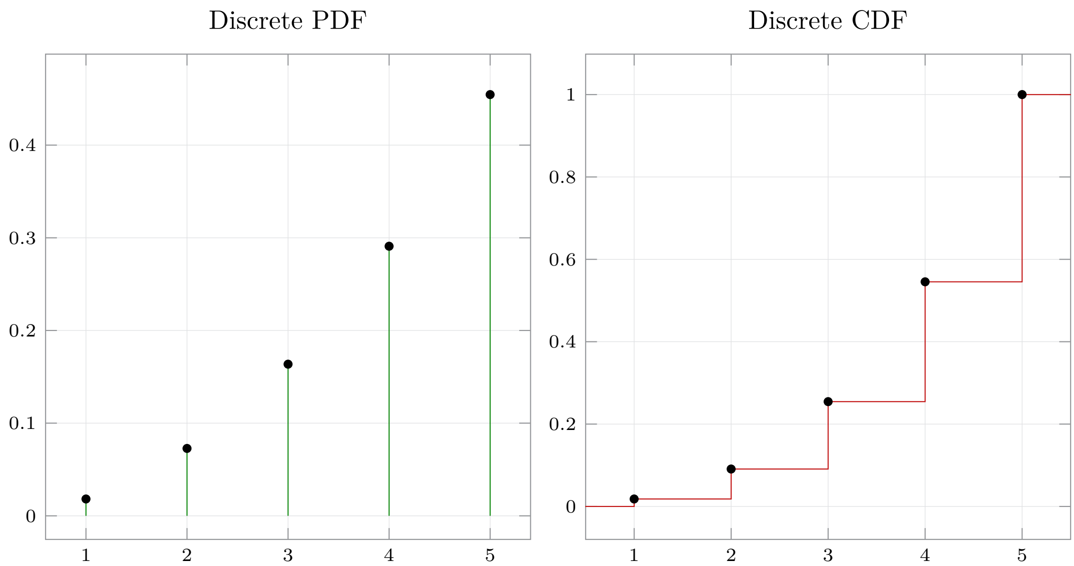
$$P(E) = \binom{365}{20} \cdot 20! \cdot \frac{1}{365^{20}} = 0.5886 \qquad \text{24.4.40}$$

The event required is the complement of E, which means the probability of at least two birthdays coinciding is 0.4114.

24.5 Random Variables. Probability Distributions

1. Using the normalization condition,

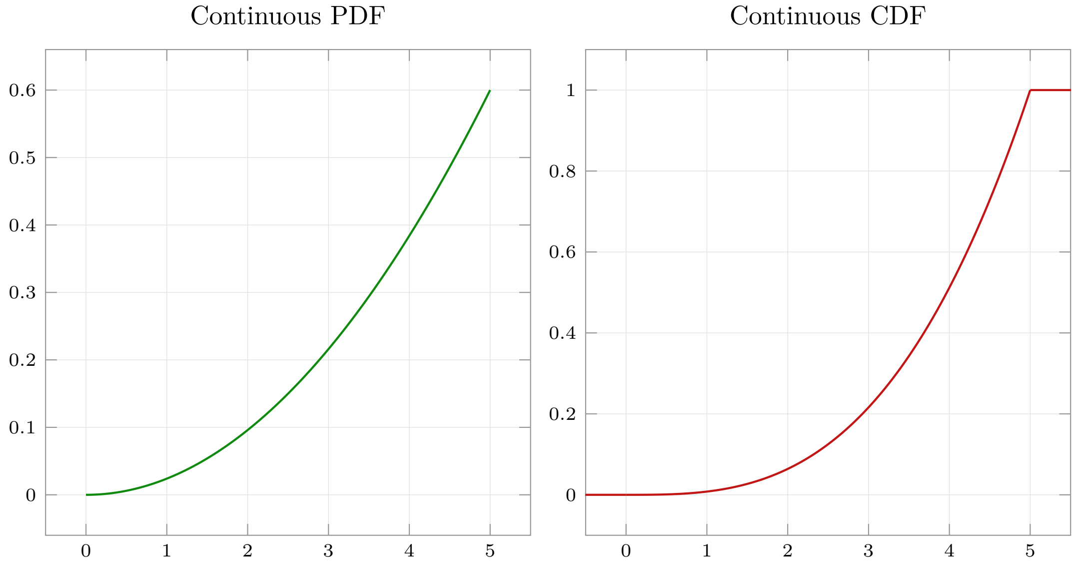
$$1 = \sum_{j=1}^{5} p_j = k(1 + 4 + 9 + 16 + 25) \qquad k = \frac{1}{55} \qquad \text{24.5.1}$$



2. Using the normalization condition,

$$1 = \int_0^5 kv^2 \ \mathrm{d}v \qquad k = \frac{3}{125} \qquad \text{24.5.2}$$

$$F(x) = \int_0^x f(v) \ \mathrm{d}v \qquad = \left[\frac{v^3}{125}\right]_0^x = \frac{x^3}{125} \qquad \text{24.5.3}$$

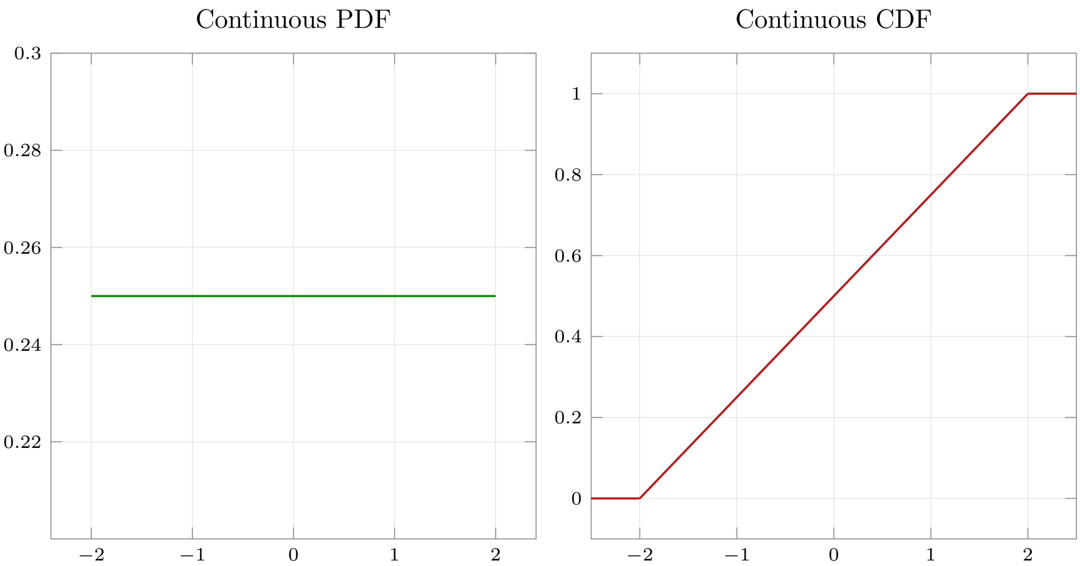


3. Using the normalization condition,

$$1 = \int_{-2}^{2} k \ \ \mathrm{d}v \qquad\qquad k = \frac{1}{4} \tag{24.5.4}$$

$$F(x) = \int_{-2}^{x} f(v) \ \ \mathrm{d}v \qquad\qquad F(x) = \left[\frac{v}{4}\right]_{-2}^{x} = \frac{x+2}{4} \tag{24.5.5}$$

$$P(0 \leq X \leq 2) = F(2) - F(0) = \frac{1}{2} \tag{24.5.6}$$



4. Finding the unknown constants,

$$P(-c < X < c) = F(c) - F(-c) = \frac{c + 2 - (-c + 2)}{4} \qquad 0.95 = \frac{c}{2} \tag{24.5.7}$$

$$c = 1.9 \tag{24.5.8}$$

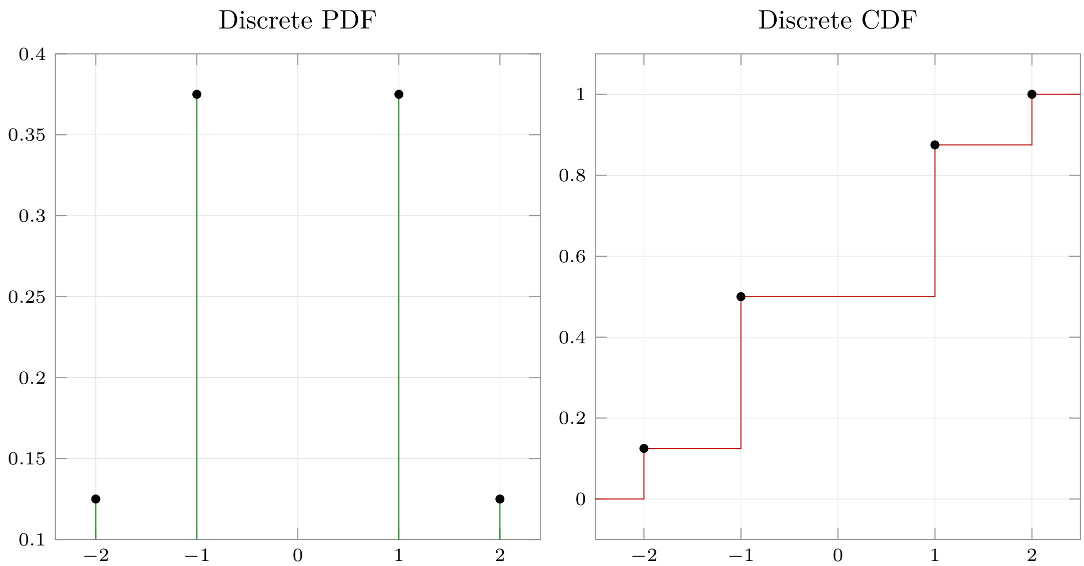
For the second part of the problem,

$$P(0 < X < d) = F(d) - F(0) = \frac{d + 2 - (2)}{4} \qquad 0.95 = \frac{d}{4} \tag{24.5.9}$$

$$d = 3.8 \tag{24.5.10}$$

Such a value d does not exist, since the sample space S is only the interval $[-2, 2]$.

5. The plots are,



No since the probabilities have already summed to unity.

