1. Chapter 5, problem 2
2. Chapter 5, problem 9
3. Chapter 5, problem 20
4. Chapter 5, problem 32
5. **Binary or ternary:** Suppose you want to transmit coarse images acquired from satellites whose pixels have values in one of 7 discrete levels 1, 2, 3, 4, 5, 6, and 7. You notice that for such a typical image, the probability $P(1) = P(2) = 1/3, P(3) = P(4) = 1/9$ and $P(5) = P(6) = P(7) = 1/27$. You wish to transmit these images from the satellite to an earth station. You have two options: 1) transmit the images over a binary channel (your discrete pixels must be represented in at most two symbols) and 2) transmit the images over a ternary channel (your discrete pixels must be represented in at most three symbols). Both channels are assumed to be noise-free (unlikely for satellite transmission). The costs per symbol for the binary and ternary channels are $1.80$ and $2.70$.

(a) Give a code for the binary channel according to Huffman and compute the average length
(b) Give a code for the ternary channel according to Fano and compute the average length
(c) If you want to minimize the cost, which channel is preferred? Compute these costs
6. **Compression of black and white image:** Consider the following matrix of 0’s and 1’s which represents a black and white image.

\[
\begin{pmatrix}
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
\end{pmatrix}
\]

To compress this image, we want to come up with a probabilistic model and derive its entropy in order to see how much we can compress. To simplify the process, we linearize the image by concatenating the three rows together into a string of 30 bits.

(a) (First model) Assume pixels (0 and 1) are i.i.d binary random variables $X_1, X_2, \ldots, X_{30}$, estimate the probability mass function $p(x)$ of $X_i$. Based on $p(x)$, compute the entropy $H(X_i)$. Using symbol coding $n = 1$, what are the average lengths resulted from using Huffman and Shannon codes to code the entire image?
(b) Assume the model in a), i.e. $p(x_1, x_2) = p(x_1)p(x_2)$, compute the joint entropy of $H(X_i, X_{i+1})$ based on $p(x_1, x_2) = p(x_1)p(x_2)$. Using block coding with $n = 2$, what are the average lengths resulted from using Huffman and Shannon codes?
(c) (Second model) Estimate the first order stationary Markov model for the image, i.e., estimate $p(x_1, x_2)$ directly from the data. Using this model, what is the joint entropy $H(X_i, X_{i+1})$? Using block coding with $n = 2$, what are the average lengths resulted from using Huffman and Shannon codes?
(d) Which of the (a), (b), and (c) has the shortest average length? Why?
7. **Biased coin:**

You saw a con-artist on a New York city sidewalk performing a trick that involves coin tossing. He claimed that the coin is fair, i.e., the probabilities of head and tail are the same at 1/2. However, you notice that the number of heads seems to come out more than the number of tails. You begin to keep track of the outcomes. Sure enough, after 50 coin tosses, you notice that 30 of them are heads. You conclude that the coin is probably biased with $p = 3/5$ of being head. But you want to make sure about your estimate, so you want to bound the probability of $p$ with the following calculations assuming i.i.d for each coin toss.

(a) Using the Markov Inequality $P(X \geq a) \leq E[X]/a$, show that $P(X \geq a) \leq \frac{\prod_i E[e^{tX_i}]}{e^{ta}}$, where $X = \sum_{i=1}^{m} X_i$, $X_i$ are i.i.d, and any $t > 0$
(b) Let $X_i \sim Bern(p)$. Find $\min_{t>0} \frac{\prod_i E[e^{tX_i}]}{e^{ta}}$
(c) Let $a = m(p + \epsilon)$, using the result in (b), show that $P(X \geq m(p + \epsilon)) \leq e^{-D(p+\epsilon||p)m}$
(d) Similarly, show that $P(X \leq m(p - \epsilon)) \leq e^{-D(p-\epsilon||p)m}$