# Harvard University
## Computer Science 20
## Problem Set 9

### PROBLEM 1

The most simple error-correction code is a repetition code. Suppose I want to send you a bit – either a 0, or a 1. But there is a probability $p$ that the bit will get flipped when I transmit it. So I instead send $2k + 1$ copies of the bit, and you use the majority response for the transmitted bit. So if you receive $00110$, you would assume I sent a 0. Suppose each bit is flipped independently with probability $p = 0.02$. Find the smallest value of $k$ so that you end with an erroneous bit with probability less than 1 in a million. Show your work.

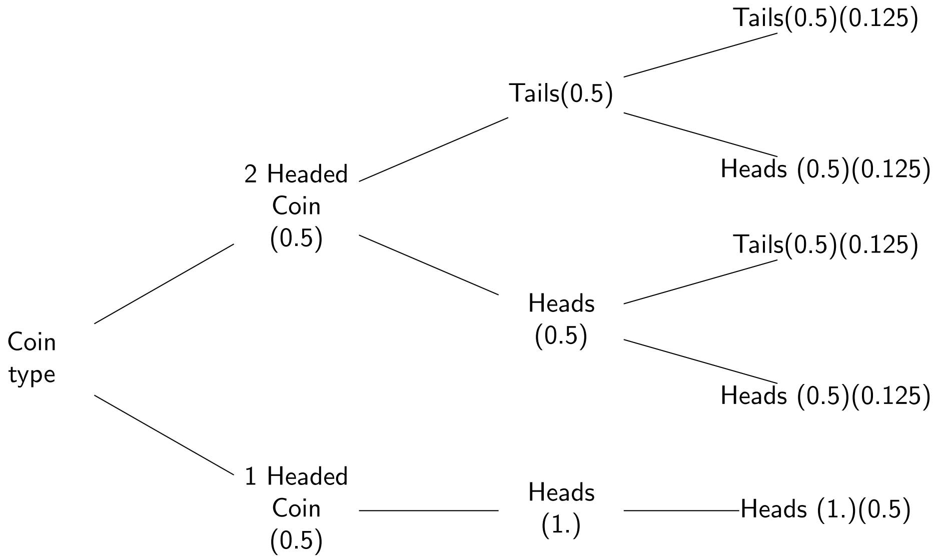
**Solution.**
An erroneous bit flip occurs when $k + 1$ (out of the $2k + 1$ bits) of the message flip. We are trying to determine the number of bits necessary such that the P($k + 1$ bits flipped) is less than 1 in a million.
P($k + 1$ bits flipped) $\leq 10^{-6}$
P(bit flip)$^{k+1} \leq 10^{-6}$
$0.02^{k+1} \leq 10^{-6}$
$k + 1 \geq \log_{0.02} 10^{-6}$
$k \geq 2.532$
Since probability must be less than $10^{-6}$, the min value of k is 3.

# PROBLEM 2

A bag contains two coins: one coin is a standard fair coin, the other has heads on both sides.

(A) Your friend picks a coin from the bag uniformly at random, and then flips it twice, resulting in two heads. If you say that the coin is the two-headed coin, what is the probability that you are wrong?
(B) Now generalize your result; if your friend picks a coin from the bag uniformly at random, and then flips it $k$ times, resulting in $k$ heads, what is the probability you would be wrong if said that the chosen coin was the two-headed coin?

**Solution.**



(A)
The probability that I am wrong would be the probability of two Head flips results from a fair coin being flipped twice. $\frac{\frac{1}{8}}{\frac{1}{8}+\frac{1}{2}} = \frac{1}{5}$
(B)
$\frac{\frac{1}{2}^{k-1}}{\frac{1}{2}^{k-1}+\frac{1}{2}}$

# PROBLEM 3

(A) One simple type of spam filter uses what are called naive Bayes classifiers, which use Bayesian inference to calculate a probability (estimate) that an email is spam or not. Suppose that an email contains the word FREE. This may be a sign that the mail is spam.

Let $S$ be the event that a new email is spam, and let $F$ be the event than a new email contains the word FREE. Given enough data, we calculate (estimates) for $Pr(S)$ (the overall probability a new message is spam), $Pr(F \mid S)$ (the probability the word FREE appears in a spam message), and $Pr(F \mid \neg S)$ (the probability the word FREE appears in a message that is not spam). Give a formula for $Pr(S \mid F)$, the probability a message is spam given the word FREE appears in it, in terms of these above quantities.

