Name: MR. YOUNG – SOLUTIONS Date: 11/7/2013 Prd: 1/2/3

PROBLEM SET: BAYESIAN PROBABILITY – HW #27

1.) A lake contains two kinds of trout, Rainbow Trout, which constitute 40% of the population, and Lake Trout. Of the Rainbow trout, 60% are mature, and of the Lake trout, 30% are mature. If a trout is caught in a trap and it appears to be mature, calculate the probability that it is a Rainbow trout.

	MATURE	NOT MATURE
R.T (40%)	60%	40%
L.T. (60%)	30%	70%

$$P(RT \mid Mature) = \frac{P(M \mid RT) \cdot P(RT)}{P(M)}$$

$$= \frac{(0.60) \cdot (0.40)}{(0.60)(0.40) + (0.60)(0.30)}$$

$$= \boxed{57.14\%}$$

2.) A factory producing headsets has three assembly lines, A, B and C, which manufacture 20%, 30% and 50% of the total production, respectively. It is known that the products from assembly lines A, B and C contain defects at rates of 2%, 1% and 3%, respectively. A headset is chosen at random from the output of the factory and is found to be defective. Calculate the probability that it came from assembly lines A, B and C.

	% of total Production	Defective Rate
A	20%	2%
B	30%	1%
C	50%	3%

$$P(A \mid Def.) = \frac{P(D \mid A) \cdot P(A)}{P(D)}$$

$$= \frac{(0.02) \cdot (0.20)}{(0.02 \cdot 0.20) + (0.01 \cdot 0.30) + (0.03 \cdot 0.5)}$$

$$= \frac{(0.02)(0.20)}{0.022}$$

$$= \boxed{18.18\%}$$

$$P(B \mid D) = \frac{0.01 \cdot 0.30}{0.022} = \boxed{13.63\%}$$

$$P(C \mid D) = \frac{(0.5 \cdot 0.03)}{0.022} = \boxed{68.18\%}$$

Name:______________________________________ Date:_________________________ Prd:__________

3.) Suppose that a Bayesian spam filter is trained on a set of 12000 spam messages and 8000 non- spam messages. The word "Viagra" appears in 3000 spam messages, and 10 messages that are not spam. The word "Western Union" appears in 1200 spam messages, and 150 messages that are not spam. Would an incoming message using Bayesian probability estimates be classified as spam if it contains the word "Viagra"? Assume the threshold for accepting as spam is 0.95?

$$P(\text{Spam} \mid \text{viagra}) = \frac{P(V|S)\cdot P(S)}{P(V)}$$

$$= \frac{\left(\frac{3000}{12000}\right)\cdot\left(\frac{12000}{20{,}000}\right)}{\left(\frac{3000}{12{,}000}\right)\left(\frac{12000}{20{,}000}\right) + \left(\frac{10}{8000}\right)\left(\frac{8000}{20{,}000}\right)} = \frac{0.15}{0.15 + \left(\frac{10}{8000}\right)\left(\frac{8000}{20{,}000}\right)} = \boxed{99.67\%}$$

Prob that it is spam.

What if it contained the word "Western Union"? Assume the threshold for accepting as spam is 0.95?

$$\frac{\left(\frac{1200}{12000}\right)\left(\frac{12{,}000}{20{,}000}\right)}{\left(\left(\frac{1200}{12000}\right)\left(\frac{12000}{20{,}000}\right) + \left(\frac{150}{8000}\right)\left(\frac{8000}{20{,}000}\right)\right)} = \frac{0.06}{0.06 + 0.0075} = \boxed{88.88\%}$$

NOT SPAM according to Threshold.

CHALLENGE: Would it determine the email is spam if BOTH words occurred?

DON'T WORRY About THIS.

4.) A hauling truck was involved in a hit and run accident at night. Two truck companies, the Uhaul and Ryder, operate in the city. It is known that 80% of the trucks in the city are Uhaul and 20% are Ryder. A witness identified the truck as Ryer. The court tested the reliability of the witness under the circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two truck types 73% of the time and failed 27% of the time.
What is the probability that the truck involved in the accident was Ryder given that the witness identified the truck as being Ryder?

	correctly identifies	incorrectly identifies
Uhaul (80%)	73%	27%
Ryder (20%)	73%	27%

$$P(\text{Ryder} \mid \text{identified as Ryder}) = \frac{P(I|R)\cdot P(R)}{P(I)}$$

$$= \frac{(0.73)(0.20)}{(0.73)(0.20) + (0.27)(0.80)}$$

$$= 0.4033$$

40.33% chance that the truck involved in the accident was Ryder, given that it was identified as Ryder.