6. The probabilities are,

$$P(X = 0) = \frac{4}{10} \cdot \frac{3}{9} = \frac{2}{15} \qquad P(X = 1) = \frac{6}{10} \cdot \frac{4}{9} + \frac{4}{10} \cdot \frac{6}{9} = \frac{8}{15} \tag{24.5.11}$$

$$P(X = 2) = \frac{6}{10} \cdot \frac{5}{9} = \frac{1}{3} \tag{24.5.12}$$

Now, the other probabilities are simply sums of the above discrete PDF.

$$P(1 < X < 2) = 0 \qquad P(X \le 1) = \frac{2}{15} + \frac{8}{15} = \frac{2}{3} \tag{24.5.13}$$

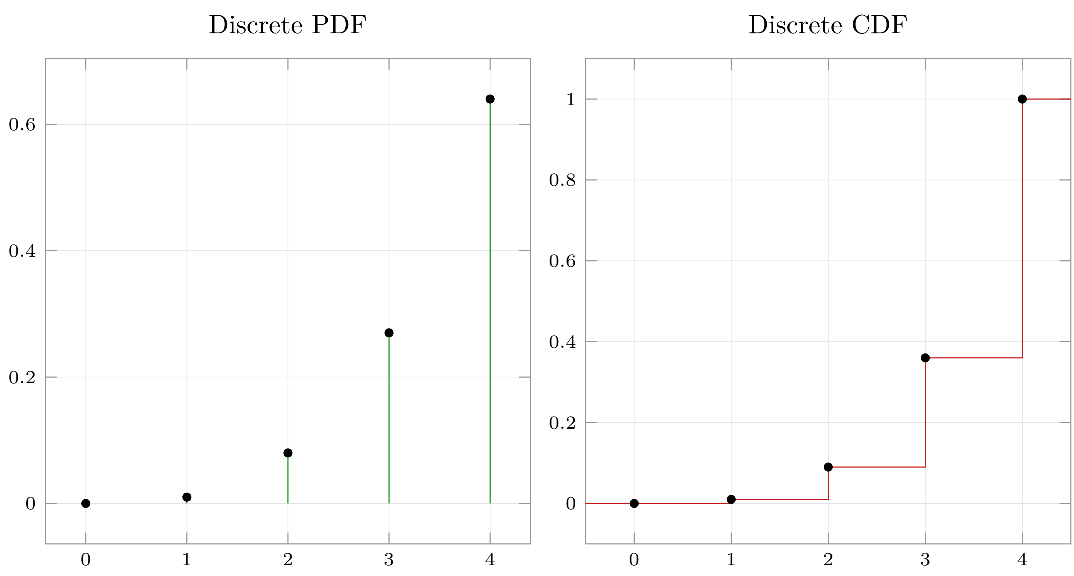
$$P(X \ge 1) = \frac{8}{15} + \frac{1}{3} = \frac{13}{15} \qquad P(X > 1) = \frac{1}{3} \tag{24.5.14}$$

$$P(0.5 < X < 10) = \frac{8}{15} + \frac{1}{3} = \frac{13}{15} \tag{24.5.15}$$

7. Using the normalization condition,

$$1 = k\,(1^3 + 2^3 + 3^3 + 4^3) \qquad k = \frac{1}{100} \tag{24.5.16}$$

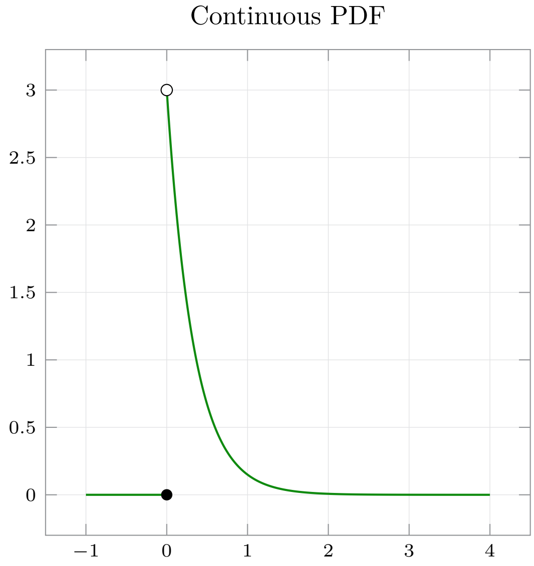
$$\tag{24.5.17}$$

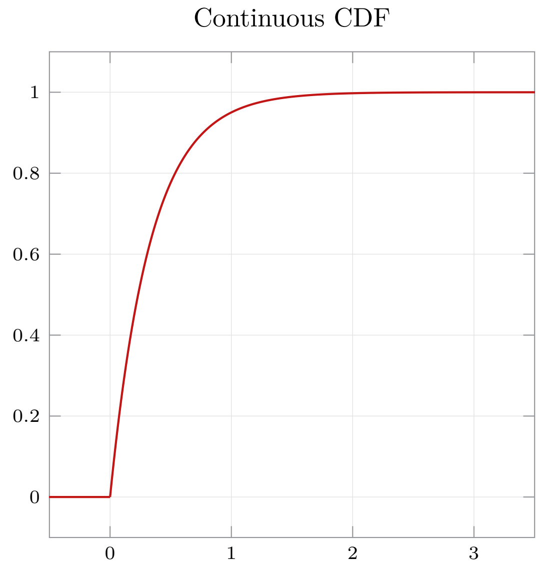


8. Using the normalization condition,

$$F(x) = \begin{cases} 1 - e^{-3x} & x > 0 \\ 0 & x \le 0 \end{cases} \qquad f(x) = F'(x) = \begin{cases} 3e^{-3x} & x > 0 \\ 0 & x \le 0 \end{cases} \tag{24.5.18}$$

$$F(x^*) = 0.9 \qquad x^* = \frac{\ln(0.1)}{-3} = 0.767 \tag{24.5.19}$$





9. Using the normalization condition

$$\int_{0.9}^{1.1} f(v) \ \mathrm{d}v = 1 \qquad 1 = \left[\frac{kv^2}{2}\right]_{0.9}^{1.1} \tag{24.5.20}$$

$$k = \frac{2}{(1.1^2 - 0.9^2)} = 5 \tag{24.5.21}$$

$$P(0.95 < X < 1.05) = \frac{5(1.05^2 - 0.95^2)}{2} = \frac{1}{2} \tag{24.5.22}$$

10. Using the normalization condition

$$\int_{119.9}^{120.1} f(v) \ \mathrm{d}v = 1 = \Big[kv\Big]_{119.9}^{120.1} \tag{24.5.23}$$

$$k = 5 \tag{24.5.24}$$

$$P(X < 119.91) = 5(119.91 - 119.9) = 0.05 \tag{24.5.25}$$

$$P(X > 120.09) = 5(120.1 - 120.09) = 0.05 \tag{24.5.26}$$

$$N = 500 * (0.05 + 0.05) = 50 \tag{24.5.27}$$

11. Let q be the probability of a single bulb not failing,

$$f(x) = \begin{cases} 6\,[0.25 - (x-1.5)^2] & x \in [1,2] \\ 0 & \text{otherwise} \end{cases} \tag{24.5.28}$$

$$P(X > 1.5) = \int_{1.5}^{2} f(v)\ \mathrm{d}v = \left[1.5v - 2(x-1.5)^3\right]_{1.5}^{2} \tag{24.5.29}$$

$$= 0.5 = q \tag{24.5.30}$$

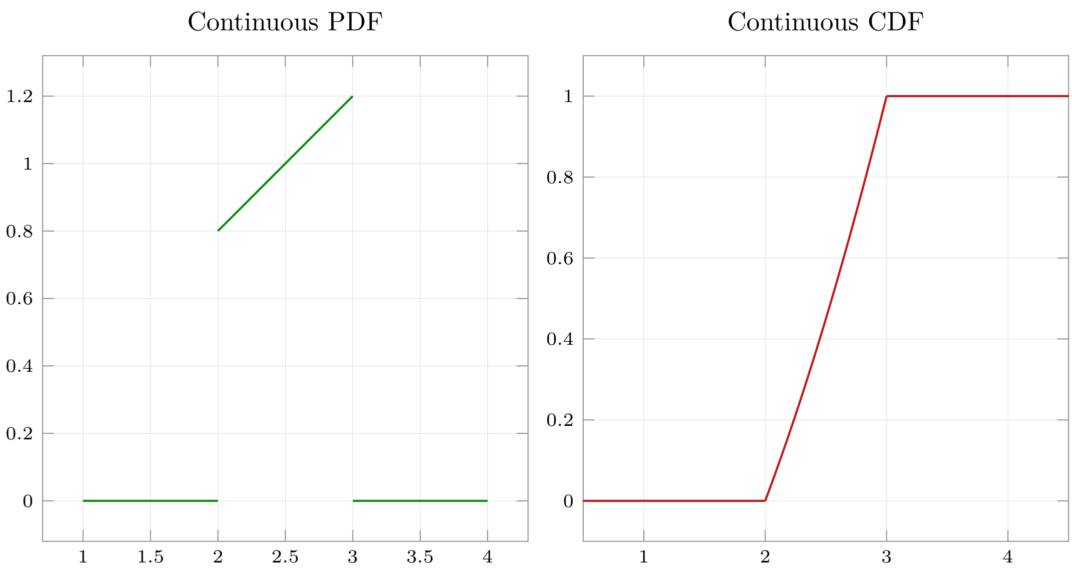
Since each bulb operates independently of the others,

$$P(E) = q^3 = \frac{1}{8} \tag{24.5.31}$$

12. The PDF is,

$$f(x) = F'(x) = \begin{cases} 0 & x < 2 \\ 0.4x & x \in [2,3) \\ 0 & x \geq 3 \end{cases} \tag{24.5.32}$$

$$P(2.5 < X < 5) = F(5) - F(2.5) = 1 - \frac{2.25}{5} = 0.55 \tag{24.5.33}$$



13. The PDF is,

$$f(x) = \begin{cases} 1 - |x| & |x| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{24.5.34}$$

$$F(x) = \int_{-1}^{x} f(v) \ \mathrm{d}v = \left[v + \frac{v^2}{2} \right]_{-1}^{x} = \frac{1}{2} + x + \frac{x^2}{2} \qquad x < 0 \tag{24.5.35}$$

