Review for final exam solutions: probability unit
MIT 18.05 Spring 2022

Problem 1. *Consider the nucleotides A, G, C, T.*

(a) *How many ways are there to make a sequence of 5 nucleotides.*

Solution: Four ways to fill each slot: $\boxed{4^5}$.

(b) *How many sequences of length 5 are there where no adjacent nucleotides are the same*

Solution: Four ways to fill the first slot and 3 ways to fill each subsequsent slot: $\boxed{4 \cdot 3^4}$.

(c) *How many sequences of length 5 have exactly one A?*

Solution: Build the sequences as follows:

Step 1: Choose which of the 5 slots gets the A: 5 ways to place the one A.

Step 2: 3^4 ways to fill the remain 4 slots.

By the rule of product there are $\boxed{5 \cdot 3^4}$ such sequences.

Problem 2. ***(a)*** *How many 5 card poker hands are there?*

(b) *How many ways are there to get a full house (3 of one rank and 2 of another)?*

(c) *What's the probability of getting a full house?*

(a) Solution: $\binom{52}{5}$.

(b) Solution: Number of ways to get a full-house: $\binom{4}{2}\binom{13}{1}\binom{4}{3}\binom{12}{1}$

(c) Solution: $\dfrac{\binom{4}{2}\binom{13}{1}\binom{4}{3}\binom{12}{1}}{\binom{52}{5}}$

Problem 3. (Counting)

(a) *How many arrangements of the letters in the word probability are there?*

(b) *Suppose all of these arrangements are written in a list and one is chosen at random. What is the probability it begins with 'b'.*

(a) Solution: There are several ways to think about this. Here is one.

The 11 letters are p, r, o, b,b, a, i,i, l, t, y. We use the following steps to create a sequence of these letters.

Step 1: Choose a position for the letter p: 11 ways to do this.
Step 2: Choose a position for the letter r: 10 ways to do this.
Step 3: Choose a position for the letter o: 9 ways to do this.
Step 4: Choose two positions for the two b's: 8 choose 2 ways to do this.
Step 5: Choose a position for the letter a: 6 ways to do this.
Step 6: Choose two positions for the two i's: 5 choose 2 ways to do this.
Step 7: Choose a position for the letter l: 3 ways to do this.

Step 8: Choose a position for the letter t: 2 ways to do this.
Step 9: Choose a position for the letter y: 1 ways to do this.

Multiply these all together we get:

$$11 \cdot 10 \cdot 9 \cdot \binom{8}{2} \cdot 6 \cdot \binom{5}{2} \cdot 3 \cdot 2 \cdot 1 \; = \; \frac{11!}{2! \cdot 2!}$$

(b) Solution: Here are two ways to do this problem.

Method 1. Since every arrangement has equal probability of being chosen we simply have to count the number that start with the letter 'b'. After putting a 'b' in position 1 there are 10 letters: p, r, o, b, a, i,i, l, t, y, to place in the last 10 positions. We count this in the same manner as part (a). That is
Choose the position for p: 10 ways.
Choose the positions for r,o,b,a,: $9 \cdot 8 \cdot 7 \cdot 6$ ways.
Choose two positions for the two i's: 5 choose 2 ways.
Choose the position for l: 3 ways.
Choose the position for t: 2 ways.
Choose the position for y: 1 ways.

Multiplying this together we get $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot \binom{5}{2} \cdot 3 \cdot 2 \cdot 1 = \frac{10!}{2!}$ arrangements start with the letter b. Therefore the probability a random arrangement starts with b is $\boxed{\frac{10!/2!}{11!/2! \cdot 2!} = \frac{2}{11}}$

Method 2. Suppose we build the arrangement by picking a letter for the first position, then the second position etc. Since there are 11 letters, two of which are b's we have a 2/11 chance of picking a b for the first letter.

Problem 4. *Let E and F be two events. Suppose the probability that at least one of them occurs is 2/3. What is the probability that neither E nor F occurs?*

Solution: We are given $P(E \cup F) = 2/3$ and asked to find $P((E \cup F)^c)$.

$P((E \cup F)^c) = 1 - P(E \cup F) = \boxed{1/3.}$

Problem 5. *Let C and D be two events with $P(C) = 0.3$, $P(D) = 0.4$, and $P(C^c \cap D) = 0.2$. What is $P(C \cap D)$?*

Solution: D is the disjoint union of $D \cap C$ and $D \cap C^c$.

So, $P(D \cap C) + P(D \cap C^c) = P(D)$. Thus, $P(D \cap C) = P(D) - P(D \cap C^c) = 0.4 - 0.2 = \boxed{0.2.}$

Problem 6. *Suppose we have 8 teams labeled T_1, ..., T_8. Suppose they are ordered by placing their names in a hat and drawing the names out one at a time.*

(a) *How many ways can it happen that all the odd numbered teams are in the odd numbered slots and all the even numbered teams are in the even numbered slots?*

Solution: Slots 1, 3, 5, 7 are filled by T_1, T_3, T_5, T_7 in any order: 4! ways.

