Quiz #1

Due: Oct. 11, 2023 at 10 p.m.

In this quiz, we list some problems that help lead you to quickly review some fundamentals of probability and algebra, which could be useful for future quizzes, programming exercises and exams. To submit your assignment, please upload to moodle a pdf file containing your write-up, which can be either

- **written by hand and scanned to pdf format**, or
- **typed via LaTeXwith the provided template file**.

For the duration of this course, each student is allocated a total of **three slip days** for assignment submissions. These slip days are designed to accommodate any personal emergencies or unforeseen circumstances. You can choose to use these days to extend the deadline for any assignment without incurring a penalty on your grade. However, the combined usage of slip days for all assignments cannot exceed three days. **If you have used up all three slip days and submit an assignment late, you will receive a score of 0 for that assignment.**

Problem 1 (15 points, Bayes' rule)
Box 1 contains 1000 light bulbs of which 10% are defective. Box 2 contains 2000 light bulbs of which 5% are defective.

1. (10 points) Suppose a box is given to you at random and you randomly select a lightbulb from the box. If that lightbulb is defective, what is the probability that you chose from Box 1?
2. (5 points) Suppose now that a box is given to you at random and you randomly select two lightbulbs from the box. If both lightbulbs are defective, what is the probability that you chose from Box 1?

Problem 2 (35 points, Moments of a Distribution)
Prove the following results:

1. (10 points) Assume X and Y are independent of each other, i.e. $P(XY) = P(X)P(Y)$. Show that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ and $\text{var}[X+Y] = \text{var}[X] + \text{var}[Y]$.
2. (5 points) Consider two variables X and Y, show that $\mathbb{E}[X] = \mathbb{E}_Y\left[\mathbb{E}_X[X \mid Y]\right]$.
3. (5 points) Consider a random variable X, prove that $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$.
4. (10 points) Consider a nonnegative random variable X and $a > 0$ show that $P[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$.

   Remark. *This property implies an important theorem:* ***Markov's Inequality.*** *It gives an upper bound for the probability that a non-negative function of a random variable is greater than or equal to some positive constant, e.g. Markov's inequality shows that no more than 1/5 of the population can have more than 5 times the average income.*

5. (5 points) The *covariance* of two random variables X and Y, which describes the degree to which X and Y are related, is defined as $\text{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$. Suppose we have two random variables X and Y such that $\text{var}[X] = \text{var}[Y]$ (i.e., they have the same variance). Let us denote $P = X + Y$ and $Q = X - Y$. Clearly P and Q are dependent since they both depend on the values of X and Y. Show that $\text{Cov}(P,Q) = 0$.

Remark. *Although two independent random variables have 0 covariance, 0 covariance does not necessarily imply independence; this is due to that covariance can only describe a limited range of relationships between two rv's and fail to characterize much more complicated dependencies.*

Problem 3 (25 points, Maximum Likelihood Estimation)
Let $X_1, \cdots, X_n \in \mathbb{R}$ be n sample points drawn independently from univariavte normal distributions such that $X_i \sim N(\mu, \sigma_i^2)$, where $\sigma_i = \sigma/\sqrt{i}$ for some parameter σ. (Every sample point comes from a distribution with a **different variance**.) Note the word "univariate; we are working in demision $d = 1$ and our point are real numbers.

1. (10 points) Derive the maximum likelihood estimates, denoted $\hat{\mu}$ and $\hat{\sigma}$, for the mean µ and the parameter σ. You may write an expression for $\hat{\sigma}^2$ rather than $\hat{\sigma}$ if you wish—it's probably simpler that way. Show all your work.

2. (5 points) Given ther true value of a statistic θ and a estimator $\hat{\theta}$ of that statistic, we define the $bias$ of the estimator to be the expected difference from the true value. That is,

$$bias(\hat{\theta}) = E[\hat{\theta}] - \theta \tag{1}$$

   We say that an estimator is $unbiased$ if its bias is 0. Either **prove** or **disprove** the following statement: *The MLE sample estimator $\hat{\mu}$ is unbiased, that is, $bias(\hat{\mu}) = 0$.*

   Hint: Neither the true μ nor true σ^2 are known when estrimating sample statustics, thus we need to plug in appropiate estimators.

3. (10 points) Either **prove** or **disprove** the following statement: *The MLE sample estimator $\hat{\sigma}^2$ is unbiased, that is, $bias(\hat{\sigma}^2) = 0$.*

   Hint: Neither the true μ nor true σ^2 are known when estrimating sample statustics, thus we need to plug in appropiate estimators.

Problem 4 (25 points, Activation Functions)
In this problem we mainly focus on properties of the sigmoid function

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{2}$$

1. (5 points) Show that $\sigma(x) + \sigma(-x) = 1$.

2. (5 points) Show that $\frac{d}{dx}\sigma(x) = \sigma(x)(1 - \sigma(x))$.

3. (5 points) Calculate $\frac{\partial \sigma(kx+b)}{\partial x}$, $\frac{\partial \sigma(kx+b)}{\partial k}$, and $\frac{\partial \sigma(kx+b)}{\partial b}$

4. (10 points) The sigmoid function is often used to constrain the input into the range $(0, 1)$. There are other constraining functions, one of which is the *tanh* function

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{3}$$

   Tanh function squashes the input $x \in \mathbb{R}$ into $(-1, 1)$, which is more suitable if we want the output range to be symmetric around 0. Show that $\tanh(x) = 2\sigma(2x) - 1$.