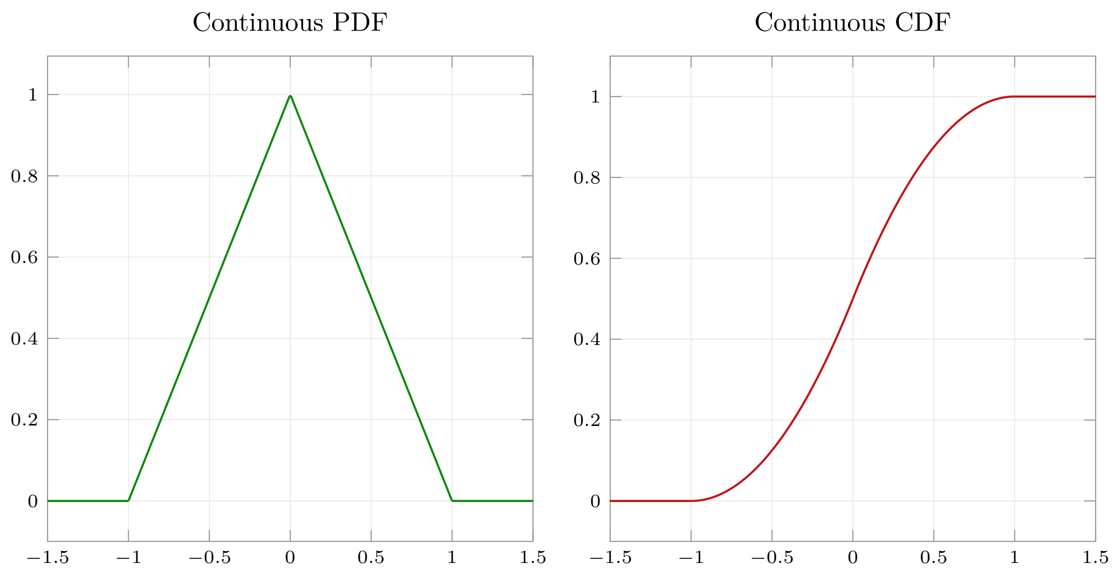
$$F(x) = \int_{-1}^{x} f(v) \ \mathrm{d}v = \frac{1}{2} + \left[v - \frac{v^2}{2} \right]_{0}^{x} = \frac{1}{2} + x - \frac{x^2}{2} \qquad x > 0 \tag{24.5.36}$$

For a total of 1000 cans,

$$P(X > 0)P(Y > 100) = F(1) - F(0) = 0.5 \qquad n_1 = 500 \tag{24.5.37}$$

$$P(X < -0.5) = P(Y < 99.5) = F(-0.5) = \frac{1}{8} \tag{24.5.38}$$

$$P(X < -1) = P(Y < 99) = F(-1) = 0 \tag{24.5.39}$$



14. Let X be the number of die rolls including the first six,

$$P(X = 1) = \frac{1}{6} \qquad P(X = 2) = \frac{5}{6} \cdot \frac{1}{6} = \frac{5}{6^2} \tag{24.5.40}$$

$$P(X = 3) = \frac{5^2}{6^2} \cdot \frac{1}{6} \qquad P(X = 4) = \frac{5^3}{6^3} \cdot \frac{1}{6} \tag{24.5.41}$$

$$P(X = k) = \frac{1}{6} \cdot \left(\frac{5}{6} \right)^{k-1} \tag{24.5.42}$$

The sum of all probabilities is,

$$\sum_{k=1}^{\infty} P(X=k) = \frac{1}{6} \sum_{k=1}^{\infty} (5/6)^{k-1} \tag{24.5.43}$$

$$= \frac{1}{6} \cdot \frac{1}{(1-5/6)} = 1 \tag{24.5.44}$$

15. The complements are,

$$X > b \qquad X \geq b \tag{24.5.45}$$

$$X < c \qquad X \leq c \tag{24.5.46}$$

$$X < b \quad \text{and} \quad X > c \qquad X \leq b \quad \text{and} \quad X > c \tag{24.5.47}$$

24.6 Mean and Variance of a Distribution

1. Using the normalization condition,

$$1 = \int_0^2 kv \ \mathrm{d}v \qquad 1 = \left[\frac{kv^2}{2}\right]_0^2 \tag{24.6.1}$$

$$k = 1/2 \qquad f(x) = \frac{x}{2} \tag{24.6.2}$$

Now, the mean and variance are,

$$\mu = \int_0^2 \frac{x^2}{2} \ \mathrm{d}x = \left[\frac{x^3}{6}\right]_0^2 = \frac{4}{3} \tag{24.6.3}$$

$$\sigma^2 = \int_0^2 \frac{x}{2} \ (x-4/3)^2 \ \mathrm{d}x = \int_0^2 \frac{x^3 - 8x^2/3 + 16x/9}{2} \ \mathrm{d}x \tag{24.6.4}$$

$$= \left[\frac{x^4}{8} - \frac{4x^3}{9} + \frac{4x^2}{9}\right]_0^2 = \frac{2}{9} \tag{24.6.5}$$

2. The PDF is,

$$f(x) = \frac{1}{6} \qquad \forall \qquad X \in \{1, 2, 3, 4, 5, 6\} \tag{24.6.6}$$

Now, the mean and variance are,

$$\mu = \sum_j x_j \cdot f(x_j) = \frac{1+2+3+4+5+6}{6} = 3.5 \tag{24.6.7}$$

$$\sigma^2 = \sum_j (x_j - \mu)\ f(x_j) = \frac{(-2.5)^2 + (-1.5)^2 + (-0.5)^2 + (0.5)^2 + (1.5)^2 + (2.5)^2}{6} \tag{24.6.8}$$

$$= \frac{35}{12} \tag{24.6.9}$$

3. Using the normalization condition,

$$1 = \int_0^{2\pi} k \ \mathrm{d}v \qquad\qquad 1 = \Big[kv\Big]_0^{2\pi} \tag{24.6.10}$$

$$k = \frac{1}{2\pi} \qquad\qquad f(x) = \frac{1}{2\pi} \tag{24.6.11}$$

Now, the mean and variance are,

$$\mu = \int_0^{2\pi} \frac{x}{2\pi} \ \mathrm{d}x = \left[\frac{x^2}{4\pi}\right]_0^{2\pi} = \pi \tag{24.6.12}$$

$$\sigma^2 = \int_0^{2\pi} \frac{1}{2\pi} \ (x-\pi)^2 \ \mathrm{d}x = \left[\frac{(x-\pi)^3}{6\pi}\right]_0^{2\pi} = \frac{\pi^2}{3} \tag{24.6.13}$$

4. Since this is a linear transformation in the RV,

$$\mu_Y = a \ \mu_X + b \qquad\qquad \mu_Y = \sqrt{3} - \sqrt{3} = 0 \tag{24.6.14}$$

$$\sigma_Y^2 = a^2 \ \sigma_X^2 \qquad\qquad \sigma_Y^2 = 1 \tag{24.6.15}$$

5. The mean and variance are,

$$f(x) = 4e^{-4x} \qquad \forall \qquad x \geq 0 \tag{24.6.16}$$

$$\mu = \int_0^{\infty} 4x \ e^{-4x} \ \mathrm{d}x = \left[\frac{e^{-4x}}{-4}\right]_0^{\infty} = \frac{1}{4} \tag{24.6.17}$$

$$\sigma^2 = \int_0^{\infty} (x - 1/4)^2 \ (4e^{-4x}) \ \mathrm{d}x = \left[\frac{-e^{-4x}}{16} \ (1 + 16x^2)\right]_0^{\infty} = \frac{1}{16} \tag{24.6.18}$$

6. Using the normalization condition,

$$1 = \int_{-1}^{1} k(1 - v^2) \ \mathrm{d}v \qquad 1 = \left[k - \frac{kv^3}{3}\right]_{-1}^{1} \tag{24.6.19}$$

$$k = \frac{1}{4/3} \qquad f(x) = \frac{3(1 - x^2)}{4} \tag{24.6.20}$$

The mean and variance are,

$$\mu = \int_{-1}^{1} \frac{3x}{4} \ (1 - x^2) \ \mathrm{d}x = \frac{3}{4}\left[\frac{x^2}{2} - \frac{x^4}{4}\right]_{-1}^{1} = 0 \tag{24.6.21}$$

$$\sigma^2 = \frac{3}{4}\int_{-1}^{1} (x - 0)^2 \ (1 - x^2) \ \mathrm{d}x = \frac{3}{4}\left[\frac{x^3}{3} - \frac{x^5}{5}\right]_{-1}^{1} = \frac{1}{5} \tag{24.6.22}$$

7. Using the normalization condition,

$$1 = \int_{0}^{\infty} Ce^{-v/2} \ \mathrm{d}v \qquad 1 = \left[- 2C \ e^{-v/2}\right]_{0}^{\infty} \tag{24.6.23}$$

$$C = \frac{1}{2} \qquad f(x) = \frac{e^{-v/2}}{2} \tag{24.6.24}$$

The mean and variance are,

$$\mu = \int_{0}^{\infty} \frac{xe^{-x/2}}{2} \ \mathrm{d}x = \left[e^{-x/2} \ (2 - x)\right]_{0}^{\infty} = 2 \tag{24.6.25}$$

$$\sigma^2 = \int_{0}^{\infty} (x - 2)^2 \ \frac{e^{-x/2}}{2} \ \mathrm{d}x = \left[- (x^2 + 4)e^{-x/2}\right]_{0}^{\infty} = 4 \tag{24.6.26}$$

8. The discrete PDF uses the fact that $(k - 1)$ tails precede the first head,

$$P(X = k) = \left(\frac{1}{2}\right)^{k-1} \cdot \frac{1}{2} \tag{24.6.27}$$

$$\mu = \sum_{k=1}^{\infty} x_j \cdot f(x_j) \qquad \mu = \frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \dots \tag{24.6.28}$$

$$\frac{\mu}{2} = \frac{1}{2^2} + \frac{2}{2^3} + \frac{3}{2^4} + \dots \qquad \mu - \frac{\mu}{2} = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots \tag{24.6.29}$$

$$\mu = 2 \tag{24.6.30}$$

9. Using the normalization condition,

$$1 = \int_{0.9}^{1.1} -k(v-0.9)(v-1.1)\ \mathrm{d}v \qquad 1 = -k\left[\frac{v^3}{3} - v^2 + 0.99v\right]_{0.9}^{1.1} \tag{24.6.31}$$

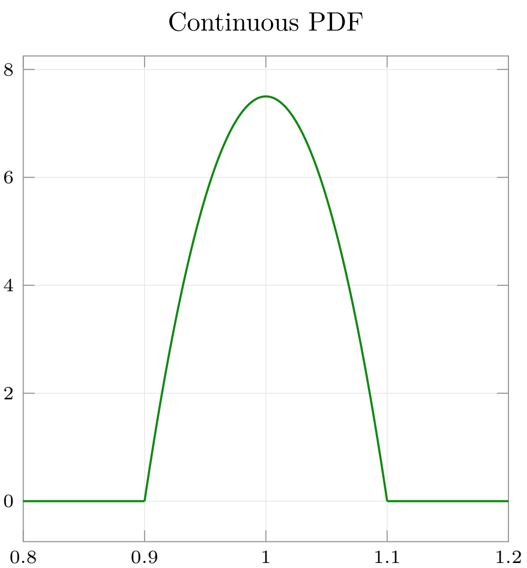
$$k = 750 \qquad f(x) = -750(x-0.9)(x-1.1) \tag{24.6.32}$$