Slots 2, 4, 6, 8 are filled by T_2, T_4, T_6, T_8 in any order: 4! ways.

Solution: $4! \cdot 4! = 576$.

(b) *What is the probability of this happening?*

Solution: There are $8!$ ways to fill the 8 slots in any way.

Since each outcome is equally likely the probabilitiy is $\dfrac{4! \cdot 4!}{8!} = \dfrac{576}{40320} = 0.143 = 1.43\%$.

Problem 7. *More cards! Suppose you want to divide a 52 card deck into four hands with 13 cards each. What is the probability that each hand has a king?*

Solution: Let H_i be the event that the i^{th} hand has one king. We have the conditional probabilities

$$P(H_1) = \frac{\binom{4}{1}\binom{48}{12}}{\binom{52}{13}}; \quad P(H_2|H_1) = \frac{\binom{3}{1}\binom{36}{12}}{\binom{39}{13}}; \quad P(H_3|H_1 \cap H_2) = \frac{\binom{2}{1}\binom{24}{12}}{\binom{26}{13}}$$

$P(H_4|H_1 \cap H_2 \cap H_3) = 1$

$$\begin{aligned} P(H_1 \cap H_2 \cap H_3 \cap H_4) &= P(H_4|H_1 \cap H_2 \cap H_3)\, P(H_3|H_1 \cap H_2)\, P(H_2|H_1)\, P(H_1) \\ &= \frac{\binom{2}{1}\binom{24}{12}\binom{3}{1}\binom{36}{12}\binom{4}{1}\binom{48}{12}}{\binom{26}{13}\binom{39}{13}\binom{52}{13}}. \end{aligned}$$

Problem 8. *Suppose we roll a fair die twice. Let A be the event 'the sum of the rolls is 5' and let B be the event 'at least one of the rolls is 4.'*

(a) *Calculate $P(A|B)$.*

(b) *Are A and B independent?*

(a) Solution: Sample space $= S = \{(1,1),\ (1,2),\ (1,3),\ \dots,\ (6,6)\} = \{(i,j) \mid i,j = 1,\ 2,\ 3,\ 4,\ 5,\ 6\}$.

(Each outcome is equally likely, with probability 1/36.)

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

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

$P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{2/36}{11/36} = \dfrac{2}{11}$.

(b) Solution: $P(A) = 4/36 \neq P(A|B)$, so they are not independent.

Problem 9. *On a quiz show the contestant is given a multiple choice question with 4 options. Suppose there is a 70% chance the contestant actually knows the answer. If they don't know the answer they guess with a 25% chance of getting it right. Suppose they get it right. What is the probability that they were guessing?*

Solution: Let C be the event the contestant gets the question correct and G the event the contestant guessed.

The question asks for $P(G|C)$.

We'll compute this using Bayes' rule: $P(G|C) = \dfrac{P(C|G)\,P(G)}{P(C)}$.

We're given: $\quad P(C|G) = 0.25, \quad P(K) = 0.7.$

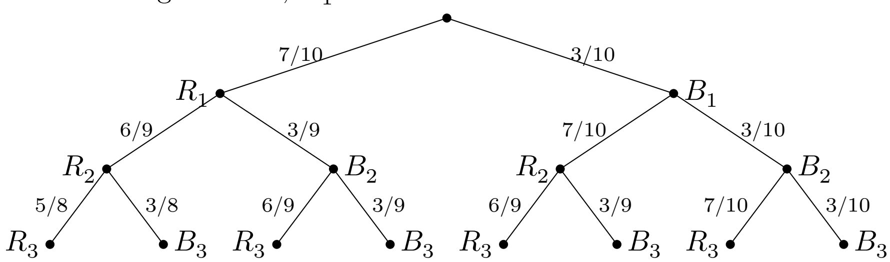
Law of total prob.: $P(C) = P(C|G)\,P(G) + P(C|G^c)\,P(G^c) = 0.25 \cdot 0.3 + 1.0 \cdot 0.7 = 0.775$

Therefore $P(G|C) = \dfrac{0.075}{0.775} = 0.097 = 9.7\%$.

Problem 10. *Suppose you have an urn containing 7 red and 3 blue balls. You draw three balls at random. On each draw, if the ball is red you set it aside and if the ball is blue you put it back in the urn. What is the probability that the third draw is blue?*

(If you get a blue ball it counts as a draw even though you put it back in the urn.)

Solution: Here is the game tree, R_1 means red on the first draw etc.



Summing the probability to all the B_3 nodes we get

$P(B_3) = \frac{7}{10} \cdot \frac{6}{9} \cdot \frac{3}{8} + \frac{7}{10} \cdot \frac{3}{9} \cdot \frac{3}{9} + \frac{3}{10} \cdot \frac{7}{10} \cdot \frac{3}{9} + \frac{3}{10} \cdot \frac{3}{10} \cdot \frac{3}{10} = 0.350.$

Problem 11. *Suppose that $P(A) = 0.4, P(B) = 0.3$ and $P((A \cup B)^C) = 0.42$. Are A and B independent?*

Solution: We have $P(A \cup B) = 1 - 0.42 = 0.58$ and we know because of the inclusion-exclusion principle that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Thus,

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.4 + 0.3 - 0.58 = 0.12 = (0.4)(0.3) = P(A)P(B).$$

So, A and B are independent.

