HW9: Probability, Expectation, Distributions, variance

Due:

Instructions:

- HW instructions
- academic integrity and collaboration

Problem 1 [24 pts: (6 each)]:
The following table lists all the different costs an average individual might have in a typical day. All costs are passed on to insurance. Please assume this table is exhaustive and distinct claims may not occur at the same time.

Claim Type	**Cost ($)**	**Prob**
Routine Check-up	150	.0025
Emergency Room Visit	3,000	.0005
Surgical Procedure	10,000	.001
Chronic Illness Treatment	5,000	.0005
no-service	0	.9955

Table 1: Health Insurance Claims

i Compute the expected cost (i.e. expected value) charged to a health insurance company

ii Suppose many people all pay 10 dollars a day for this insurance. Describe, in one or two easily understood sentences, a pricing problem this insurance company has. Be sure to reference any probability terms you deem relevant.

iii Compute the variance associated with healthcare costs on a given day.

iv Which of the following customers would you expect to have a lower cost variance. Justify your response with one sentence.

- a customer who charges a cheap bottle of tylenol to their insurance weekly
- a customer who gets a major, expensive operation every 10 years

Problem 2 [24 pts: (6 each)]:
Let us model the cars arriving at a traffic light each hour with a Poisson distribution.

i State the independence assumption of the Poisson distribution in this context with one sentence which is easily understood by a non-technical reader.

ii Give an example of a real-life situation which might occur at the traffic light which violates the assumption directly above.

iii State the constant rate assumption of the Poisson distribution in this context with one sentence which is easily understood by a non-technical reader.

iv Give an example of a real-life situation which might occur at the traffic light which violates the assumption directly above.

v After counting the passing cars at the light for ten one hour sessions, you observe the following number of cars in each one-hour timeslot:

$$33, 44, 36, 32, 45, 41, 29, 34, 38, 39$$

Estimate[1] the expected number of cars which pass the intersection each hour from this data.

vi A friend of yours goes out to do data collection and reports that no cars passed through the traffic light in an hour! You suspect that your friend may have gone to the wrong traffic light. Compute the probability that this last observation (X=0) came from your Poisson model from the first 10 observations (use the estimated rate λ from the previous part). Write one sentence which uses this probability to explain if it was possible that your friend was at this traffic light.

Problem 3 Problem 4 Permutation Expectation

We choose a random permutation $a[]$ of the numbers of indices $1 : n$.
(A) What is the expected value of X= the number of elements such that $a_i = i$ (number of fixed points of the permutation)?

[1] Hint: your first intuition is very likely a great estimate!

(B)(optional, no credit) ★★ Calculate $\mathbf{var}[X]$

(C) What is the expected number of $Y =$ inversions (pairs $i < j$ with $a_i > a_j$)?

(D) (optional, no credit) ★★★ Calculate $\mathbf{var}[Y]$

Problem 4 Problem 3 At crossroads

★ A lost tourist arrives at a point with 3 roads. The first road brings him back to the same point after 1 hours of walk. The second road brings him back to the same point after 6 hours of walk. The last road leads to the city after 2 hours of walk. There are no signs on the roads. Assuming that the tourist chooses a road equally likely at all times, what is the average time until the tourist arrives to the city ?

Problem 5 Problem 6 Entropy and Codes

You are given the following 2 codes for a, b, c, d, e, f, g, h:

Letter	Code 1
a	000
b	001
c	010
d	011
e	100
f	101
g	110
h	111

Letter	Code 2
a	11
b	10
c	011
d	010
e	0011
f	0010
g	0001
h	0000

1. Encode the following strings into bits using the 2 different codes (sequences of 0's and 1's).

   (a) abcdefgh

   (b) abbbadgh

   Which code is better?

2. Which code is better if you know that the frequency in which the letters appear is given by:

Letter	a	b	c	d	e	f	g	h
Frequency	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

3. Which code is better if you know that the frequency in which the letters appear is given by:

Letter	a	b	c	d	e	f	g	h
Frequency	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{7}{48}$

4. Which code is better if you know that the frequency in which the letters appear is given by:

Letter	a	b	c	d	e	f	g	h
Frequency	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$

5. Compute the entropy function $H(p)$ for all the above distributions. If the probability is not a power of 2 you don't need to evaluate the $\log_2$.

6. Decode the following sequence of bits into letters using code 1 and using code 2 (It is not always the case that a sequence of bits can be decoded by 2 different codes).

111000111010