The mean and variance are,

$$\mu = \int_{0.9}^{1.1} (-750)(x-0.9)(x-1.1)x\ \mathrm{d}x = \left[\frac{-750x^4 + 2000x^3 - 1485x^2}{4}\right]_{0.9}^{1.1} = 1 \tag{24.6.33}$$

$$\sigma^2 = \int_{0.9}^{1.1} (x-1)^2\ (-750)(x-0.9)(x-1.1)\ \mathrm{d}x \tag{24.6.34}$$

$$= \left[\frac{-300x^5 + 1500x^4 - 2995x^3 + 2985x^2 - 1485x}{2}\right]_0^{\infty} = \frac{1}{500} \tag{24.6.35}$$



10. Using the symmetry of the PDF,

$$P(E) = 2 \cdot P(X > 1.06) = 2\int_{1.06}^{1.1} (-750)(x-0.9)(x-1.1)\ \mathrm{d}x \tag{24.6.36}$$

$$= \frac{26}{125} = 20.8\ \% \tag{24.6.37}$$

11. Using the symmetry of the PDF, let the maximum permissible deviation be k,

$$P(E) = 2 \cdot P(X > 1 + k) = 2 \int_{1+k}^{1.1} (-750)(x - 0.9)(x - 1.1) \;\mathrm{d}x \tag{24.6.38}$$

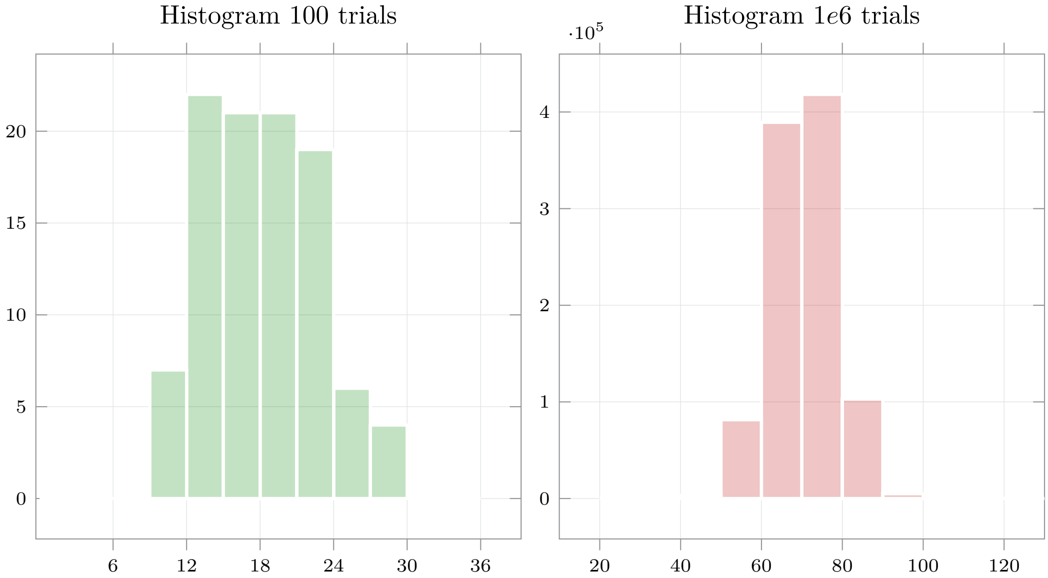
$$= \Big[-500x^3 + 1500x^2 - 1485x \Big]_{1+k}^{1.1} = 500k^3 - 15k + 1 \tag{24.6.39}$$

If this probability is to be 0.1, then the solutions of this cubic equation are,

$$P(E) = 0.1 \qquad\qquad k = \{-0.1977, 0.073, 0.125\} \tag{24.6.40}$$

Since $k \in [0, 0.1]$, the only usable value is $k^* = 0.073$

12. Plotting a histogram of the sum of 6 fair dice rolled together,



13. The expected number of units sold per day is,

$$\mathbb{E}[X] = 0.1(10) + 0.3(11) + 0.4(12) + 0.2(13) = 11.7 \tag{24.6.41}$$

$$P = 55(11.7) = 643.5 \text{ \$} \tag{24.6.42}$$

14. Finding the expected value,

$$g(X) = X^2 \qquad\qquad \mathbb{E}[g(X)] = \mathbb{E}[X^2] \tag{24.6.43}$$

$$\mathbb{E}[X^2] = \int_{-1}^{1} x^2 \cdot \frac{1}{2} \;\mathrm{d}x \qquad\qquad = \frac{1}{3} \tag{24.6.44}$$

15. The mean and variance are,

$$\mu = \int_0^1 6x(1-x)x \, \mathrm{d}x = \left[2x^3 - \frac{3x^4}{2}\right]_0^1 = \frac{1}{2} \tag{24.6.45}$$

$$\sigma^2 = \int_0^1 (x-0.5)^2 \, (6x)(1-x) \, \mathrm{d}x \tag{24.6.46}$$

$$= \left[\frac{-24x^5 + 60x^4 - 50x^3 + 15x^2}{20}\right]_0^1 = \frac{1}{20} \tag{24.6.47}$$

$$Z = \frac{X - 0.5}{\sqrt{0.05}} \tag{24.6.48}$$

16. Using the heuristic for normally distributed variables,

$$P(X > k) = \int_k^1 6x(1-x) \, \mathrm{d}x = 2k^3 - 3k^2 + 1 \tag{24.6.49}$$

$$P(X > k) = 0.05 \quad \Longrightarrow \quad k = 0.865 \tag{24.6.50}$$

This is the only root of the cubic equation that lies in the interval [0, 1].

17. Since this is a discrete PDF,

$$\mathbb{E}[X] = \frac{1 + 2\cdot 2 + 3\cdot 2 + 4\cdot 3 + 5\cdot 2 + 6\cdot 4}{36} \tag{24.6.51}$$

$$+ \frac{8\cdot 2 + 9\cdot 1 + 10\cdot 2 + 12\cdot 4}{36} \tag{24.6.52}$$

$$+ \frac{15\cdot 2 + 16\cdot 1 + 18\cdot 2}{36} + \frac{20\cdot 2 + 24\cdot 2 + 25\cdot 1 + 30\cdot 2 + 36\cdot 1}{36} \tag{24.6.53}$$

$$= \frac{49}{4} \tag{24.6.54}$$

18. The mean life is,

$$\mathbb{E}[X] = \int_0^\infty 0.001x \; e^{-0.001x} \, \mathrm{d}x = \left[-(x + 1000)e^{-0.001x}\right]_0^\infty \tag{24.6.55}$$

$$= 1000 \text{ h} \tag{24.6.56}$$

19. The expected value is,

$$\mathbb{E}[X^3] = \frac{0^3 \cdot 1 + 1^3 \cdot 3 + 2^3 \cdot 3 + 3^3 \cdot 1}{8} = \tag{24.6.57}$$

$$= \frac{27}{4} \tag{24.6.58}$$

20. Deriving,

(a) The relation is,

$$\mathbb{E}[X - \mu] = \int_{-\infty}^{\infty} (x - \mu)\ f(x)\ \ \mathrm{d}x \tag{24.6.59}$$

$$= \int_{-\infty}^{\infty} x\ f(x)\ \ \mathrm{d}x - \mu\ \int_{-\infty}^{\infty} f(x)\ \ \mathrm{d}x \tag{24.6.60}$$

$$= \mu - \mu(1) = 0 \tag{24.6.61}$$

(b) Proving,

$$\mathbb{E}[X^1] = \int_{-\infty}^{\infty} x^1\ f(x)\ \ \mathrm{d}x = \mu \tag{24.6.62}$$

$$\mathbb{E}[(X - \mu)^2] = \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2)\ f(x)\ \ \mathrm{d}x = \mathbb{E}[X^2] - \mu^2 \tag{24.6.63}$$

$$\mathbb{E}[1] = \mathbb{E}[X^0] = \int_{-\infty}^{\infty} x^0\ f(x)\ \ \mathrm{d}x = 1 \tag{24.6.64}$$

(c) The moments are,

$$\mathbb{E}[X^k] = \int_{-\infty}^{\infty} \frac{x^k}{b - a}\ \ \mathrm{d}x = \left[\frac{x^{k+1}}{(b-a)(k+1)}\right]_b^a \tag{24.6.65}$$

$$= \frac{1}{(k+1)} \cdot \frac{b^{k+1} - a^{k+1}}{b - a} \tag{24.6.66}$$

(d) A symmetric distribution is symmetric about its mean.

$$\mathbb{E}[(X-\mu)^3] = \int_{-\infty}^{\infty} (x-\mu)^3 \; f(x) \; \mathrm{d}x \qquad v = x - \mu \tag{24.6.67}$$

$$\mathbb{E}[(X-\mu)^3] = \int_{-\infty}^{0} v^3 f(v+\mu) \; \mathrm{d}v + \int_{0}^{\infty} v^3 f(v+\mu) \; \mathrm{d}v \tag{24.6.68}$$

$$= \int_{0}^{\infty} (-v)^3 f(-v+\mu) \; \mathrm{d}v + \int_{0}^{\infty} v^3 f(v+\mu) \; \mathrm{d}v \tag{24.6.69}$$

$$= 0 \tag{24.6.70}$$

provided, $f(\mu + v) = f(\mu - v)$, which is a result of the symmetry of the PDF.

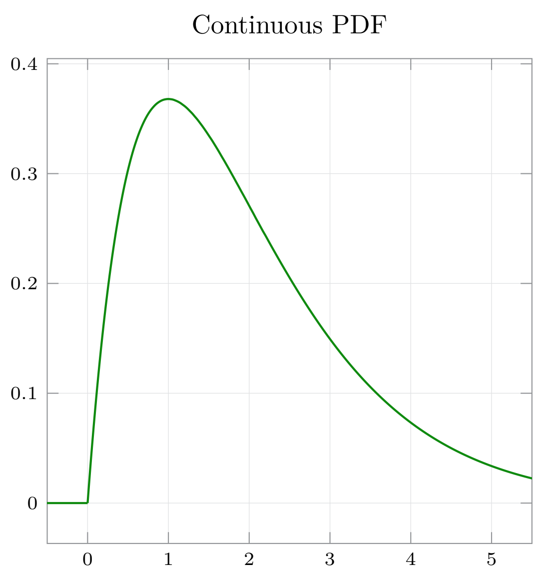
(e) The skewness is,

$$\mathbb{E}[X] = \int_{0}^{\infty} x^2 e^{-x} \; \mathrm{d}x = \Big[- e^{-x}(x^2 + 2x + 2) \Big]_0^{\infty} = 2 = \mu \tag{24.6.71}$$

$$\mathbb{E}[(X-\mu)^3] = \int_{0}^{\infty} (x-2)^3 \; (xe^{-x}) \; \mathrm{d}x = 4 \tag{24.6.72}$$

$$\mathbb{E}[(X-\mu)^2] = \int_{0}^{\infty} (x-2)^2 \; (xe^{-x}) \; \mathrm{d}x = 2 \tag{24.6.73}$$

$$\gamma = \frac{4}{2\sqrt{2}} = \sqrt{2} \tag{24.6.74}$$



(f) Coded in `sympy`. TBC.