Problem 12. *Suppose now that events A, B and C are mutually independent with*

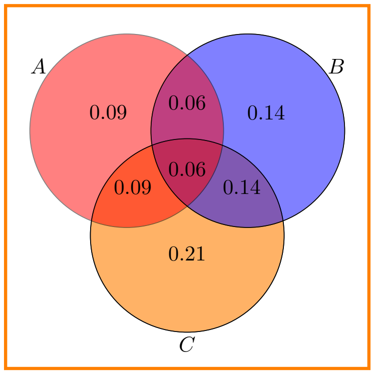
$$P(A) = 0.3, \quad P(B) = 0.4, \quad P(C) = 0.5.$$

Compute the following: (Hint: Use a Venn diagram)
(i) $P(A \cap B \cap C^c)$ $\quad$ (ii) $P(A \cap B^c \cap C)$ $\quad$ (iii) $P(A^c \cap B \cap C)$

Solution: By the mutual independence we have

$$P(A \cap B \cap C) = P(A)P(B)P(C) = 0.06 \qquad P(A \cap B) = P(A)P(B) = 0.12$$
$$P(A \cap C) = P(A)P(C) = 0.15 \qquad P(B \cap C) = P(B)P(C) = 0.2$$

We show this in the following Venn diagram



Note that, for instance, $P(A \cap B)$ is split into two pieces. One of the pieces is $P(A \cap B \cap C)$ which we know and the other we compute as $P(A \cap B) - P(A \cap B \cap C) = 0.12 - 0.06 = 0.06$. The other intersections are similar.

We can read off the asked for probabilities from the diagram.
(i) $P(A \cap B \cap C^c) = 0.06$
(ii) $P(A \cap B^c \cap C) = 0.09$
(iii) $P(A^c \cap B \cap C) = 0.14$.

Problem 13. *Suppose A and B are events with $0 < P(A) < 1$ and $0 < P(B) < 1$.*

(a) *If A and B are disjoint can they be independent?*

(b) *If A and B are independent can they be disjoint?*

(c) *If $A \subset B$ can they be independent?*

Solution: The answer to all three parts is 'No'. Each of these answers relies on the fact that the probabilities of A and B are strictly between 0 and 1.

To show A and B are not independent we need to show either $P(A \cap B) \neq P(A) \cdot P(B)$ or $P(A|B) \neq P(A)$.

(a) No, they cannot be independent: $A \cap B = \emptyset \;\Rightarrow\; P(A \cap B) = 0 \neq P(A) \cdot P(B)$.

(b) No, they cannot be disjoint: same reason as in part (a).

(c) No, they cannot be independent: $A \subset B \;\Rightarrow\; A \cap B = A$
$\Rightarrow\; P(A \cap B) = P(A) > P(A) \cdot P(B)$. The last inequality follows because $P(B) < 1$.

Problem 14. *Directly from the definitions of expected value and variance, compute $E[X]$ and $Var(X)$ when X has probability mass function given by the following table:*

X	*-2*	*-1*	*0*	*1*	*2*
pmf	*1/15*	*2/15*	*3/15*	*4/15*	*5/15*

Solution: We compute

$$E[X] = -2 \cdot \frac{1}{15} + -1 \cdot \frac{2}{15} + 0 \cdot \frac{3}{15} + 1 \cdot \frac{4}{15} + 2 \cdot \frac{5}{15} = \frac{2}{3}.$$

Thus

$$\begin{aligned}
\mathrm{Var}(X) &= E[(X - \frac{2}{3})^2] \\
&= \left(-2 - \frac{2}{3}\right)^2 \cdot \frac{1}{15} + \left(-1 - \frac{2}{3}\right)^2 \cdot \frac{2}{15} + \left(0 - \frac{2}{3}\right)^2 \cdot \frac{3}{15} + \left(1 - \frac{2}{3}\right)^2 \cdot \frac{4}{15} + \left(2 - \frac{2}{3}\right)^2 \cdot \frac{5}{15} \\
&= \frac{14}{9}.
\end{aligned}$$

Problem 15. *Suppose that X takes values between 0 and 1 and has probability density function $2x$. Compute $\mathrm{Var}(X)$ and $\mathrm{Var}(X^2)$.*

Solution: We will make use of the formula $\mathrm{Var}(Y) = E[Y^2] - E[Y]^2$. First we compute

$$E[X] = \int_0^1 x \cdot 2x dx = \frac{2}{3}$$

$$E[X^2] = \int_0^1 x^2 \cdot 2x dx = \frac{1}{2}$$

$$E[X^4] = \int_0^1 x^4 \cdot 2x dx = \frac{1}{3}.$$

Thus,

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

and

$$\mathrm{Var}(X^2) = E[X^4] - (E[X^2])^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Problem 16. *The pmf of X is given by the following table*