Supose we have $Pr(S) = 0.01$, $Pr(F \mid S) = 0.05$ $Pr(F \mid \neg S) = 0.001$. Use your formula to calculate $Pr(S \mid F)$.

(B) (Bonus, you do not need to turn in, this is a bit harder.) Naive Bayes classifiers actually use several the probability calculation for multiple words, treating each word as independent, ignoring any correlation among words. (This is why they are called naive.) So suppose a naive Bayes classifer tests for the words FREE, PRIZE, BANK. Let $P$ be the event than a new email contains the word PRIZE and $B$ be the event that a new email contains the word BANK.

Let us suppose we find the following quantities:

$Pr(S) = 0.01$,<br>
$Pr(F \mid S) = 0.05$,<br>
$Pr(F \mid \neg S) = 0.001$,<br>
$Pr(B \mid S) = 0.04$,<br>
$Pr(B \mid \neg S) = 0.002$,<br>
$Pr(P \mid S) = 0.08$,<br>
$Pr(P \mid \neg S) = 0.001$.

What would be the probability a naive Bayes classifier would give to a new email being spam, if it contains all three words FREE BANK PRIZE.

**Solution.**
(A)

$P(S \mid F) = \frac{P(F \mid S)P(S)}{P(F \mid S)P(S)+P(F \mid \neg S)P(\neg S)}$

$P(S \mid F) = \frac{0.05*0.01}{0.05*0.01+0.001*(1-0.01)} =$ 0.3355704 [solution,letterpaper]cs20 enumerate tikz pgf hyperref multicol float

# Harvard University
## Computer Science 20
## Problem Set 9

### PROBLEM 4

The most simple error-correction code is a repetition code. Suppose I want to send you a bit – either a 0, or a 1. But there is a probability $p$ that the bit will get flipped when I transmit it. So I instead send $2k+1$ copies of the bit, and you use the majority response for the transmitted bit. So if you receive $00110$, you would assume I sent a 0. Suppose each bit is flipped independently with probability $p = 0.02$. Find the smallest value of $k$ so that you end with an erroneous bit with probability less than 1 in a million. Show your work.

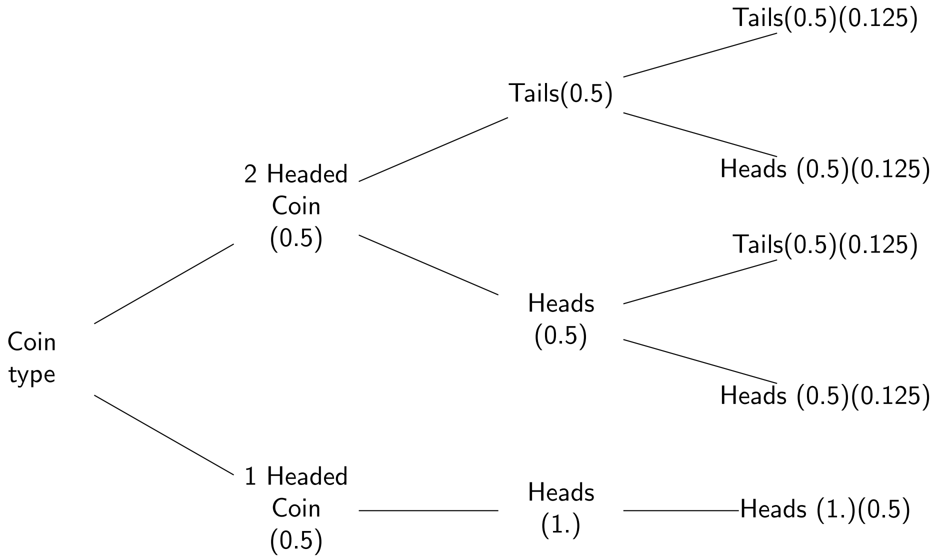
**Solution.**<br>
An erroneous bit flip occurs when $k+1$ (out of the $2k+1$ bits) of the message flip. We are trying to determine the number of bits necessary such that the P($k+1$ bits flipped) is less than 1 in a million.
P($k+1$ bits flipped) $\leq 10^{-6}$
P(bit flip)$^{k+1} \leq 10^{-6}$
$0.02^{k+1} \leq 10^{-6}$
$k+1 \geq \log_{0.02} 10^{-6}$
$k \geq 2.532$
Since probability must be less than $10^{-6}$, the min value of k is 3.