(g) Let the function be discrete at just three points.

$$f(-5) = \frac{1}{6} \qquad f(-1) = a \tag{24.6.75}$$

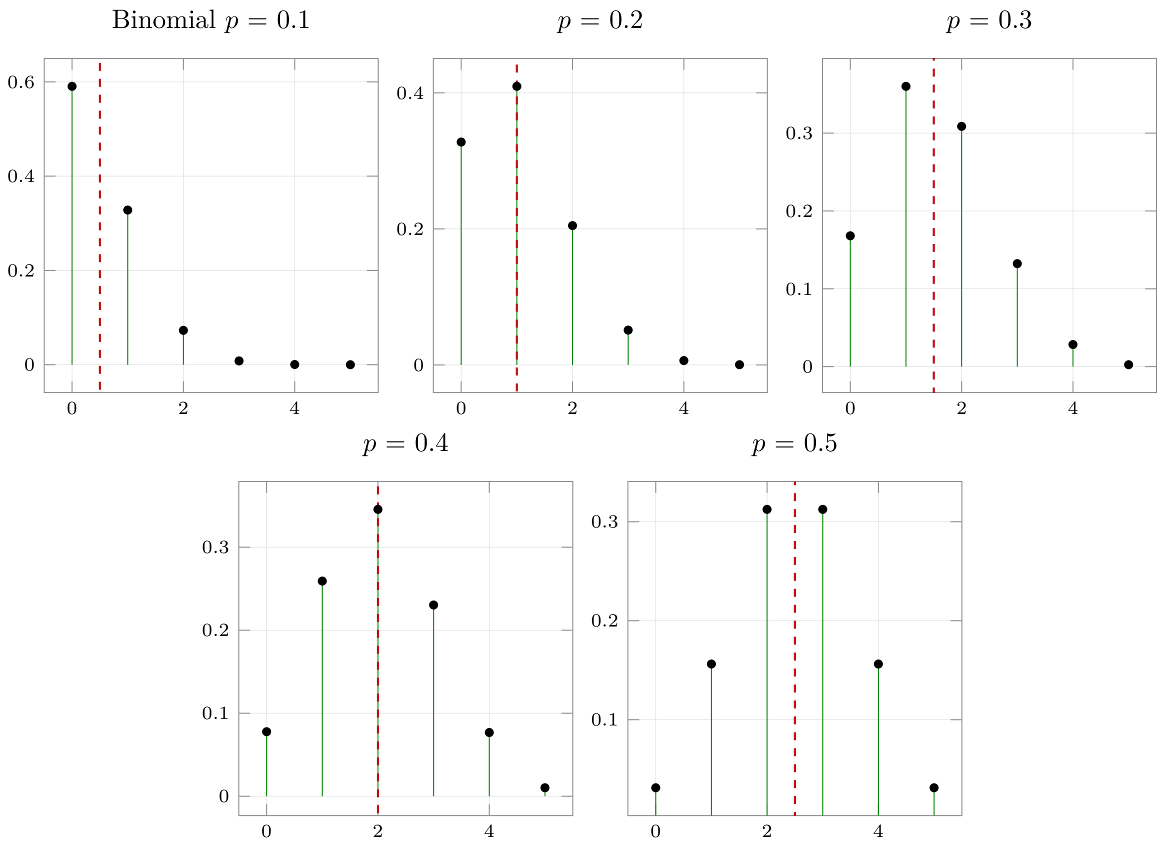
$$f(4) = b \tag{24.6.76}$$

$$a + b + \frac{1}{6} = 1 \qquad \frac{-5}{6} - a + 4b = 0 \tag{24.6.77}$$

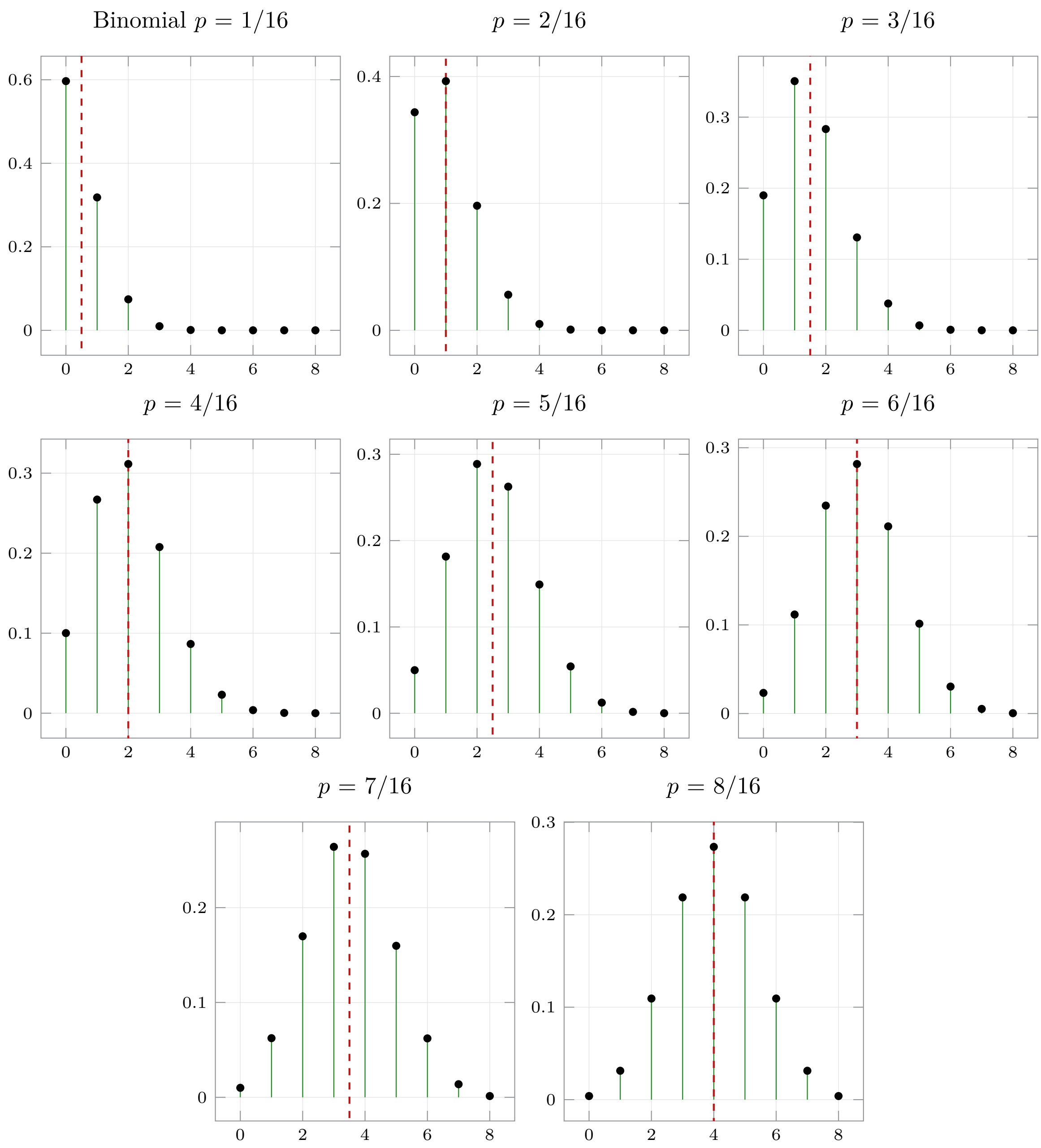
This gives the solution $a = 1/2$ and $b = 1/3$ which makes the PDF non-symmetric.

24.7 Binomial, Poisson, and Hypergeometric Distributions

1. Marking the positions of the mean, at $\mu = np$



2. Marking the positions of the mean, at $\mu = np$, with $n = 8$



3. Using the Poisson distribution with $\lambda = 5$

$$P(X \leq 5) = e^{-5}\left[\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!} + \frac{5^5}{5!}\right] \tag{24.7.1}$$

$$= \frac{1097}{12}\, e^{-5} = 61.6\,\% \tag{24.7.2}$$

$$P(X \geq 6) = 38.4\,\% \tag{24.7.3}$$

The probability is larger than in Example 3 since the value is closer to the mean.

4. Finding the probabilities, noting that the probabilities of success are

Number (x)	$f(x)$ with rep	$f(x)$ without rep
0	0.2401	0.2066
1	0.4116	0.45077
2	0.2646	0.28173
3	0.0756	0.05779
4	0.0081	0.0031

using the binomial distribution and hypergeometric distribution respectively.

5. For a fair coin, $p = q = 0.5$ and X being the number of heads,

$$f(x) = \binom{5}{x}(0.5)^x\ (0.5)^{5-x} \qquad f(x) = \frac{1}{32}\binom{5}{x} \tag{24.7.4}$$

$$P(X=0) = \frac{1}{32} \qquad P(X=1) = \frac{5}{32} \tag{24.7.5}$$

$$P(X \geq 1) = \frac{31}{32} \qquad P(X \leq 4) = \frac{31}{32} \tag{24.7.6}$$

$$\tag{24.7.7}$$

6. Finding the probabilities, noting that the probabilities of success are

$$\lambda = np = 100(0.04) = 4 \qquad f(x) = e^{-4}\frac{4^x}{x!} \tag{24.7.8}$$

Number (x)	$f(x)$
0	0.018
1	0.073
2	0.147
3	0.195
4	0.195
5	0.156

7. Finding the probabilities, noting that the probabilities of success are

$$\lambda = np = 100(0.04) = 4 \qquad f(x) = e^{-5}\frac{5^x}{x!} \tag{24.7.9}$$

$$P(X \leq 4) = e^{-5}\left[\frac{5^0}{0!} + \frac{5}{1!} + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!}\right] \qquad = 0.440 \tag{24.7.10}$$

8. Finding the probabilities, noting that the calls received per minute are Poisson distributed with $\lambda = 5$,

$$f(x) = e^{-5}\frac{5^x}{x!} \qquad P(X > 10) = 1 - P(X \leq 10) = 0.014 \tag{24.7.11}$$

9. Finding the probabilities, noting that the particles emitted per second are Poisson distributed with $\lambda = 0.5$,

$$f(x) = e^{-0.5}\frac{0.5^x}{x!} \qquad P(X \geq 2) = 1 - P(X < 2) = 0.09 \tag{24.7.12}$$