Value of X	*-1*	*0*	*1*
Probability	*1/3*	*1/6*	*1/2*

(a) *Compute $E[X]$.*

(b) *Give the pdf of $Y = X^2$ and use it to compute $E[Y]$.*

(c) *Instead, compute $E[X^2]$ directly from an extended table.*

(d) *Compute $\mathrm{Var}(X)$.*

(a) Solution: We have the extended table

X values:	-1	0	1
prob:	1/3	1/6	1/2
X^2	1	0	1

So, $\boxed{E[X] = -1/3 + 1/2 = 1/6.}$

(b) Solution:

Y values:	0	1
prob:	1/6	5/6

$\Rightarrow$ $\boxed{E[Y] = 5/6.}$

(c) Solution: Using the table in part (a) $E[X^2] = 1 \cdot (1/3) + 0 \cdot (1/6) + 1 \cdot (1/2) = 5/6$ (same as part (b)).

(d) Solution: $\boxed{\text{Var}(X) = E[X^2] - E[X]^2 = 5/6 - 1/36 = 29/36.}$

Problem 17. *Compute the expectation and variance of a Bernoulli(p) random variable.*

Solution: Make a table:

X:	0	1
prob:	(1-p)	p
X^2	0	1.

From the table, $\boxed{E[X] = 0 \cdot (1-p) + 1 \cdot p = p.}$

Since X and X^2 have the same table $E[X^2] = E[X] = p$.

Therefore, $\boxed{\text{Var}(X) = p - p^2 = p(1-p).}$

Problem 18. *Suppose 100 people all toss a hat into a box and then proceed to randomly pick out of a hat. What is the expected number of people to get their own hat back.*

Hint: express the number of people who get their own hat as a sum of random variables whose expected value is easy to compute.

Solution: Let X be the number of people who get their own hat.

Following the hint: let X_j represent whether person j gets their own hat. That is, $X_j = 1$ if person j gets their hat and 0 if not.

We have, $X = \sum_{j=1}^{100} X_j$, so $E[X] = \sum_{j=1}^{100} E[X_j]$.

Since person j is equally likely to get any hat, we have $P(X_j = 1) = 1/100$. Thus, $X_j \sim$ Bernoulli$(1/100)$ $\Rightarrow$ $E[X_j] = 1/100$ $\Rightarrow$ $\boxed{E[X] = 1.}$

Problem 19. *Suppose that $X \sim Bin(n, 0.5)$. Find the probability mass function of $Y = 2X$.*

Solution: For $y = 0, 2, 4, \ldots, 2n$,

$$P(Y = y) = P(X = \frac{y}{2}) = \binom{n}{y/2} \left(\frac{1}{2}\right)^n.$$

Problem 20. ***(a)** Suppose that X is uniform on $[0,1]$. Compute the pdf and cdf of X.*

***(b)** If $Y = 2X + 5$, compute the pdf and cdf of Y.*

(a) Solution: We have $f_X(x) = 1$ for $0 \le x \le 1$. The cdf of X is

$$F_X(x) = \int_0^x f_X(t)dt = \int_0^x 1dt = x.$$

(b) Solution: Since X is between 0 and 1 we have Y is between 5 and 7. Now for $5 \leq y \leq 7$, we have

$$F_Y(y) = P(Y \leq y) = P(2X + 5 \leq y) = P(X \leq \frac{y-5}{2}) = F_X(\frac{y-5}{2}) = \frac{y-5}{2}.$$

Differentiating $P(Y \leq y)$ with respect to y, we get the probability density function of Y, for $5 \leq y \leq 7$,

$$f_Y(y) = \frac{1}{2}.$$

Problem 21. ***(a)*** *Suppose that X has probability density function $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. Compute the cdf, $F_X(x)$.*

(b) *If $Y = X^2$, compute the pdf and cdf of Y.*

(a) Solution: We have cdf of X,

$$F_X(x) = \int_0^x \lambda \mathrm{e}^{-\lambda x} dx = 1 - \mathrm{e}^{-\lambda x}.$$

Now for $y \geq 0$, we have

(b) Solution:

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = 1 - \mathrm{e}^{-\lambda\sqrt{y}}.$$

Differentiating $F_Y(y)$ with respect to y, we have

$$f_Y(y) = \frac{\lambda}{2} y^{-\frac{1}{2}} \mathrm{e}^{-\lambda\sqrt{y}}.$$

Problem 22. *Suppose that X is a random variable that takes on values 0, 2 and 3 with probabilities 0.3, 0.1, 0.6 respectively. Let $Y = 3(X-1)^2$.*

(a) *What is the expectation of X?*

(b) *What is the variance of X?*

(c) *What is the expection of Y?*

(d) *Let $F_Y(t)$ be the cumulative density function of Y. What is $F_Y(7)$?*

(a) Solution: We first make the probability tables

X	0	2	3
prob.	0.3	0.1	0.6
Y	3	3	12

So, $E[X] = 0 \cdot 0.3 + 2 \cdot 0.1 + 3 \cdot 0.6 = 2$

(b) Solution: $E[X^2] = 0 \cdot 0.3 + 4 \cdot 0.1 + 9 \cdot 0.6 = 5.8 \;\Rightarrow\; \mathrm{Var}(X) = E[X^2] - E[X]^2 = 5.8 - 4 = 1.8$.

