**University of Illinois** **Fall 2025**

# ECE 313: Problem Set 5: Problems and Solutions

**Due:** Sunday, October 12 at 07:00:00 p.m. **Note the later due date.**
**Reading:** *ECE 313 Course Notes,* Section 2.8, 2.9.

**Note on reading:** For most sections of the course notes, there are short-answer questions at the end of the chapter. We recommend that after reading each section, you try answering the short-answer questions. Do not hand in; answers to the short answer questions are provided in the appendix of the notes.

**Note on turning in homework:** You must upload handwritten homework to Gradescope. Alternatively, you can typeset the homework in LaTeX. However, no additional credit will be awarded to typeset submissions. No late homework will be accepted. Please write on the top right corner of the first page:
NAME
NETID
SECTION
PROBLEM SET #
Page numbers are encouraged but not required. Five points will be deducted for improper headings.

1. [**A Variant of Geometric Distribution**]
In class, we defined a geometric random variable ($X_{\text{Trials}}$) as the number of *trials* needed to get the first success in a sequence of independent Bernoulli trials with probability of success $p$. Another way of defining a geometric random variable ($X_{\text{Failures}}$) could be to count the number of *failures* before the first success. Do the following.

   (a) Find the pmf of $X_{\text{Failures}}$ and verfiy that it is indeed a pmf.
   **Solution:** A realization $X_{\text{Failures}} = k$ means $k$ failures followed by one success:

   $$P(X_{\text{Failures}} = k) = (1-p)^k\, p, \qquad k = 0, 1, 2, \ldots$$

   Verification:

   $$(1-p)^k\, p \geq 0, k = 0, 1, 2, \ldots$$
   $$\sum_{k=0}^{\infty} (1-p)^k p = p \cdot \frac{1}{1-(1-p)} = 1.$$

   (b) Express $X_{\text{Failures}}$ in terms of $X_{\text{Trials}}$. Calculate the mean and variance of $X_{\text{Failures}}$.
   **Solution:** Relationship:

   $$X_{\text{Trials}} = X_{\text{Failures}} + 1.$$

   Since $X_{\text{Trials}} \sim \text{Geom}(p)$,

   $$\mathbb{E}[X_{\text{Trials}}] = \frac{1}{p}, \qquad \text{Var}(X_{\text{Trials}}) = \frac{1-p}{p^2}.$$

   Therefore,

   $$\mathbb{E}[X_{\text{Failures}}] = \mathbb{E}[X_{\text{Trials}}] - 1 = \frac{1-p}{p}, \qquad \text{Var}(X_{\text{Failures}}) = \text{Var}(X_{\text{Trials}}) = \frac{1-p}{p^2}.$$

(c) Calculate the maximum likelihood estimate of $p$ based on a single observation from $X_{\text{Failures}}$. How does it compare to the maximum likelihood estimate of $p$ based on a single observation from $X_{\text{Trials}}$?

**Solution:** For one observation $k$ from $X_{\text{Failures}}$, the likelihood is

$$L(p) = (1-p)^k p, \quad 0 < p < 1.$$

Log-likelihood:

$$\ell(p) = k \ln(1-p) + \ln p, \qquad \ell'(p) = -\frac{k}{1-p} + \frac{1}{p}.$$

Setting $\ell'(p) = 0$ gives

$$\hat{p}_{\text{MLE}} = \frac{1}{k+1}.$$

For a single observation $t$ from $X_{\text{Trials}}$, the likelihood is

$$L(p) = (1-p)^{t-1} p, \quad 0 < p < 1.$$

Log-likelihood:

$$\ell(p) = (t-1) \ln(1-p) + \ln p, \qquad \ell'(p) = -\frac{t-1}{1-p} + \frac{1}{p}.$$

Setting $\ell'(p) = 0$ gives

$$\hat{p}_{\text{MLE}} = \frac{1}{t}.$$

Since $t = k+1$, the two MLEs agree numerically:

$$\frac{1}{k+1} = \frac{1}{t}.$$

Thus, both formulations yield the same maximum likelihood estimate of $p$.

2. **[Maximum Likelihood Parameter Estimation]**
A biased coin when tossed shows a Heads with probability $p$ and Tails with probability $1-p$.

(a) The biased coin is tossed 10 times, and 6 Heads are observed. What is the maximum likelihood estimate $\hat{p}_{\text{ML}}$ of $p$ given this observation?