10. Since each bulb is independent, this is a binomial PDF with $n = 15$, $p = 0.02$,

$$P(X = 0) = \binom{15}{0} (0.02)^0 \, (0.98)^{15} = 0.739 \tag{24.7.13}$$

11. The new probability would be,

$$P(E_2) = (0.98)^1 00 = 0.1326 \tag{24.7.14}$$

12. The number of defects per 600 m is Poisson distributed with parameter $\lambda = 6$,

$$f(x) = e^{-6}\,\frac{6^x}{x!} \qquad f(0) = 0.0025 \tag{24.7.15}$$

13. This is a hypergeometric distribution with $N = 13$ and desirable draws $M = 3$.

$$f(x) = \frac{\binom{3}{x}\binom{10}{3-x}}{\binom{13}{3}} \tag{24.7.16}$$

$$\tag{24.7.17}$$

Number (x)	$f(x)$
0	0.4196
1	0.472
2	0.1049
3	0.0035

14. Finding the probabilities, noting that the customers per minute are Poisson distributed with $\lambda = 2$,

$$f(x) = e^{-2}\frac{2^x}{x!} \qquad P(X > 4) = 1 - P(X \leq 4) = 0.0527 \tag{24.7.18}$$

15. Finding the probabilities, noting that the number of defects per lot are Poisson distributed with $\lambda = 0.2$,

$$f(x) = e^{-0.2}\,\frac{0.2^x}{x!} \qquad P(X > 0) = 1 - P(X = 0) = 0.18127 \tag{24.7.19}$$

16. MGF

(a) Using differentiation under the integral sign,

$$\frac{\mathrm{d}^k G}{\mathrm{d}t^k} = \int_{-\infty}^{\infty} \frac{\mathrm{d}^k}{\mathrm{d}t^k}\Big[\exp(xt)\Big]\ f(x)\ \ \mathrm{d}x \qquad = \int_{-\infty}^{\infty} x^k\ e^{kt}\ f(x)\ \ \mathrm{d}x \tag{24.7.20}$$

$$\left.\frac{\mathrm{d}^k G}{\mathrm{d}t^k}\right|_{t=0} = \int_{-\infty}^{\infty} x^k\ \ f(x)\ \ \mathrm{d}x = \mathbb{E}[X^k] \qquad G'(0) = \mathbb{E}[X^1] = \mu \tag{24.7.21}$$

(b) Looking at $G^{(k)}(t=0)$,

$$G(t) = (pe^t + q)^n \qquad \frac{\mathrm{d}^k G}{\mathrm{d}t^k} = \sum_{x=0}^{n} x^k\ e^{tx}\ \binom{n}{x}\ p^x\ q^{n-x} \tag{24.7.22}$$

$$\frac{\mathrm{d}^k G}{\mathrm{d}t^k} = \sum_{x=0}^{n} x^k\ f_b(x) = \mathbb{E}[X^k] \tag{24.7.23}$$

Since the moments uniquely describe a PDF, this is the moment generating function of the binomial distribution $f_b(x)$.

(c) Using b,

$$\mu = \mathbb{E}[X] = \left.\frac{\mathrm{d}G}{\mathrm{d}t}\right|_{t=0} = \Big[n(pe^t + q)^{n-1}\ (p)\Big]_{t=0} \tag{24.7.24}$$

$$= np \tag{24.7.25}$$

(d) Using b,

$$\mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2] \qquad = \mathbb{E}[X^2] - 2\mu\ \mathbb{E}[X] + \mu^2\ \mathbb{E}[X^0] \tag{24.7.26}$$

$$= \mathbb{E}[X^2] - 2\mu^2 + \mu^2(1) \qquad = \mathbb{E}[X^2] - \mu^2 \tag{24.7.27}$$

$$\frac{\mathrm{d}G}{\mathrm{d}t} = (npe^t)\ (pe^t + q)^{n-1} \qquad \frac{\mathrm{d}^2 G}{\mathrm{d}t^2} = (np)(np - p + 1) \tag{24.7.28}$$

Subsittuting into the expression for the variance,

$$\sigma^2 = np\ (np - p + 1 - np) = npq \tag{24.7.29}$$

(e) From the definition of the MGF for the discrete case,

$$G(t) = \sum_{j=0}^{\infty} e^{tx_j}\ e^{-\mu}\ \frac{\mu^{x_j}}{x_j!} = e^{-\mu}\ \sum_{j} \frac{(\mu e^t)^{x_j}}{x_j!} = e^{-\mu}\ \ \exp(\mu e^t) \tag{24.7.30}$$

$$\mathbb{E}[X] = G'(0) = e^{-\mu}\ \mu e^t\ \ \exp(\mu e^t)\Big|_{t=0} = \mu \tag{24.7.31}$$

$$\mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - 2\mu\ \mathbb{E}[X] + \mu^2\ \mathbb{E}[1] \tag{24.7.32}$$

This requires the second moment,

$$G''(0) = e^{-\mu}\ (\mu e^t)(1 + \mu e^t)\ \ \exp(\mu e^t)\Big|_{t=0} = \mu(1 + \mu) \tag{24.7.33}$$

$$\sigma^2 = \mu(1 + \mu) + 2\mu^2 - \mu^2 = \mu \tag{24.7.34}$$

(f) Proving the relation,

$$x \binom{M}{x} = x \cdot \frac{M!}{x!\ (M - x)!} = \frac{M!}{(x - 1)!\ (M - x)!} = M \cdot \frac{(M - 1)!}{(x - 1)!\ (M - x)!} \tag{24.7.35}$$

$$= M \binom{M - 1}{x - 1} \tag{24.7.36}$$

(g) Since each of the outcomes A_j have to occur x_j times out of a total of n trials,

$$p(E) = p_1^{x_1} \cdot p_2^{x_2} \ \ldots\ p_j^{x_j} \tag{24.7.37}$$

Since this outcome is counted multiple times, the number of outcomes disregarding order is,

$$f(x_1, x_2, \ldots, x_j) = \frac{n!}{x_1!\ x_2!\ \ \ldots\ \ x_j!}\ p_1^{x_1} \cdot p_2^{x_2}\ \ldots\ p_j^{x_j} \tag{24.7.38}$$

The factorial expression is the number of ways of arranging n items, such that x_j is the number of items belonging to class j, (all of which are identical).

17. This is a binomial distribution with $p = 0.1$,

$$P(E) = P(X \geq 1) = 1 - P(X = 0) \qquad\qquad P(E) = 0.95 \tag{24.7.39}$$

$$0.05 = \binom{n}{0}\ p^0\ (1 - p)^n \qquad\qquad n = 28 \tag{24.7.40}$$

24.8 Normal Distribution

1. Converting to the standard normal,

$$Z = \frac{x - 10}{2} \tag{24.8.1}$$

$$P(X > 12) = P(Z > 1) = 0.1587 \tag{24.8.2}$$

$$P(X < 10) = P(Z < 0) = 0.5 \tag{24.8.3}$$

$$P(X < 11) = P(Z < 0.5) = 0.6915 \tag{24.8.4}$$

$$P(9 < X < 13) = P(-0.5 < Z < 1.5) = 0.62465 \tag{24.8.5}$$

2. Converting to the standard normal,

$$Z = \frac{x - 105}{5} \tag{24.8.6}$$

$$P(X \leq 112.5) = P(Z \leq 1.5) = 0.93319 \tag{24.8.7}$$

$$P(X > 100) = P(Z > -1) = 0.84134 \tag{24.8.8}$$

$$P(110.5 < X < 111.25) = P(1.1 < Z < 1.25) = 0.03002 \tag{24.8.9}$$

3. Converting to the standard normal,

$$Z = \frac{x - 50}{3} \tag{24.8.10}$$

$$P(X < c_1) = P(Z < c^*) = 0.05 \qquad c^* = \frac{c_1 - 50}{3} = -1.64485 \tag{24.8.11}$$

$$c_1 = 45.065 \tag{24.8.12}$$

$$P(X > c_2) = P(Z > c^*) = 0.01 \qquad c^* = \frac{c_1 - 50}{3} = 2.32635 \tag{24.8.13}$$

$$c_2 = 56.979 \tag{24.8.14}$$

$$P(50 - c_3 < X < 50 + c_3) = P(-c^* < Z < c^*) = 0.5 \qquad c^* = \frac{c_3}{3} = 0.67449 \tag{24.8.15}$$

$$c_3 = 2.0235 \tag{24.8.16}$$

4. Converting to the standard normal,

$$Z = \frac{x - 3.6}{0.1} \tag{24.8.17}$$

$$P(X \leq c_1) = P(Z \leq c^*) = 0.5 \qquad c^* = \frac{c_1 - 3.6}{0.1} = 0 \tag{24.8.18}$$

$$c_1 = 3.6 \tag{24.8.19}$$

$$P(X > c_2) = P(Z > c^*) = 0.1 \qquad c^* = \frac{c_1 - 3.6}{0.1} = 1.28155 \tag{24.8.20}$$

$$c_2 = 3.728 \tag{24.8.21}$$

$$P(-c_3 < X - 3.6 < c_3) = P(-c^* < Z < c^*) = 0.999 \qquad c^* = \frac{c_3}{0.1} = 3.29053 \tag{24.8.22}$$

$$c_3 = 0.3291 \tag{24.8.23}$$

5. Using the lookup tables for the standard normal,

$$P(X < 4) = P\left(\frac{X - 5}{1} < -1\right) = P(Z < -1) = 0.15866 \tag{24.8.24}$$

6. If $\sigma \downarrow$, then the limit $c \uparrow$ in $P(Z < c)$, which makes the probability $\downarrow$.

7. Using the lookup tables for the standard normal,

$$Z = \frac{X - 150}{5} \tag{24.8.25}$$

$$P(148 < X < 152) = P\left(- 0.4 < Z < 0.4\right) = 0.31084 \tag{24.8.26}$$

$$P(140 < X < 160) = P\left(- 2 < Z < 2\right) = 0.9545 \tag{24.8.27}$$

8. Using the lookup tables for the standard normal,

$$Z = \frac{X - 1500}{50} \tag{24.8.28}$$

$$P(X > c) = P(Z > c^*) = 0.95 \tag{24.8.29}$$

$$c^* = -1.64495 = \frac{c - 1500}{50} \tag{24.8.30}$$

$$c = 1418 \tag{24.8.31}$$

9. Using the lookup tables for the standard normal,

$$P(X < 500) = P\left(\frac{X - 480}{100} < 0.2\right) = P(Z < 0.2) = 0.5793 \tag{24.8.32}$$

10. A binomial PDF with $p = 0.01$ and $n = 1000$ is approximated by a normal PDF with,

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{X - 10}{\sqrt{9.9}} P(X < 10) \qquad = P(Z < 0) = 0.5 \tag{24.8.33}$$

A guess is $p = 0.5$ since this is the probability that an RV is less than its mean value.