(c) Solution: $E[Y] = 3 \cdot 0.3 + 3 \cdot 0.1 + 12 \cdot 6 = 8.4$.

(d) Solution: From the table we see that $F_Y(7) = P(Y \leq 7) = 0.4$.

Problem 23. *Suppose you roll a fair 6-sided die 25 times (independently), and you get \$3 every time you roll a 6. Let X be the total number of dollars you win.*

(a) *What is the pmf of X.*

(b) *Find $E[X]$ and $\mathrm{Var}(X)$.*

(c) *Let Y be the total won on another 25 independent rolls. Compute and compare $E[X+Y]$, $E[2X]$, $\mathrm{Var}(X+Y)$, $\mathrm{Var}(2X)$.*

Explain briefly why this makes sense.

(a) Solution: There are a number of ways to present this. Here's one:

$X \sim 3\,\text{binomial}(25, 1/6)$, so

$$P(X = 3k) = \binom{25}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{25-k}, \quad \text{for } k = 0,\ 1,\ 2,\ \ldots,\ 25.$$

(b) Solution: $X \sim 3\,\text{binomial}(25, 1/6)$.

Recall that the mean and variance of binomial(n, p) are np and $np(1-p)$. So,

$$E[X] = 3np = 3 \cdot 25 \cdot \frac{1}{6} = 75/6, \text{ and } \mathrm{Var}(X) = 9np(1-p) = 9 \cdot 25(1/6)(5/6) = 125/4.$$

(c) Solution: $E[X+Y] = E[X] + E[Y] = 150/6 = 25.$, $E[2X] = 2E[X] = 150/6 = 25.$

$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 250/4$. $\mathrm{Var}(2X) = 4\mathrm{Var}(X) = 500/4$.

The means of $X+Y$ and $2X$ are the same, but $\mathrm{Var}(2X) > \mathrm{Var}(X+Y)$.

This makes sense because in $X+Y$ sometimes X and Y will be on opposite sides from the mean so distances to the mean will tend to cancel, However in $2X$ the distance to the mean is always doubled.

Problem 24. *A continuous random variable X has PDF $f(x) = x + ax^2$ on [0,1]*

Find a, the CDF and $P(0.5 < X < 1)$.

Solution: First we find the value of a:

$$\int_0^1 f(x)\,dx = 1 = \int_0^1 x + ax^2\,dx = \frac{1}{2} + \frac{a}{3} \;\Rightarrow\; a = 3/2.$$

The CDF is $F_X(x) = P(X \le x)$. We break this into cases:

(i) $b < 0$, so $F_X(b) = 0$.

(ii) $0 \le b \le 1$, so $F_X(b) = \int_0^b x + \frac{3}{2}x^2\,dx = \frac{b^2}{2} + \frac{b^3}{2}$.

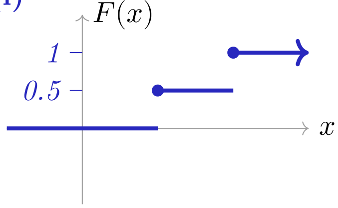
(iii) $1 < x$, so $F_X(b) = 1$.

Using F_X we get

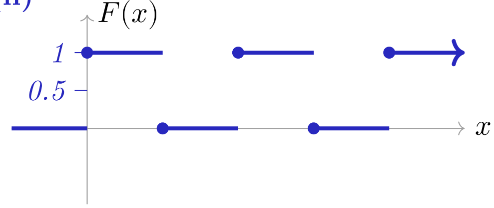
$$P(0.5 < X < 1) = F_X(1) - F_X(0.5) = 1 - \left(\frac{0.5^2 + 0.5^3}{2}\right) = \frac{13}{16}.$$

Problem 25. *For each of the following say whether it can be the graph of a cdf. If it can be, say whether the variable is discrete or continuous.*

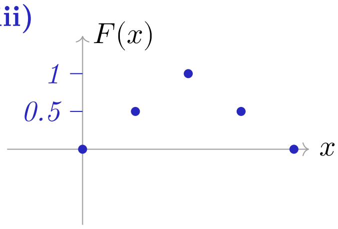
(i)



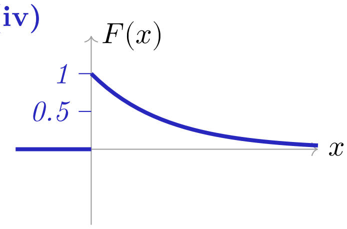
(ii)



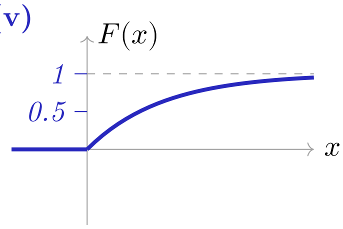
(iii)



