HOMEWORK 7
STATISTICS & MLE *

10-606 MATHEMATICAL FOUNDATIONS FOR MACHINE LEARNING

START HERE: Instructions

- **Collaboration Policy**: Please read the collaboration policy in the syllabus.
- **Late Submission Policy:** See the late submission policy in the syllabus.
- **Submitting your work:** You will use Gradescope to submit answers to all questions.
  - **Written:** For written problems such as short answer, multiple choice, derivations, proofs, or plots, please use the provided template. Submissions can be handwritten onto the template, but should be labeled and clearly legible. If your writing is not legible, you will not be awarded marks. Alternatively, submissions can be written in LaTeX. Each derivation/proof should be completed in the boxes provided. To receive full credit, you are responsible for ensuring that your submission contains exactly the same number of pages and the same alignment as our PDF template.
  - **Latex Template:** https://www.overleaf.com/read/hcdnrzsznjwm#082efb

Question	Points
Random Variables	7
Gaussian Random Variables	15
Maximum Likelihood Estimation	8
Total:	30

*Compiled on Monday 6th October, 2025 at 20:45

Instructions for Specific Problem Types

For "Select One" questions, please fill in the appropriate bubble completely:

Select One: Who taught this course?

- ● Matt Gormley
- ○ Marie Curie
- ○ Noam Chomsky

If you need to change your answer, you may cross out the previous answer and bubble in the new answer:

Select One: Who taught this course?

- ● Henry Chai
- ○ Marie Curie
- ~~●~~ Noam Chomsky

For "Select all that apply" questions, please fill in all appropriate squares completely:

Select all that apply: Which are scientists?

- ■ Stephen Hawking
- ■ Albert Einstein
- ■ Isaac Newton
- □ I don't know

Again, if you need to change your answer, you may cross out the previous answer(s) and bubble in the new answer(s):

Select all that apply: Which are scientists?

- ■ Stephen Hawking
- ■ Albert Einstein
- ■ Isaac Newton
- □ I don't know

For questions where you must fill in a blank, please make sure your final answer is fully included in the given space. You may cross out answers or parts of answers, but the final answer must still be within the given space.

Fill in the blank: What is the course number?

$\boxed{\text{10-606}}$ $\boxed{\text{10-60}\cancel{6}\text{7}}$

1 Random Variables (7 points)

Consider two random variables $X \in \{1, 2, 3, 4\}$ and $Y \in \{1, 2\}$. The marginal distribution of X is $P(X) = [0.1, 0.4, 0.3, 0.2]$

The conditional distribution of $Y = 1$ for the four values of X is given by $P(Y = 1 \mid X) = [0.7, 0.7, 0.1, 0.1]$

(Note that both of the above equations represent probability tables, as described in lecture. The free variable in both cases is X.)

1. (2 points) What is $P(X = 2 \mid Y = 1)$?

2. (2 points) Under the same assumptions as above, what is $P(X = 1 \mid Y = 2)$?

3. (1 point) In the probability table $P(X, Y, Z)$, suppose that $X \in \{1, 2\}$, $Y \in \{1, 2, 3\}$, and $Z \in \{1, 2, 3\}$. How many entries does this table have?

4. (2 points) Consider that we flip 5 fair coins. What is probability of at least 4 heads?

2 Gaussian Random Variables (15 points)

Consider a mixture of two Gaussian distributions:

$$0.4\mathcal{N}\left(\begin{pmatrix}10\\2\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right) + 0.6\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}8.4 & 2\\2 & 1.7\end{pmatrix}\right).$$

1. (4 points) Compute the marginal distributions for each dimension.

2. (6 points) Compute the mean, mode and median for each marginal distribution.

3. (5 points) Compute the mean and mode for the two-dimensional (joint) distribution.

3 Maximum Likelihood Estimation (8 points)

Introduction

The Poisson distribution models the number of events that occur within a fixed time interval or space. It is particularly useful when events happen randomly and independently at some average rate λ.

If a random variable X follows a Poisson distribution with parameter λ, then:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, ...$$

The parameter λ represents the expected number of events in the fixed interval, and $P(X = k)$ represents the probability of k events in a fixed interval.

Suppose you are managing a call center and want to understand the pattern of incoming calls during the evening shift. You need to estimate the average call rate λ to properly staff the center. Over the course of one week, you record the number of calls received during the 7pm-8pm hour on different days.

Your observations are: On Monday you received 12 calls, Tuesday 8 calls, Wednesday 15 calls, Thursday 10 calls, and Friday 11 calls. Let X_1, X_2, X_3, X_4, X_5 represent the number of calls on each respective day. Assume each day's call count follows a Poisson distribution with parameter λ, and that days are independent.

1. (2 points) Suppose you knew the true parameter λ. What is the probability of observing a specific sequence of calls, for example $X_1 = 12, X_2 = 8, X_3 = 15, X_4 = 10, X_5 = 11$? Express your answer in terms of λ, X_i, and summation notation.

2. (3 points) At this point, you make three clever observations about finding the value of λ that best explains your data:

   (a) As the number of days n grows very large, the true value of λ should maximize the likelihood function $\mathcal{L}(\lambda)$ above.

   (b) The value of λ that maximizes $\mathcal{L}(\lambda)$ also maximizes the natural log of the likelihood, $\log \mathcal{L}(\lambda)$, since logarithm is a monotonically increasing function.

   (c) At the maximum, the derivative of the log-likelihood with respect to λ equals zero: $\frac{d}{d\lambda} \log \mathcal{L}(\lambda) = 0$.

   These observations motivate us to work with the log-likelihood function $\ell(\lambda) = \log \mathcal{L}(\lambda)$ instead of the likelihood itself, since it's easier to differentiate sums than products.

   Derive the log-likelihood function $\ell(\lambda)$. Simplify your expression.

   Hint The likelihood function for a Poisson distribution is:

$$\mathcal{L}(\lambda) = \prod_{i=1}^{n} \left(\frac{e^{-\lambda} \lambda^{X_i}}{X_i!} \right)$$

3. (3 points) Differentiate the log-likelihood with respect to λ and set it equal to zero. Solve for λ to find the maximum likelihood estimator $\hat{\lambda}_{MLE}$.

4 Collaboration Questions

After you have completed all other components of this assignment, report your answers to these questions regarding the collaboration policy. Details of the policy can be found in the syllabus.

1. Did you receive any help whatsoever from anyone in solving this assignment? If so, include full details.
2. Did you give any help whatsoever to anyone in solving this assignment? If so, include full details.
3. Did you find or come across code that implements any part of this assignment? If so, include full details.

Your Answer