**Solution:** Let $X$ denote the number of Heads observed in 10 independent tosses. Then

$$X \sim \text{Binomial}(n = 10, p).$$

The pmf is

$$P(X = k) = \binom{10}{k} p^k (1-p)^{10-k}.$$

For the observed data $X = 6$, the likelihood is

$$L(p) = \binom{10}{6} p^6 (1-p)^4.$$

Taking the log-likelihood:

$$\ell(p) = \ln L(p) = \ln \binom{10}{6} + 6 \ln p + 4 \ln(1-p).$$

Differentiating and setting equal to zero:

$$\ell'(p) = \frac{6}{p} - \frac{4}{1-p} = 0 \quad \Rightarrow \quad 6(1-p) = 4p.$$

Hence,

$$6 - 6p = 4p \quad \Rightarrow \quad 10p = 6 \quad \Rightarrow \quad \hat{p}_{\text{ML}} = 0.6.$$

(b) Suppose it is known that $p = 0.05$. The biased coin is now tossed an unknown number $n$ times during which 6 Heads are observed. What is the maximum likelihood estimate $\hat{n}_{\text{ML}}$ of $n$ given this observation?

**Solution:** If $X$ has binomial distributed with parameters $(n, 0.05)$, and we observe $X = 6$. Thus, the likelihood of observing $X = 6$ is zero if $n < 6$. The likelihood function $L(n)$ for $n \geq 7$ is given by:

$$L(n) = P\{X = 6\} = \binom{n}{6} (0.05)^6 (0.95)^{n-6}. \tag{1}$$

Taking the ratio:

$$\frac{L(n)}{L(n+1)} > 1 \implies n - 5 > (n+1)(0.95) \implies n > 119. \tag{2}$$

This implies that $L(n)$ strictly decreases for $n \geq 120$. Similarly, one can show that $L(n)$ strictly increases for $n \leq 119$. Thus, $\hat{n}_{\text{ML}} = 119$ or $120$.

3. **[Markov Inequality]**
Let $X$ denote the outcome of rolling a fair die. We define two random variables $Y = X^2$ and $Z = X^2 - 15$.

(a) Find $E[Y]$ and $E[Z]$

**Solution:** By LOTUS, we have

$$E[Y] = \sum_{x=1}^{6} x^2 p_X(x) = \frac{1}{6} \sum_{x=1}^{6} x^2 = 15.17$$
$$E[Z] = E[Y - 15] = E[Y] - 15 = 0.17$$

(b) Find the exact probability for $\{Y \geq c\}$ as a function of $c$ for $c \in \{1, 10, 100\}$, and verify if the Markov inequality holds for all these $c$.

**Solution:** We can compute the target probability by summing all the probabilities of outcomes satisfying the condition.

$$P\{Y \geq 1\} = \sum_{\{x:x^2 \geq 1\}} p_X(x) = \sum_{x=1}^{6} p_X(x) = 1$$
$$P\{Y \geq 10\} = \sum_{\{x:x^2 \geq 10\}} p_X(x) = \sum_{x=4}^{6} p_X(x) = \frac{3}{6} = 0.5$$
$$P\{Y \geq 100\} = \sum_{\{x:x^2 \geq 100\}} p_X(x) = 0$$

The corresponding Markov inequalities are

$$P\{Y \geq 1\} = 1 \leq \frac{E[Y]}{1} = 15.17 \qquad \checkmark$$
$$P\{Y \geq 10\} = 0.5 \leq \frac{E[Y]}{10} = 1.517 \qquad \checkmark$$
$$P\{Y \geq 100\} = 0 \leq \frac{E[Y]}{100} = 0.1517 \qquad \checkmark$$

(c) Find the exact probability for $\{Z \geq c\}$ as a function of $c$ for $c \in \{1, 10, 100\}$, and verify if the Markov inequality holds for all these $c$. Why does Markov inequality not hold for $Z$?

