**University of Illinois** **Fall 2023**

# ECE 313: Problem Set 4: Solutions

**Due:** Friday, September 22 at 7:00:00 p.m.
**Reading:** *ECE 313 Course Notes,* Sections 2.8 - 2.10

**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:** Homework is assigned on a weekly basis on Fridays, and is due by 7 p.m. on the following Friday. 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 AS IT APPEARS ON Canvas
NETID
SECTION
PROBLEM SET #
Page numbers are encouraged but not required. Five points will be deducted for improper headings.
**5 points will be deducted for a submission with incorrectly-assigned page numbers.**

1. [**Chip Testing**]

   (a) 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., $E \sim \mathrm{Bi}(n, 1-p_a)$.

   (b) 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) 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):

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

   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:

   $$1 - \frac{1}{a^2} = 0.8 \implies a = \sqrt{5}.$$

   Thus,

   $$\frac{\sqrt{5}}{2\sqrt{n}} = 0.01 \implies n = 12500.$$

Thus, Alice needs to test with $12,500$ images. It seems she will need to relax the confidence interval as well in order to reduce the number of initial pass test images further.

2. **[Maximum Likelihood Parameter Estimation]**

   (a) The likelihood function $L(p) = P\{6 \text{ Heads in 10 flips}\} = p^6(1-p)^4$. Hence, taking the derivative of the likelihood function $L(p)$:

$$\frac{\partial(p^6(1-p)^4)}{dp} = 6p^5(1-p)^4 - 4p^6(1-p)^3. \tag{2}$$

   Setting the R.H.S. to zero and solving for $p$ gives $p = \frac{6}{10}$. It is easy to see that the derivative is positive for $p < \frac{6}{10}$ and negative for $p > \frac{6}{10}$. Hence, $\hat{p}_{ML} = \frac{6}{10}$.

   (b) 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:

$$L(n) = P\{X = 6\} = \binom{n}{6}(0.05)^6(0.95)^{n-6}. \tag{3}$$

   Taking the ratio:

$$\frac{L(n)}{L(n+1)} > 1 \implies n - 5 > (n+1)(0.95) \implies n > 119. \tag{4}$$

   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, $\hat{n}_{ML} = 119$ or $120$.

   (c) If $X$ is geometrically distributed random variable with parameters $p$, we observe $X = 10$. The likelihood function $L(p)$ for this observation is given by:

$$L(p) = (1-p)^9 p. \tag{5}$$

   Taking the derivative of the likelihood function $L(p)$:

$$\frac{\partial((1-p)^9 p)}{dp} = -9(1-p)^8 p + (1-p)^9. \tag{6}$$

   Setting the R.H.S. to zero and solving for $p$ gives $p = \frac{1}{10}$. It is easy to see that the derivative is positive for $p < \frac{1}{10}$ and negative for $p > \frac{1}{10}$. Hence, $\hat{p}_{ML} = \frac{1}{10}$.

3. **[Message Sources]**

   (a) There are $4^8$ possible words that $M_4$ can produce, out of which $2^8$ are bytes and therefore the probability that $M_4$ produces a byte is $2^{-8}$.

   (b) Applying the Law of Total Probability:

$$P(\text{byte}) = P(\text{byte}|M_2)P(M_2) + P(\text{byte}|M_4)P(M_4) = (1)\frac{1}{2} + \frac{1}{2}\frac{1}{2^8} = 2^{-1} + 2^{-9}.$$

   (c)

$$P(M_2|\text{byte}) = \frac{P(\text{byte}|M_2)P(M_2)}{P(\text{byte})} = \frac{2^{-1}}{2^{-1} + 2^{-9}} = \frac{2^8}{1 + 2^8}.$$

4. **[Debugging a Program]**

   (a) Let $E1$, $E2$, $B1$, and $B2$ denote the relevant events. $B1$ and $B2$ form a partition of the $\Omega$, i.e., $B1 \cap B2 = \phi$ and $B1 \cup B2 = \Omega$. Then, from the law of total probability:

   $$P(E1) = P(E1|B1)P(B1) + P(E1|B2)P(B2) = \frac{1}{5} \times \frac{5}{6} + \frac{2}{3} \times \frac{1}{6} = \frac{5}{18}.$$

   (b) We need to determine $P(B1|E1)$. Using Bayes rule, we get:

   $$P(B1|E1) = \frac{P(E1|B1)P(B1)}{P(E1)} = \frac{\frac{1}{5} \times \frac{5}{6}}{\frac{5}{18}} = \frac{3}{5}.$$

5. **[Coin Draw and Toss]**

   (a) Let $C1$, $C2$, and $C3$ be the events referring to choosing choosing coins numbered 1, 2, and 3, respectively. Let coin 3 be the biased coin. Using the Law of Total Probability,

   $$\begin{aligned} P(T) = \frac{7}{12} &= P(T|C1)P(C1) + P(T|C2)P(C2) + P(T|C3)P(C3) \\ &= \frac{1}{2} \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3} + (1-p) \times \frac{1}{3} \implies p = \frac{1}{4}. \end{aligned}$$

   (b) Let $FF$ be the event that both coins are fair. Let $HT$ denote the event that a head and a tail is observed. We need to determine $P(FF|HT)$. Then,

   $$P(HT|FF) = \frac{1}{2} \times \frac{1}{2} \times 2 = \frac{1}{2}.$$

   Similarly,

   $$P(HT|FF^c) = \frac{1}{2}(1-p) + \frac{1}{2}p = \frac{1}{2}.$$

   Using Law of Total Probability,

   $$P(HT) = P(HT|FF)P(FF) + P(HT|FF^c)P(FF^c) = \frac{1}{2} \times \frac{1}{3} + \frac{1}{2} \times \frac{2}{3} = \frac{1}{2}.$$

   Using Bayes rule,

   $$P(FF|HT) = \frac{P(HT|FF)P(FF)}{P(HT)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}.$$

   This is same as the unconditional probability $P(FF)$. What does this tell you about events $FF$ and $HT$?