## PROBLEM 5

A bag contains two coins: one coin is a standard fair coin, the other has heads on both sides.

(A) Your friend picks a coin from the bag uniformly at random, and then flips it twice, resulting in two heads. If you say that the coin is the two-headed coin, what is the probability that you are wrong?
(B) Now generalize your result; if your friend picks a coin from the bag uniformly at random, and then flips it $k$ times, resulting in $k$ heads, what is the probability you would be wrong if said that the chosen coin was the two-headed coin?

**Solution.**



(A)
The probability that I am wrong would be the probability of two Head flips results from a fair coin being flipped twice. $\frac{\frac{1}{8}}{\frac{1}{8}+\frac{1}{2}} = \frac{1}{5}$
(B)
$\frac{\frac{1}{2}^{k-1}}{\frac{1}{2}^{k-1}+\frac{1}{2}}$

## PROBLEM 6

(A) One simple type of spam filter uses what are called naive Bayes classifiers, which use Bayesian inference to calculate a probability (estimate) that an email is spam or not. Suppose that an email contains the word FREE. This may be a sign that the mail is spam.

Let $S$ be the event that a new email is spam, and let $F$ be the event than a new email contains the word FREE. Given enough data, we calculate (estimates) for $Pr(S)$ (the overall probability a new message is spam), $Pr(F \mid S)$ (the probability the word FREE appears in a spam message), and $Pr(F \mid \neg S)$ (the probability the word FREE appears in a message that is not spam). Give a formula for $Pr(S \mid F)$, the probability a message is spam given the word FREE appears in it, in terms of these above quantities.

Supose we have $Pr(S) = 0.01$, $Pr(F \mid S) = 0.05$ $Pr(F \mid \neg S) = 0.001$. Use your formula to calculate $Pr(S \mid F)$.

(B) (Bonus, you do not need to turn in, this is a bit harder.) Naive Bayes classifiers actually use several the probability calculation for multiple words, treating each word as independent, ignoring any correlation among words. (This is why they are called naive.) So suppose a naive Bayes classifer tests for the words FREE, PRIZE, BANK. Let $P$ be the event than a new email contains the word PRIZE and $B$ be the event that a new email contains the word BANK.

Let us suppose we find the following quantities:

$Pr(S) = 0.01$,<br>
$Pr(F \mid S) = 0.05$,<br>
$Pr(F \mid \neg S) = 0.001$,<br>
$Pr(B \mid S) = 0.04$,<br>
$Pr(B \mid \neg S) = 0.002$,<br>
$Pr(P \mid S) = 0.08$,<br>
$Pr(P \mid \neg S) = 0.001$.

What would be the probability a naive Bayes classifier would give to a new email being spam, if it contains all three words FREE BANK PRIZE.

**Solution.**
(A)

$P(S \mid F) = \frac{P(F \mid S)P(S)}{P(F \mid S)P(S)+P(F \mid \neg S)P(\neg S)}$

$P(S \mid F) = \frac{0.05*0.01}{0.05*0.01+0.001*(1-0.01)} = 0.3355704$

## PROBLEM 7

You are a contestant on the all new "Let's Make A Deal" with Montana Hill. Montana brings you up to play the 6 doors game. Behind one door is $1200. Behind another door is $120. Behind the other 4 doors is $0. The game works just like Monty's version in that the prizes are randomly distributed and, after you choose a door, Montana will open a door (which will always hide $0) and give you the opportunity to switch your choice.

(A) What is your probability of getting a prize (of either $1200 or $120) if you switch your guess after Montana opens one door? Show your work.
(B) After playing the game, Montana gives you the opportunity to buy a chance to play again. What is the fair price you should be willing to pay to play this game? Show your work.