**Solution:** We can compute the target probability by summing all the probabilities of outcomes satisfying the condition.

$$P\{Z \geq 1\} = \sum_{\{x:x^2-15 \geq 1\}} p_X(x) = \sum_{x=4}^{6} p_X(x) = \frac{3}{6} = 0.5$$
$$P\{Z \geq 10\} = \sum_{\{x:x^2-15 \geq 10\}} p_X(x) = \sum_{x=5}^{6} p_X(x) = \frac{2}{6} = \frac{1}{3}$$
$$P\{Z \geq 100\} = \sum_{\{x:x^2-15 \geq 100\}} p_X(x) = 0$$

The corresponding Markov inequalities are

$$P\{Z \geq 1\} = 0.5 \nleq \frac{E[Z]}{1} = 0.17 \qquad ✗$$
$$P\{Z \geq 10\} = \frac{1}{3} \nleq \frac{E[Z]}{10} = 0.017 \qquad ✗$$
$$P\{Z \geq 100\} = 0 \leq \frac{E[Z]}{100} = 0.0017 \qquad \checkmark$$

The Markov inequality does not hold because $Z$ is not always non-negative. For $X \leq 3$, $Z = X^2 - 15 < 0$.

4. **[Chip Testing]**
Alice is a graduate student who has designed an integrated circuit (IC) implementing an machine learning accelerator IC in a 45 nm semiconductor process as part of her graduate research. She has just received 50 packaged chips and is getting ready to test them to see if it is working properly. Alice wants to show that her design can classify images with high accuracy $p_a$. To do that she tests her chip with $n$ images and counts the number $E$ that are incorrectly classified. She obtains an accuracy estimate $\hat{p}_a = 1 - \frac{E}{n}$. Alice hopes to write-up a research paper on her design and submit it to ISSCC, a top circuits conference. All she needs is one working chip that classifies images from the test set with high accuracy in order to report the results (yield is not an issue in papers from academia). However, testing a chip is a slow process and Alice wants to minimize the testing time so she can submit the paper before the deadline.

   (a) Determine the probability distribution of the random variable $E$ representing the misclassification error count.

   **Solution:** The error count $E$ can be obtained by running $n$ independent trials of a Bernoulli random variable with parameter $p_a$. Thus, $E$ is a binomial random variable with parameters $(n, 1-p_a)$, i.e., $E \sim \text{Bi}(n, 1-p_a)$.

   (b) Alice tests the first chip using $n = 100$ test images and finds that 95 images are correctly classified. Is it ok for Alice to report that her design gives an accuracy of $p_a = 0.95$? Give reasons.

   **Solution:** No. Suppose the true accuracy is $0 \leq p_a \leq 1$, then testing 100 images will provide an estimate $\hat{p}_a$ which will differ from $p_a$. Alice needs to test with sufficiently large number of test images so that the $p_a$ lies in a small interval (confidence interval) around $\hat{p}_a$ with high probability (high confidence level) and report both.

   (c) How many test vectors should Alice test her chip with so that she can report that the true accuracy of her design $p_a$ lies in the interval $\hat{p}_a \pm 1\%$ with a confidence level greater than 95%?

   **Solution:** Treating each decision of the machine learning accelerator as the outcome of an independent Bernoulli trail with an unknown parameter $p_a$, we have (see equation (2.15) in the course notes):

$$\Pr\left\{p_a \in \left(\hat{p}_a - \frac{a}{2\sqrt{n}}, \hat{p}_a + \frac{a}{2\sqrt{n}}\right)\right\} \geq 1 - \frac{1}{a^2}. \tag{3}$$

   Thus, a 95% confidence level implies $a = \sqrt{20}$. Furthermore, a 1% confidence interval around the estimated accuracy implies the number of test images required are $n = (\sqrt{20})^2/(4 \times (0.01)^2) = 50000$.

   (d) Since testing is a slow process, Alice decides to do a quick pass through all of her 50 chips by testing with 2500 images to find "good parts" which she will then test with large number of vectors as in Part (c). She would still like to achieve a high confidence level of 96%? What confidence interval can she achieve?

   **Solution:** A 96% confidence interval implies $a = 5$. Hence, with $n = 2500$ images, Alice can obtain a confidence interval of:

$$\frac{5}{2\sqrt{2500}} = \pm 5\%.$$

   Thus, if $\hat{p}_a = 90\%$, then her test ensures that the true accuracy $p_a \in (85\%, 95\%)$ with a probability greater than 95%.

5. **[Airline industries]**
Each airplane has capacity for 150 passengers, and overbooking is a common practice in these industries.

(a) To sell more tickets than available seats, the airline needs to estimate the probability that each passenger will attend the flight. Suppose that each passenger will attend the flight with probability $p$. The airline uses $\hat{p}_n = X/n$ as the estimate of $p$, where $n$ is the number of sold tickets and $X$ is the number of people who attended the flight. How large $n$ should be to estimate $p$ within 0.1 with confidence of 0.99?