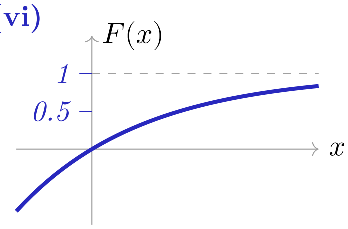
(iv)



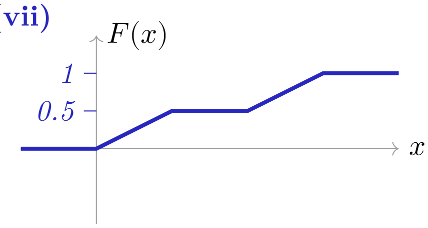
(v)



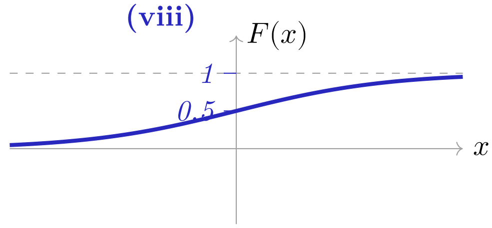
(vi)



(vii)



(viii)



Solution:
(i) yes, discrete, (ii) no, (iii) no, (iv) no, (v) yes, continuous

(vi) no (vii) yes, continuous, (viii) yes, continuous.

Problem 26. Correlation
Flip a coin 5 times. Use properties of covariance to compute the covariance and correlation between the number of heads on the first 3 and last 3 flips.

Solution: As usual let X_i = the number of heads on the i^{th} flip, i.e. 0 or 1.

Let $X = X_1 + X_2 + X_3$ the sum of the first 3 flips and $Y = X_3 + X_4 + X_5$ the sum of the last 3. Using the algebraic properties of covariance we have

$$\begin{aligned}
\text{Cov}(X, Y) &= \text{Cov}(X_1 + X_2 + X_3, X_3 + X_4 + X_5) \\
&= \text{Cov}(X_1, X_3) + \text{Cov}(X_1, X_4) + \text{Cov}(X_1, X_5) \\
&+ \text{Cov}(X_2, X_3) + \text{Cov}(X_2, X_4) + \text{Cov}(X_2, X_5) \\
&+ \text{Cov}(X_3, X_3) + \text{Cov}(X_3, X_4) + \text{Cov}(X_3, X_5)
\end{aligned}$$

Because the X_i are independent the only non-zero term in the above sum is $\text{Cov}(X_3X_3) = \text{Var}(X_3) = \dfrac{1}{4}$
Therefore, $\text{Cov}(X, Y) = \frac{1}{4}$.

We get the correlation by dividing by the standard deviations. Since X is the sum of 3 independent Bernoulli(0.5) we have $\sigma_X = \sqrt{3/4}$

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{1/4}{(3)/4} = \frac{1}{3}.$$

Problem 27. Exponential Distribution
Suppose that buses arrive are scheduled to arrive at a bus stop at noon but are always X minutes late, where X is an exponential random variable with probability density function $f_X(x) = \lambda e^{-\lambda x}$. Suppose that you arrive at the bus stop precisely at noon.

(a) *Compute the probability that you have to wait for more than five minutes for the bus to arrive.*

Solution: We compute

$$P(X \geq 5) = 1 - P(X < 5) = 1 - \int_0^5 \lambda \mathrm{e}^{-\lambda x} dx = 1 - (1 - \mathrm{e}^{-5\lambda}) = \mathrm{e}^{-5\lambda}.$$

(b) *Suppose that you have already waiting for 10 minutes. Compute the probability that you have to wait an additional five minutes or more.*

Solution: We want $P(X \geq 15 | X \geq 10)$. First observe that $P(X \geq 15, X \geq 10) = P(X \geq 15)$. From similar computations in (a), we know

$$P(X \geq 15) = \mathrm{e}^{-15\lambda} \qquad\qquad P(X \geq 10) = \mathrm{e}^{-10\lambda}.$$

From the definition of conditional probability,

$$P(X \geq 15 | X \geq 10) = \frac{P(X \geq 15, X \geq 10)}{P(X \geq 10)} = \frac{P(X \geq 15)}{P(X \geq 10)} = \mathrm{e}^{-5\lambda}$$

Note: This is an illustration of the memorylessness property of the exponential distribution.

