CPSC 633-600 (Total 100 points)
Hypothesis Testing/Bayesian Learning

See course web page for the **due date**.

Instructor: Yoonsuck Choe

April 19, 2009

1 Hypothesis Testing

Problem 1 (Written: 10 pts): Solve exercise 5.2.

Problem 2 (Program: 15 pts): Suppose you have three random variables X, Y, and Z where each can have values drawn from $\in 1, 2, 3, ..., 10$. (1) If all possible combinations of X, Y, and Z are enumerated, e.g., $\{\langle 1, 1, 1\rangle, \langle 1, 1, 2\rangle, \langle 1, 1, 3\rangle, ..., \langle 10, 10, 10\rangle\}$, how would the distribution of the sum $w = x + y + z$ of each triplet $\langle x, y, z\rangle$ look like? (For example, the sequence of values would be $\{3, 4, 5, ..., 30\}$.) Write a short program to calculate and plot the results. (2) How do the results compare to the results in slide09.pdf, page 9? (3) Which theorem in probability theory is this result related to?

2 MDL

Problem 3 (Program: 15 pts): Write a program to generate random values $x \in \{-2, -1, 0, 1, 2\}$ that have the following probability distribution: Use the *rejection method*, with uniform distribution being the

Table 1: Probability Distribution

v	$P(X = v)$
-2	0.12
1	0.18
0	0.4
1	0.18
2	0.12

comparison function. Basically, this amounts to:

1. Pick a random number $a \in \{-2, -1, 0, 1, 2\}$ (pick with equal probability).
2. Pick another random number b from $[0..1]$ (to speed it up, you can pick it from $[0..0.4]$).
3. If $b \leq P(X = a)$ (see Table 1 for the probabilities),
   - then accept a as a new random value,
   - else reject a and repeat from step 1.

Experiment: pick 100 values and plot the frequency of occurrence for the five possible values. Divide the frequency with the sum of frequencies in each experiment to show the empirical probability for each value. Repeat the experiment for 1,000 and 10,000 samples.

Problem 4 (Program: 15 pts): Suppose there are 128 messages m_1, m_2, ... m_{128} each occurring with probability $P(m_i) = c \cdot \exp\left(-\frac{(i-64)^2}{500}\right)$, where c is a normalizing constant that makes $\sum_i P(m_i) = 1$.

(1) How many bits are needed to represent all 128 messages if all message lengths are the same?

(2) How many bits are needed on average, assuming an optimal coding (message length for $m_i = -\log_2 P(m_i)$)? This is simply:

$$\sum_{i=1}^{128} -P(m_i) \log_2 P(m_i)$$

3 Conditional Independence

Problem 5 (Written: 15 pts): Consider three binary variables $a, b, c \in \{0, 1\}$ having the joint probability distribution given in the table below. Show by direct evaluation that this distribution has the property that a and b are dependent, so that $P(a, b) \neq P(a)p(b)$, but that they become independent when conditioned on c, so that $P(a, b|c) = P(a|c)p(b|c)$ for both $c = 0$ and $c = 1$ [from C. M. Bishop, *Pattern Recognition and Machine Learning*, Singapore: Springer, 2006].

Table 2: Joint Probability

a	b	c	P(a,b,c)
0	0	0	0.192
0	0	1	0.144
0	1	0	0.048
0	1	1	0.216
1	0	0	0.192
1	0	1	0.064
1	1	0	0.048
1	1	1	0.096

Problem 6 (Written: 10 pts): How is the above result related to the concept of conditional independence $P(a|b, c) = P(a|c)$? Show the relation analytically.

4 Naive Bayes Classifier

Answer the following questions given the data set in the following table.

Table 3: Classification

a_1	a_2	v
0	0	-
0	1	+
1	1	-
1	2	+
2	2	-
2	3	+
3	3	-
3	4	+
4	4	-

Problem 7 (Written: 10 pts): Use the Naive Bayes classifier to classify instance $(a_1, a_2) = (2, 2)$. What are the estimated posterior probabilities $P(+|a_1 = 2, a_2 = 2)$ and $P(-|a_1 = 2, a_2 = 2)$ under Naive Bayes?

Problem 8 (Written: 10 pts): (1) Is the result consistent with the true class (-)? (2) Why not? (hint: check if the conditional independence assumption holds).

5 Bayesian Belief Network (This part is optional: solve for extra credit)

Problem 9 (Written: extra credit 5 pts): Based on Table 2 draw a Bayesian Belief Network (BBN) that correctly represents the conditional independence relation.

Problem 10 (Written: extra credit 5 pts): For each node in the BBN, calculate the conditional probability table, again from Table 2.

Problem 11 (Program: extra credit 10 pts): Write a program to do Monte Carlo estimation of the conditional probability $P(a|b)$ (generate three sets with 10, 100, and 1000 examples). Compare the simulation results from the three sets, and also calculate (by hand) the true $P(a|b)$ from Table 2. Are the results comparable?