11. Using the lookup tables for the standard normal,

$$P(X > t) = P(Z > t^*) = 0.2 \qquad t^* = 0.84162 \tag{24.8.34}$$

$$t^* \frac{t - 1000}{100} = 0.84162 \qquad t = 1084.16 \tag{24.8.35}$$

12. Using the lookup tables for the standard normal,

$$P(X > 15000) = P\left(\frac{X - 12000}{2000} > 1.5\right) = P(Z > 1.5) = 0.06681 \tag{24.8.36}$$

13. Using the lookup tables for the standard normal,

$$P(0.009 < X < 0.011) = P(-1.0 < Z < 1) = 0.68268 \tag{24.8.37}$$

$$np = 1000(68.268) = 683 \tag{24.8.38}$$

14. Normal distribution

(a) Using the loopkup table,

$$P(\mu - k\sigma < X \leq \mu + k\sigma) = P\left(-k < \frac{X - \mu}{\sigma} < k\right) \tag{24.8.39}$$

$$= P(-k < Z < k) = \Phi(k) - \Phi(-k) \tag{24.8.40}$$

k	$\Phi(-k)$	$\Phi(k)$	$P(-k < Z < k)$
1	0.15866	0.84134	68.27 %
2	0.02275	0.97725	95.45 %
3	0.00135	0.99865	99.73 %

$$P(\mu - k\sigma < X \leq \mu + k\sigma) = P\left(-k < \frac{X-\mu}{\sigma} < k\right) \tag{24.8.41}$$

$$= P(-k < Z < k) = \Phi(k) - \Phi(-k) \tag{24.8.42}$$

$$= 2\Phi(k) - 1 \tag{24.8.43}$$

$P(-k < Z < k)$	k
0.95	1.95996
0.99	2.57583
0.999	3.29053

(b) Proving the relation, using the symmetry of $f(x)$ about the y axis, which makes $f(-x) = f(x)$,

$$\Phi(-z) = \int_{-\infty}^{-z} f(x)\ \mathrm{d}x = \int_{\infty}^{z} f(-x)\ (-1)\ \mathrm{d}x = \int_{z}^{\infty} f(x)\ \mathrm{d}x \tag{24.8.44}$$

$$= \int_{-\infty}^{\infty} f(x)\ \mathrm{d}x - \int_{-\infty}^{z} f(x)\ \mathrm{d}x \tag{24.8.45}$$

$$= 1 - \Phi(z) \tag{24.8.46}$$

(c) Equating the second derivative to zero,

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{-1}{\sigma^2\sqrt{2\pi}}\ e^{-u^2/2}\ \left(\frac{x-\mu}{\sigma}\right) \tag{24.8.47}$$

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} = \frac{-1}{\sigma^3\sqrt{2\pi}}\ e^{-u^2/2} + \frac{1}{\sigma^3\sqrt{2\pi}}\ e^{-u^2/2}\ \left(\frac{x-\mu}{\sigma}\right)^2 = 0 \tag{24.8.48}$$

$$1 = \left(\frac{x-\mu}{\sigma}\right)^2 \tag{24.8.49}$$

$$x^* = \mu \pm \sigma \tag{24.8.50}$$

(d) Using polar coordinates,

$$\Phi^2(\infty) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2+y^2}{2}\right) \ \mathrm{d}x \ \mathrm{d}y \tag{24.8.51}$$

$$r^2 = x^2 + y^2 \tag{24.8.52}$$

$$\mathrm{d}x \ \mathrm{d}y = \mathrm{d}r \ (r) \ \mathrm{d}\theta \tag{24.8.53}$$

$$I = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} re^{-r^2/2} \ \mathrm{d}r \ \mathrm{d}\theta \tag{24.8.54}$$

$$= \int_0^{\infty} re^{-r^2/2} \ \mathrm{d}r = \left[-e^{-r^2/2} \right]_0^{\infty} = 1 \tag{24.8.55}$$

$$\Phi(\infty) = \sqrt{1} = 1 \tag{24.8.56}$$

(e) Using the definition of the variance,

$$\mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (v-\mu)^2 \ f(v) \ \mathrm{d}v \tag{24.8.57}$$

$$z = \frac{v-\mu}{\sigma} \tag{24.8.58}$$

$$\mathbb{E}[(X-\mu)^2] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 z^2 e^{-z^2/2} \ \sigma \ \mathrm{d}z \tag{24.8.59}$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} \ \mathrm{d}z = \frac{\sigma^2}{\sqrt{2\pi}} \left[\sqrt{\pi/2}\,\mathrm{erf}(z/\sqrt{2}) - xe^{-z^2/2} \right]_{-\infty}^{\infty} \tag{24.8.60}$$

$$= \sqrt{2\pi} \cdot @\frac{\sigma^2}{\sqrt{2\pi}} = \sigma^2 \tag{24.8.61}$$

The variance is indeed σ^2, which means that the term σ appearing in the formula is in fact that standard deviation.

(f) Proving the relation,

$$\mathrm{Var}[X_i] = pq \qquad \mathbb{E}[X_i] = p \tag{24.8.62}$$

$$\mathrm{Var}[\bar{X}] = \frac{(pq)^2}{n} \qquad \mathbb{E}[\bar{X}] = p \tag{24.8.63}$$

$$P\left(\left|\bar{X}-p\right| \geq \epsilon\right) \leq \frac{(pq)^2}{n\epsilon^2} \tag{24.8.64}$$

Since this probability goes to zero as $n \to \infty$, the complement probability approaches 1. This uses Chebyshev's inequality.

(g) Using the transformed variable,

$$Y = g(x) = c_1 X + c_2 \tag{24.8.65}$$

$$\mathbb{E}[Y] = \mathbb{E}[g(x)] = c_1\ \mathbb{E}[X] + c_2\ \mathbb{E}[1] = c_1 \mu + c_2 \tag{24.8.66}$$

$$\mathbb{E}[(Y - c_1\mu - c_2)^2] = \mathbb{E}[(c_1 X - c_1 \mu)^2] = c_1^2\ \ \mathbb{E}[(X - \mu)^2] \tag{24.8.67}$$

$$= c_1^2\ \sigma^2 \tag{24.8.68}$$

15. Since the tables only show the CDF of the function, they are used as,

$$P(Z < b) = \Phi(b) \qquad\qquad P(Z > a) = 1 - \Phi(a) \tag{24.8.69}$$

$$P(a < Z < b) = \Phi(b) - \Phi(a) \tag{24.8.70}$$

The inverse normal CDF lookup tables work similarly.

24.9 Distributions of Several Random Variables

1. Using the normalization condition,

$$1 = \int_0^2 \int_8^{12} k\ \ \mathrm{d}x\ \ \mathrm{d}y = 8k \qquad\qquad k = \frac{1}{8} \tag{24.9.1}$$

$$P(E_1) = \int_1^{1.5} \int_8^{11} (1/8)\ \ \mathrm{d}x\ \ \mathrm{d}y \qquad\qquad = \frac{3}{16} \tag{24.9.2}$$

$$P(E_2) = \int_0^{1} \int_9^{12} (1/8)\ \ \mathrm{d}x\ \ \mathrm{d}y \qquad\qquad = \frac{3}{8} \tag{24.9.3}$$

2. Since the PDF is defined nonzero over the region

$$x \geq 0, \quad y \geq 0 \qquad\qquad x + y \leq 8 \tag{24.9.4}$$

$$P(E_1) = \int_4^{\infty} \int_4^{\infty} f(x, y)\ \ \mathrm{d}x\ \ \mathrm{d}y \qquad\qquad = 0 \tag{24.9.5}$$

$$P(E_2) = \int_0^{1} \int_0^{1} (1/32)\ \ \mathrm{d}x\ \ \mathrm{d}y \qquad\qquad = \frac{1}{32} \tag{24.9.6}$$

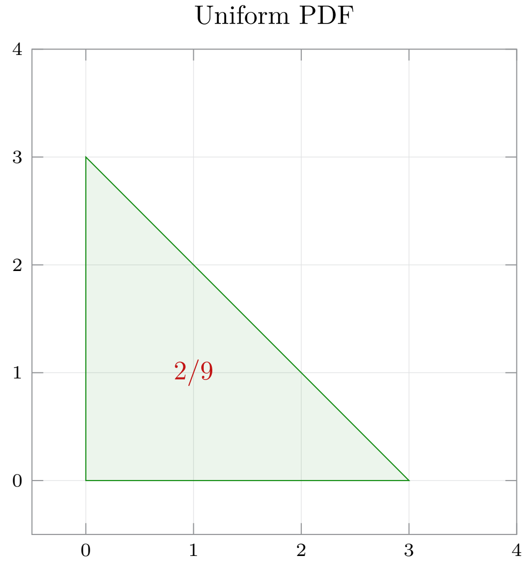
3. Using the normalization condition,

$$1 = \int_0^3 \int_0^{3-y} k \ \mathrm{d}x \ \mathrm{d}y \qquad = k \left[3y - \frac{y^2}{2}\right]_0^3 \tag{24.9.7}$$

$$k = \frac{2}{9} \tag{24.9.8}$$

$$P(E_1) = \int_0^1 \int_0^{1-y} (2/9) \ \mathrm{d}x \ \mathrm{d}y \qquad = \frac{2}{9}\left[y - \frac{y^2}{2}\right]_0^1 = \frac{1}{9} \tag{24.9.9}$$

$$P(E_2) = \int_0^{1.5} \int_x^{3-x} (2/9) \ \mathrm{d}y \ \mathrm{d}x \qquad = \frac{2}{9}\left[3x - x^2\right]_0^{1.5} = \frac{1}{2} \tag{24.9.10}$$



4. The marginal PDF of x is,

$$f_1(x) = \int_0^{8-x} \frac{1}{32} \ \mathrm{d}y = \frac{8-x}{32} \qquad f_1(x) = \begin{cases} \dfrac{8-x}{32} & x \in [0, 8] \\ 0 & \text{otherwise} \end{cases} \tag{24.9.11}$$

5. The marginal PDF of x is,

$$f_2(y) = \int_{\alpha_1}^{\beta_1} \frac{1}{k} \ \mathrm{d}x = \frac{1}{(\beta_2 - \alpha_2)} \tag{24.9.12}$$

$$f_2(y) = \begin{cases} \dfrac{1}{(\beta_2 - \alpha_2)} & y \in [\alpha_2, \beta_2] \\ 0 & \text{otherwise} \end{cases} \tag{24.9.13}$$

6. Using the sum of variances of I.I.D. RVs

$$\mu = 10 \text{ g} \qquad \sigma = 0.05 \text{ g} \tag{24.9.14}$$

$$n = 10000 \qquad Y = nX \tag{24.9.15}$$

$$\mu_Y = n\mu = 100 \text{ kg} \qquad \sigma_Y = \sqrt{n}\sigma = 5 \text{ g} \tag{24.9.16}$$