Name:______________________________ Date:____________________ Prd:__________

5.) A man was initially thought to have a 0.1% risk of lung cancer but ended up with a positive test result (the test accurately classifies about 90% of cancerous tumors and 95% of benign tumors.) What is the probability that the man who tested positive actually has lung cancer?

$$P\left(\begin{smallmatrix}\text{actually}\\ \text{has}\\ \text{lung}\\ \text{cancer}\end{smallmatrix} \middle| +\right) = \frac{P(+|C)\cdot P(C)}{P(+)} = \frac{(0.9)(0.001)}{(0.9)(0.001)+(0.999)(0.05)}$$

	+	−
has Cancer	90%	10%
NOT Cancer (benign)	5%	95%

$= 0.017$

1.8% chance that the man who tested positive has lung cancer.

6.) An auto insurance company charges younger drivers a higher premium than it does older drivers because younger drivers as a group tend to have more accidents. The company has 3 age groups: Group A includes those under 25 years old, 36% of all its policyholders. Group B includes those 25-39 years old, 41% of all its policyholders, Group C includes those 40 years old and older. Company records show that in any given one-year period, 19% of its Group A policyholders have an accident. The percentages for groups B and C are 2.5% and 6%, respectively.
(a) What percent of the company's policyholders are expected to have an accident during the next year?
(b) Suppose a person has just had an accident. If he/she is one of the policyholders, calculate the probability that he/she is over 40 years old.

	% of Policyholders	% who have an accident
GROUP A	36%	19%
Group B	41%	2.5%
GROUP C	23%	6%

a.) $(0.36)(0.19)+(0.41)(0.025)+(0.23)(0.06)$

$= 0.09245$

9.2% of All policy holders have an accident in a year.

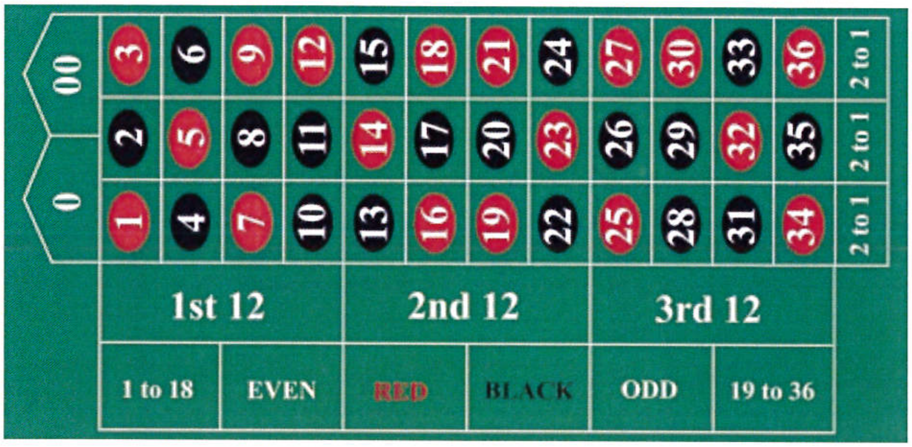
b.) $P\left(\overset{\text{Group}}{C} \middle| \text{Accident}\right) = \frac{P(A|C)\cdot P(C)}{P(A)}$

$= \frac{0.06\cdot 0.23}{0.092} = 0.15$

15% chance that the person involved in the accident was 40 or older.

Name:______________________________ Date:____________________ Prd:________

Below is a roulette table.



Calculate the odds of each below. Keep in mind that when you pick red for instance (0 and 00 are neither red or black or even or odd). This is important for all of the probabilities

38 Total numbers

1.) P(RED) $\boxed{18/38}$

2.) P(EVEN) $\boxed{18/38}$

3.) P(1^{st} 12) $\boxed{12/38}$

4.) P(Picking the number 5) $\boxed{1/38}$

5.) P(a bet covering 4 numbers (10, 11, 13, 14) $\boxed{4/38}$

6.) P(A U B) Where A is defined as EVEN and B is defined as 2^{nd} 12.

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{18}{38} + \frac{12}{38} - \frac{6}{38} = \frac{24}{38} = \boxed{\frac{12}{19}}$ ie. even and 2ND 12

7.) When you pick RED for instance, the payout is double (in other words, if you bet \$100, the casino will pay you \$100 if it lands on red). How much should they pay you (if the game was FAIR)?

Odds: $\frac{18}{38}$ FAIR Payout $\frac{38}{18} = 2.11 \cdot 100 = 211.11$
-100
$\boxed{\$ 111.11}$

should pay out a bit more than doubling your money.

8.) Why might you potentially make money in the short run, but if you play the game 1000 times with similar sized bets, you are almost definitely going to lose money?

IN the short run, you could get lucky and win a handful of times. IN the long run, the actual probability will begin to approach the theoretical probability (i.e. 18/38 for Red). Because the casino is not paying a "fair" payout, you will lose money in the long run.