**Solution.**
(A)

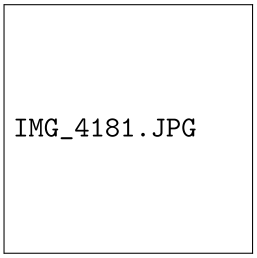


Figure 1: probability of getting prize w/ no switch (collapsed branches)

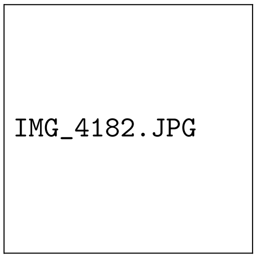


Figure 2: probability of getting prize w/ switch (collapsed branches)

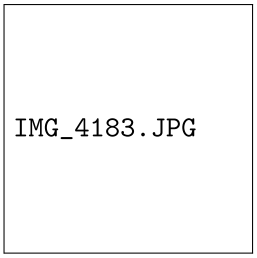


Figure 3: simplified model w/o loss of generality showing all options

Without loss of generality let us consider make it possible to reorder all doors such that we consider the player's initial choice to be the 'first' door.

First layer represents whether or not on your first door selection you chose a winning door. Probabilities were calculated by $\frac{\text{\# ways to select a winning option}}{\text{\# ways to distribute prizes}}$.
Layer 2 represents the (unknown) state of the doors after 1 (empty) door was revealed by Monty hall.
Figure 1 and 2 simplify winning/losing branches by summing probabilities.
Probability of getting a prize if you switch your guess is $\frac{1}{3} \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{1}{2} = \frac{5}{12}$

(B)
The amount we can expect to win (if the switch strategy is used) is $\frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot 1200 + \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot 120 + \frac{2}{3} \cdot \frac{1}{4} \cdot 1200 + \frac{2}{3} \cdot \frac{1}{4} \cdot 120 = 275$. Therefore am I willing to pay anything under \$275 (probabilities used are ffrom figure 3).

## PROBLEM 8

You hold a share of stock in a student-run company ColdX that has invented a drug that cures the common cold. Tomorrow the FDA is going to rule on whether the company can start clinical trials. If they rule "Yes," your share will be worth \$60; if they rule "No," your share will be worth \$20. Shares of ColdX are trading at \$30. So the value of a ColdX share is a random variable $X$, and the market thinks that the expectation of $X$ is \$30.
(A) Determine the probability mass function for $X$, assuming that the market has priced ColdX shares fairly.
(B) A "call option" on ColdX gives its owner the right (but not the obligation) to purchase a share of ColdX for \$40 after the FDA issues its ruling. The value $Y$ of this option is also a random variable. Show that its expectation, which is the fair price of the option, is \$5.

**Solution.**
(A) $\text{PMF}_x(X) = \begin{cases} \frac{1}{4} & x = 60 \\ \frac{3}{4} & x = 20 \end{cases}$
(B)
$E[Y] = \left(\frac{1}{4} \cdot \max(60-40, 0)\right) + \left(\frac{3}{4} \cdot \max(20-40, 0)\right) \because$ possible-gain $\cdot$ probability of possible-gain
$E[Y] = \left(\frac{1}{4} \cdot 20\right) + \left(\frac{3}{4} \cdot 0\right)$
$E[Y] = \frac{20}{4}$
$E[Y] = 5$
we can expect to make \$5 from the option.

## PROBLEM 9

Shuffle a standard 52-card deck, so that the ordering of the cards is perfectly random. You deal the cards out to 4 people (one of whom is you), each getting 13 cards; each set of 13 cards is called a hand. A player gets 4 points for each Ace in the hand, 3 points for each King, 2 for each Queen, 1 for each Jack, and nothing for the other cards. Let the random variable $X$ denote the total number of points in your hand.
(A) What is the probability that you have 0 points in your hand?
(B) What is the expected number of points in your hand?
(C) A friend comes by before you pick up your hand, looks at it, and tells you you have 7 hearts, 2 clubs, 2 diamonds, and 2 spades in your hand. Now what is the expected number of points in your hand.

**Solution.**
(A)

$\Pi_{i=0}^{12}\frac{36-i}{52-i}$
(B)
$\frac{(4+3+2+1)\cdot 4}{4} = 10$
(C)
Expected number of points per card can be calculated by dividing the total points by the number of cards

$$\frac{4\cdot(4+3+2+1)}{52} = \frac{20}{13}\text{points}$$

Since we are dealt $13$ cards we can multiply $13\cdot\frac{10}{13}$ (by linearity of expectation) to get the expected number of points as 10.