Problem 28. Normal Distribution: *Throughout these problems, let ϕ and Φ be the pdf and cdf, respectively, of the standard normal distribution Suppose Z is a standard normal random variable and let $X = 3Z + 1$.*

(a) *Express $P(X \leq x)$ in terms of Φ*

Solution: We have

$$F_X(x) = P(X \leq x) = P(3Z + 1 \leq x) = P(Z \leq \frac{x-1}{3}) = \Phi\left(\frac{x-1}{3}\right).$$

(b) *Differentiate the expression from (a) with respect to x to get the pdf of X, $f(x)$. Remember that $\Phi'(z) = \phi(z)$ and don't forget the chain rule*

Solution: Differentiating with respect to x, we have

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{dx}} F_X(x) = \frac{1}{3}\phi\left(\frac{x-1}{3}\right).$$

Since $\phi(x) = (2\pi)^{-\frac{1}{2}}\mathrm{e}^{-\frac{x^2}{2}}$, we conclude

$$f_X(x) = \frac{1}{3\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-1)^2}{2\cdot 3^2}},$$

which is the probability density function of the $N(1,9)$ distribution. **Note:** The arguments in (a) and (b) give a proof that $3Z+1$ is a normal random variable with mean 1 and variance 9. See Problem Set 3, Question 5.

(c) *Find $P(-1 \le X \le 1)$*

Solution: We have

$$P(-1 \le X \le 1) = P\left(-\frac{2}{3} \le Z \le 0\right) = \Phi(0) - \Phi\left(-\frac{2}{3}\right) \approx 0.2475$$

(d) *Recall that the probability that Z is within one standard deviation of its mean is approximately 68%. What is the probability that X is within one standard deviation of its mean?*

Solution: Since $E[X] = 1$, $\mathrm{Var}(X) = 9$, we want $P(-2 \le X \le 4)$. We have

$$P(-2 \le X \le 4) = P(-3 \le 3Z \le 3) = P(-1 \le Z \le 1) \approx 0.68.$$

Problem 29. Transforming Normal Distributions
Suppose $Z \sim$ N(0,1) and $Y = \mathrm{e}^Z$.

(a) *Find the cdf $F_Y(a)$ and pdf $f_Y(y)$ for Y. (For the CDF, the best you can do is write it in terms of Φ the standard normal cdf.)*

Solution: Note, Y follows what is called a *log-normal distribution.*

$F_Y(a) = P(Y \le a) = P(\mathrm{e}^Z \le a) = P(Z \le \ln(a)) = \boxed{\Phi(\ln(a)).}$

Differentiating using the chain rule:

$$f_y(a) = \frac{d}{da}F_Y(a) = \frac{d}{da}\Phi(\ln(a)) = \frac{1}{a}\phi(\ln(a)) = \boxed{\frac{1}{\sqrt{2\pi}\,a}\mathrm{e}^{-(\ln(a))^2/2}.}$$

(b) *We don't have a formula for $\Phi(z)$ so we don't have a formula for quantiles. So we have to write quantiles* in terms *of Φ^{-1}.*

(i) Write the 0.33 quantile of Z in terms of Φ^{-1}

(ii) Write the 0.9 quantile of Y in terms of Φ^{-1}.

(iii) Find the median of Y.

Solution: (i) The 0.33 quantile for Z is the value $q_{0.33}$ such that $P(Z \le q_{0.33}) = 0.33$. That is, we want

$$\Phi(q_{0.33}) = 0.33 \;\Leftrightarrow\; \boxed{q_{0.33} = \Phi^{-1}(0.33)}.$$

(ii) We want to find $q_{0.9}$ where

$$F_Y(q_{0.9}) = 0.9 \;\Leftrightarrow\; \Phi(\ln(q_{0.9})) = 0.9 \;\Leftrightarrow\; \boxed{q_{0.9} = \mathrm{e}^{\Phi^{-1}(0.9)}}.$$

(iii) As in (ii) $q_{0.5} = \mathrm{e}^{\Phi^{-1}(0.5)} = \mathrm{e}^0 = \boxed{1}$.

Problem 30. (Random variables derived from normal random variables)
Let X_1, X_2, ...X_n be i.i.d. $N(0,1)$ random variables.

Let $Y_n = X_1^2 + \ldots + X_n^2$.

(a) *Use the formula $\mathrm{Var}(X_j) = E[X_j^2] - E[X_j]^2$ to show $E[X_j^2] = 1$.*

Solution: $\mathrm{Var}(X_j) = 1 = E[X_j^2] - E[X_j]^2 = E[X_j^2]$. QED

(b) *Set up an integral in x for computing $E[X_j^4]$.*

For 3 extra credit points, use integration by parts show $E[X_j^4] = 3$.

(If you don't do this, you can still use this result in part c.)

Solution: $E[X_j^4] = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\infty} x^4 \mathrm{e}^{-x^2/2}\, dx$.

(Extra credit) By parts: let $u = x^3$, $v' = x\mathrm{e}^{-x^2/2} \;\Rightarrow\; u' = 3x^2$, $v = -\mathrm{e}^{-x^2/2}$

$$E[X_j^4] = \frac{1}{\sqrt{2\pi}} \left[x^3 \mathrm{e}^{-x^2/2} \Big|_{infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 3x^2 \mathrm{e}^{-x^2/2}\, dx \right]$$

The first term is 0 and the second term is the formula for $3E[X_j^2] = 3$ (by part (a)). Thus, $E[X_j^4] = 3$.