7. Using the sum of variances of I.I.D. RVs

$$n_1 = 50, \quad n_2 = 49 \qquad Z = n_1 X + n_2 Y \tag{24.9.17}$$

$$\mu_Z = n_1 \mu_X + n_2 \mu_Y = 27.45 \text{ mm} \tag{24.9.18}$$

$$\sigma_Z^2 = n_1 \sigma_X^2 + n_2 \sigma_Y^2 = 0.1446 \text{ mm}^2 \qquad \sigma_Z = 0.38 \text{ mm} \tag{24.9.19}$$

8. The marginal PDFs are,

$$f_1(x) = \int_1^{1.04} f(x, y) \ \text{d}y \qquad = 25 \quad \forall \quad x \in [0.98, 1.02] \tag{24.9.20}$$

$$f_2(y) = \int_{0.98}^{1.02} f(x, y) \ \text{d}x \qquad = 25 \quad \forall \quad y \in [1, 1.04] \tag{24.9.21}$$

$$P(X < 1) = \int_{0.98}^{1} f_1(x) \ \text{d}x = 0.5 \tag{24.9.22}$$

9. The binomial theorem is a set of n I.I.D. RVs, each of which has $\mu = p$ and $\sigma^2 = p(1 - p)$

$$\mathbb{E}[X_1 + X_2 + \cdots + X_n] = np \tag{24.9.23}$$

$$\text{Var}[X_1 + X_2 + \cdots + X_n] = np(1 - p) \tag{24.9.24}$$

These match the mean and variance of a binomial distribution.

10. Consider the expected value of the second hypergeometric trial,

$$\mathbb{E}[X_2] = \frac{M}{N} \cdot \frac{M-1}{N-1} + \frac{N-M}{N} \cdot \frac{M}{N-1} = \frac{MN - M}{N(N-1)} = \frac{M}{N} \tag{24.9.25}$$

This means that the second trial is identically distributed to the first Trial in spite of these trials not being independent.

Using the addition of means,

$$\mathbb{E}[X_1 + X_2 + \cdots + X_n] = n \cdot \frac{M}{N} \tag{24.9.26}$$

The addition of variances cannot be used since the trials are not independent.

11. Using the sum of variances of I.I.D. RVs

$$n_1 = 5, \quad n_2 = 4 \qquad\qquad Z = n_1 X + n_2 Y \tag{24.9.27}$$

$$\mu_Z = n_1 \mu_X + n_2 \mu_Y = 25.26 \text{ cm} \tag{24.9.28}$$

$$\sigma_Z^2 = n_1 \sigma_X^2 + n_2 \sigma_Y^2 = 6.1 \times 10^{-5} \text{ cm}^2 \qquad\qquad \sigma_Z = 0.0078 \text{ cm} \tag{24.9.29}$$

12. Using the sum of variances of I.I.D. RVs

$$Z = X + Y \tag{24.9.30}$$

$$\mu_Z = \mu_X + \mu_Y = 105 \tag{24.9.31}$$

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 = 0.29 \qquad\qquad \sigma_Z = 0.5385 \tag{24.9.32}$$

13. The probability is,

$$P(X > Y) = \int_0^\infty \int_y^\infty 0.25\, e^{-0.5(x+y)} \, dx \, dy \tag{24.9.33}$$

$$= \int_0^\infty 0.25 e^{-0.5y} \left[\int_y^\infty e^{-0.5x} \, dx \right] dy \tag{24.9.34}$$

$$I_1 = \left[-2e^{-0.5x} \right]_y^\infty = 2e^{-0.5y} \tag{24.9.35}$$

$$I_2 = 0.5 \int_0^\infty e^{-y} = 0.5 \tag{24.9.36}$$

Heuristically, the PDF is symmetric about the line $y = x$, which means the probability is half.

14. Solving,

(a) They are independent since the joint PDF is a product of individual PDFs $f_1(x)$ and $f_2(y)$.

(b) The marginal distributions are,

$$f_1(x) = \int_0^\infty f(x, y) \, dy = 4e^{-2x} \left[\frac{e^{-2y}}{-2} \right]_0^\infty = 2e^{-2x} \tag{24.9.37}$$

$$f_2(y) = 2e^{-2y} \tag{24.9.38}$$

using a similar calculation.

(c) Using the marginal PDFs,

$$P(X > 2) = \int_2^\infty f_1(x) \, dx = \left[-e^{-2x} \right]_2^\infty = e^{-4} = 0.0183 \tag{24.9.39}$$

15. Let $f(x, y)$ and $g(x, y)$ be two PDFs,

$$f_1(x) = \begin{cases} \frac{1}{2\pi} & x \in [0, 2\pi] \\ 0 & \text{otherwise} \end{cases} \qquad g_1(x) = \begin{cases} \frac{1}{2\pi} + \sin(x) & x \in [0, 2\pi] \\ 0 & \text{otherwise} \end{cases} \tag{24.9.40}$$

Clearly, the marginal PDF in y is the same.

$$f_2(y) = \int_0^{2\pi} f_1(x)\ f_2(y)\ \ \mathrm{d}x = g_2(y) \tag{24.9.41}$$

By a similar process, the other marginal PDFs can also be made equal.

16. Deriving, using the four areas as defined,

$$Z_3 : x < a_1, \qquad y \in [a_2, b_2] \qquad\qquad Z_4 : x \in [a_1, b_1], \qquad y < b_1 \tag{24.9.42}$$

$$Z_2 : x < a_1, \qquad y < b_1 \qquad\qquad Z_4 : x \in [a_1, b_1], \qquad y \in [a_2, b_2] \tag{24.9.43}$$

$$F(a_1, a_2) = Z_2 \qquad\qquad F(b_1, b_2) = Z_1 + Z_2 + Z_3 + Z_4 \tag{24.9.44}$$

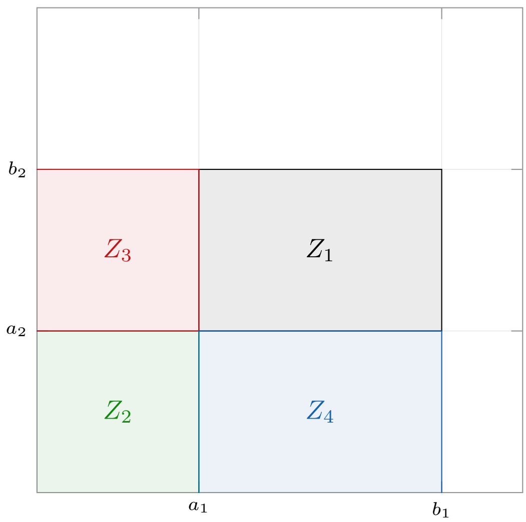
$$F(b_1, a_2) = Z_2 + Z_4 \qquad\qquad F(a_1, b_2) = Z_2 + Z_4 \tag{24.9.45}$$

Clearly, there is a double subtraction of Z_2, which means

$$Z_1 = (Z_1 + Z_2 + Z_3 + Z_4) - (Z_2 + Z_3) - (Z_2 + Z_4) + Z_2 \tag{24.9.46}$$

$$= F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \tag{24.9.47}$$

Uniform PDF



17. Using the marginal distribution,

$$f_1(x) = \begin{cases} 1/2 & x = 0 \\ 1/2 & x = 1 \end{cases} \qquad f_2(y) = \begin{cases} 1/2 & y = 0 \\ 1/2 & y = 1 \end{cases} \tag{24.9.48}$$

$$f_1(0) \cdot f_2(0) = \frac{1}{4} \neq f(0, 0) \tag{24.9.49}$$

This means the RVs are not independent.

18. Using the normalization condition,

$$1 = \int_0^{2\pi} \int_0^1 k \ \mathrm{d}r \ (r) \ \mathrm{d}\theta \qquad 1 = 2k\pi \left[\frac{r^2}{2}\right]_0^1 \tag{24.9.50}$$

$$k = \frac{1}{\pi} \tag{24.9.51}$$

$$P(r < 1/2) = 2\pi \int_0^{1/2} \frac{r}{\pi} \ \mathrm{d}r \qquad P(r < 1/2) = \frac{1}{4} \tag{24.9.52}$$

This is the fraction of the area of a circle contained within the inner half radius.

19. Finding the marginal PDF,

$$f_1(x) = \int_0^1 f(x, y) \ \mathrm{d}y \qquad = \left[xy + \frac{y^2}{2}\right]_0^1 = x + \frac{1}{2} \tag{24.9.53}$$

$$f_2(y) = \int_0^1 f(x, y) \ \mathrm{d}x \qquad = \left[xy + \frac{x^2}{2}\right]_0^1 = y + \frac{1}{2} \tag{24.9.54}$$

$$g_1(x) = \int_0^1 g(x, y) \ \mathrm{d}y \qquad = (x + 1/2) \left[\frac{y + y^2}{2}\right]_0^1 = x + \frac{1}{2} \tag{24.9.55}$$

$$g_2(y) = \int_0^1 g(x, y) \ \mathrm{d}x \qquad = (y + 1/2) \left[\frac{x + x^2}{2}\right]_0^1 = y + \frac{1}{2} \tag{24.9.56}$$

Both $f(x, y)$ and $g(x, y)$ have the same marginal PDFs.

20. Let the condition hold,

$$f(x,\, y) = f_1(x) \cdot f_2(y) \tag{24.9.57}$$

$$F(x,\, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_1(u) \cdot f_2(v) \;\; \mathrm{d}u \;\; \mathrm{d}v \tag{24.9.58}$$

$$= \int_{-\infty}^{x} f_1(u) \left[\int_{-\infty}^{y} f_2(v) \;\; \mathrm{d}y \right] \; \mathrm{d}x \qquad\qquad = F_2(y) \; \int_{-\infty}^{x} f_1(u) \;\; \mathrm{d}x \tag{24.9.59}$$

$$= F_2(y) \cdot F_1(x) \tag{24.9.60}$$

This means the condition is sufficient.
For the inverse, assume the RVs are independent,

$$F(x,\, y) = F_1(x) \cdot F_2(y) \tag{24.9.61}$$

$$F(x,\, y) = \left[\int_{-\infty}^{x} f_1(u) \;\; \mathrm{d}u \right] \left[\int_{-\infty}^{y} f_2(v) \;\; \mathrm{d}y \right] \tag{24.9.62}$$

$$= \int_{-\infty}^{x} \int_{-\infty}^{y} f_1(u) \cdot f_2(v) \;\; \mathrm{d}u \;\; \mathrm{d}v \tag{24.9.63}$$

$$f(x,\, y) = f_1(x) \cdot f_2(y) \tag{24.9.64}$$

This means the condition is necessary.