**Solution:** Notice that

$$P\left(p \in \left(\hat{p} - \frac{a}{2\sqrt{n}}, \hat{p} + \frac{a}{2\sqrt{n}}\right)\right) \geq 1 - \frac{1}{a^2}.$$

Hence to get a confidence of 0.99, we should have $1 - \frac{1}{a^2} = 0.99 \quad \Rightarrow \quad a = 10$.
To estimate within 0.1, we have

$$\frac{a}{2\sqrt{n}} \geq 0.1 \quad \Rightarrow \quad \sqrt{n} \geq 50 \quad \Rightarrow \quad n \geq 2500.$$

(b) According to the historical data, each passenger will attend a flight with probability $p_{\text{attend}} = 0.9$. What is the maximum number of tickets the airline can sell to ensure that no one is left behind with probability 0.75? (Hint: Use Chebyshev's inequality, roots of $0.9x^2 + 0.6x - 150 = 0$ are $-13.25$ and $12.58$.)

**Solution:** Let $X$ denote the number of passengers that will attend the flight. Let $n$ denote the number of sold tickets.
By Chebyshev's inequality we have:

$$P(|X - np| \geq a\sigma) \leq \frac{1}{a^2} \quad \Rightarrow \quad P(X \in (np - a\sigma, np + a\sigma)) \geq 1 - \frac{1}{a^2},$$

where $\sigma = \sqrt{np(1-p)}$.
If we want to make sure that with probability 0.75, no one is left behind, we should have:

$$\text{i) } 1 - \frac{1}{a^2} = 0.75 \quad \Rightarrow \quad a = 2,$$

$$\text{ii) } np + a\sigma \leq 150 \quad \Rightarrow \quad 0.9n + 2\sqrt{0.09n} \leq 150.$$

Let $x = \sqrt{n}$. Then the condition becomes:

$$0.9x^2 + 0.6x = 150.$$

Roots are $-13.25$ and $12.58$. For any $x \in (-13.25, 12.58)$, we have $0.9x^2 + 0.6x - 150 \leq 0$. Hence $n$ is the largest integer smaller than $(12.58)^2$, i.e.

$$n = 158.$$

6. **[Discrete Random Variable on Even Integers]**
Let $X$ denote a discrete random variable that takes on even integer values $0, 2, 4, \ldots, n$, and zero otherwise.

(a) Let the pmf of $X$ be given by

$$p_X(k) = \frac{3(2^k)}{4(2^n) - 1}, \quad \text{for even integer values } k \in \{0, 2, 4, \ldots, n\},$$

where the value of $n$ is unknown. Find the maximum-likelihood estimate $\hat{n}_{\text{ML}}$ from the observation that $X = 10$ on a trial of the experiment.

**Solution:** The likelihood of observing $X = 10$ is

$$p_X(10) = \frac{3(2^{10})}{4(2^n) - 1}.$$

As $n$ increases, the denominator increases, so $p_X(10)$ decreases. Thus, to maximize the likelihood, we choose the smallest even $n$ such that $X = 10$ is possible. That is $n = 10$.

$$\hat{n}_{\text{ML}} = 10.$$

(b) Now, let

$$p_X(k) = a, \quad \text{for even integer values } k \in \{0, 2, 4, \ldots, n\},$$

and zero otherwise. Find the constant $a$ that makes this a valid pmf and compute its mean.

**Solution:** For a valid pmf we must have

$$1 = \sum_{k=0}^{n/2} p_X(2k) = \sum_{k=0}^{n/2} a = a\left(\frac{n}{2} + 1\right).$$

So

$$a = \frac{1}{\frac{n}{2} + 1} = \frac{2}{n + 2}.$$

Since all values are equally likely, the mean is the midpoint:

$$\mu_X = \mathbb{E}[X] = \frac{n}{2}.$$

Alternatively, using the definition:

$$\mathbb{E}[X] = \sum_{k=0}^{n/2} (2k) \cdot a = 2a \sum_{k=0}^{n/2} k = 2a \cdot \frac{\frac{n}{2}\left(\frac{n}{2} + 1\right)}{2} = a \cdot \frac{n(n+2)}{4} = \frac{n}{2}.$$