(c) *Deduce from parts (a) and (b) that $\mathrm{Var}(X_j^2) = 2$.*

Solution: $\mathrm{Var}(X_j^2) = E[X_j^4] - E[X_j^2]^2 = 3 - 1 = 2$. QED

(d) *Use the Central Limit Theorem to approximate $P(Y_{100} > 110)$.*

Solution: $E[Y_{100}] = E[100X_j^2] = 100$. $\quad \mathrm{Var}(Y_{100}) = 100\mathrm{Var}(X_j) = 200$.

The CLT says Y_{100} is approximately normal. Standardizing gives

$P(Y_{100} > 110) = P\left(\dfrac{Y_{100} - 100}{\sqrt{200}} > \dfrac{10}{\sqrt{200}} \right) \approx P(Z > 1/\sqrt{2}) = \boxed{0.24}$.

This last value was computed using R: `1 - pnorm(1/sqrt(2),0,1)`.

Problem 31. More Transforming Normal Distributions
(a) *Suppose Z is a standard normal random variable and let $Y = aZ + b$, where $a > 0$ and b are constants.*

Show $Y \sim N(b, a^2)$ (remember our notation for normal distributions uses mean and variance).

Solution: Let $\phi(z)$ and $\Phi(z)$ be the PDF and CDF of Z.
$F_Y(y) = P(Y \le y) = P(aZ + b \le y) = P(Z \le (y-b)/a) = \Phi((y-b)/a)$.
Differentiating:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \Phi((y-b)/a) = \frac{1}{a} \phi((y-b)/a) = \frac{1}{\sqrt{2\pi}\, a}\, \mathrm{e}^{-(y-b)^2/2a^2}.$$

Since this is the density for $\mathrm{N}(b, a^2)$ we have shown $Y \sim \mathrm{N}(b, a^2)$.

(b) *Suppose $Y \sim N(\mu, \sigma^2)$. Show $\dfrac{Y - \mu}{\sigma}$ follows a standard normal distribution.*

Solution: By part (a), $Y \sim \mathrm{N}(\mu, \sigma^2) \Rightarrow Y = \sigma Z + \mu$. But, this implies $(Y - \mu)/\sigma = Z \sim \mathrm{N}(0, 1)$. QED

Problem 32. **(Sums of normal random variables)**
Let X, Y be independent random variables where $X \sim N(2, 5)$ and $Y \sim N(5, 9)$ (we use the notation $N(\mu, \sigma^2)$). Let $W = 3X - 2Y + 1$.

(a) *Compute $E[W]$ and $\mathrm{Var}(W)$.*

Solution: $E[W] = 3E[X] - 2E[Y] + 1 = 6 - 10 + 1 = \boxed{-3}$

$\mathrm{Var}(W) = 9\mathrm{Var}(X) + 4\mathrm{Var}(Y) = 45 + 36 = \boxed{81}$

(b) *It is known that the sum of independent normal distributions is normal. Compute $P(W \leq 6)$.*

Solution: Since the sum of independent normal is normal part (a) shows: $W \sim N(-3, 81)$. Let $Z \sim N(0, 1)$. We *standardize* W: $P(W \leq 6) = P\left(\frac{W+3}{9} \leq \frac{9}{9}\right) = P(Z \leq 1) \approx \boxed{0.84.}$

Problem 33. *Let $X \sim U(a, b)$. Compute $E[X]$ and $\mathrm{Var}(X)$.*

Solution: Method 1

$U(a, b)$ has density $f(x) = \dfrac{1}{b-a}$ on $[a, b]$. So,

$$E[X] = \int_a^b x f(x)\, dx = \frac{1}{b-a} \int_a^b x\, dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b^2 - a^2}{2(b-a)} = \boxed{\frac{a+b}{2}.}$$
$$E[X^2] = \int_a^b x^2 f(x)\, dx = \frac{1}{b-a} \int_a^b x^2\, dx = \left.\frac{x^3}{3(b-a)}\right|_a^b = \frac{b^3 - a^3}{3(b-a)}.$$

Finding $\mathrm{Var}(X)$ now requires a little algebra,

$$\begin{aligned}\mathrm{Var}(X) &= E[X^2] - E[X]^2 = \frac{b^3 - a^3}{3(b-a)} - \frac{(b+a)^2}{4} \\ &= \frac{4(b^3 - a^3) - 3(b-a)(b+a)^2}{12(b-a)} = \frac{b^3 - 3ab^2 + 3a^2b - a^3}{12(b-a)} = \frac{(b-a)^3}{12(b-a)} = \boxed{\frac{(b-a)^2}{12}.}\end{aligned}$$

Method 2

There is an easier way to find $E[X]$ and $\mathrm{Var}(X)$.

Let $U \sim \mathrm{U}(a, b)$. Then the calculations above show $E[U] = 1/2$ and $(E[U^2] = 1/3 \Rightarrow \mathrm{Var}(U) = 1/3 - 1/4 = 1/12$.

Now, we know $X = (b-a)U + a$, so $E[X] = (b-a)E[U] + a = (b-a)/2 + a = (b+a)/2$ and $\mathrm{Var}(X) = (b-a)^2\mathrm{Var}(U) = (b-a)^2/12$.