Probability Final Exam

Instructor: Prof. Shou-De Lin
14:30 ∼ 17:30, Wed., June 19th, 2008

• Total score: 120 points
• Final exam grades and final grades will be e-mailed in early July.
  You have three days to make petitions.

Problems

1. Let \( X \) have the following probability density function:

\[
f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.
\]

What is the probability density function of \( Y = e^X \)? (5 points)

Ans.

\[
P(Y \leq y) = P(e^X \leq y) = P(X \leq \ln y),
\]

and therefore

\[
f_Y(y) = \frac{dP(Y \leq y)}{dy} = \frac{dP(X \leq \ln y)}{d\ln y} \cdot \frac{d\ln y}{dy} = \frac{1}{y\sqrt{2\pi\sigma}} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}}, \quad 0 < y.
\]

2. Person A throws an unbiased dice \( n \) times and B throws the same dice \( n + 1 \) times.
   We care about how many ‘6’ s they throw. If you are told that

\[
P(B \text{ has more ‘6’ s than } A) = \frac{5}{12},
\]

then what is the probability that A and B have equally many ‘6’ s after throwing the dice \( n \) times? (7 points)

Hint: conditioning on which player has more ‘6’ s after each has thrown \( n \) times

Ans. Let A and B respectively throw \( A_n \) and \( B_n \) ‘6’ s in \( n \) times. The probability that B has more ‘6’ s than A is

\[
\frac{5}{12} = P(B_n > A_n) + P(B_n = A_n)P(B \text{ throws a ‘6’ in the } (n+1) \text{th time})
\]

\[
= \frac{1 - P(B_n = A_n)}{2} + \frac{P(B_n = A_n)}{6},
\]

so \( P(A_n = B_n) = 1/4. \)
3. Your company must make a sealed bid for a construction project. Your company will win if your bid is lower than other companies. If you win the bid, then you plan to pay another firm 100 thousand dollars to do the work. You are competing with two other companies, and you believe their bids are two independent random variables uniformly distributed in [70, 250] and [140, 300], respectively.

(a) Suppose your bid is $x$, what is the probability that you win? (4 points)

**Ans.**

When $x \in [0, 70)$, we will win.

When $x \in [70, 140)$, we only have to beat the company whose bid is in [70, 250]. Therefore, the probability to win is

$$\frac{250 - x}{180}.$$  

When $x \in [140, 250]$, we need to beat both of the competitors, so the winning probability is

$$\left(\frac{250 - x}{180}\right) \left(\frac{300 - x}{160}\right).$$

When $x > 250$, we will lose.

(b) Suppose your bid is $x$, what is the expected profit? (2 points)

**Ans.**

$$\text{expected profit} =  
\begin{cases} 
  x - 100 & \text{if } x \in [0, 70), \\
  \frac{(x-100)(250-x)}{180} & \text{if } x \in [70, 140), \\
  \frac{(x-100)(250-x)(300-x)}{180\cdot 160} & \text{if } x \in [140, 250], \\
  0 & \text{if } x > 250. 
\end{cases}$$

(c) Determine the $x$ that maximizes your profit. (2 points)

**Ans.** $x = (1300 - 100\sqrt{13})/6 \approx 156.57$, the maximum expected profit is $(156.57 - 100)(250 - 156.57)(300 - 156.57)/(180\cdot 160) \approx 26.32$ thousand dollars.

4. $X$, $Y$, and $Z$ are three random variables. Can you propose a real-world example of them that satisfy the following conditions (6 points):

(a) $X$ and $Y$ are independent.

(b) $X$ and $Y$ become dependent given $Z$.

**Ans.** $X$: father’s blood type, $Y$: mother’s blood type, $Z$: their child’s blood type.

5. $T_1$ and $T_2$ are two positive continuous random variables that satisfy:

- $T_1 > T_2$.
- $T_1 + T_2 < 2$. 

2
Their joint density function is uniform in the above region, and is zero elsewhere. What is $P(T_1 + T_2 > 1)$? (6 points)

**Ans.** According to the following figure,

$$
P(T_1 + T_2 > 1) = \frac{\text{Area of the shaded region}}{\text{Area of the triangle } \Delta_{(0,0),(1,1),(2,0)}} = \frac{2 \cdot 1/2 - 1 \cdot 0.5/2}{2 \cdot 1/2} = \frac{3}{4}.
$$

6. Correlation Coefficient:
Let $X$ and $Y$ be random variables of the continuous type having the joint p.d.f. $f(x, y) = 2$, $0 \leq y \leq x \leq 1$.

(a) What are the means of $X$ and $Y$? (4 points)

**Ans.**

$$
\mu_X = \int_0^1 x \cdot 2 dy = \frac{2}{3},
$$

$$
\mu_Y = \int_0^1 2(1-y) dy = \frac{1}{3}.
$$

(b) What is the covariance of $X$ and $Y$? (4 points)

**Ans.**

$$
\text{Cov}[X,Y] = E[XY] - \mu_X \mu_Y = \int_0^1 \int_0^x 2xy dy dx - \frac{2}{3} \cdot \frac{1}{3} = \frac{1}{3}.
$$

7. Hypothesis Test:
A public poll was taken to determine whether we should allow tourists from Mainland China. Let $p$ equal the proportion of people who favor this decision. We shall test $H_0 : p = 0.65$ against $H_1 : p > 0.65$. 
(a) Given $\alpha = 0.025$, what is the critical region? (5 points)

**Ans.** Let $n$ be the sample size and $y$ be the number of positive votes. The critical region for $\alpha = 0.025$ is

$$z = \frac{y/n - 0.65}{\sqrt{(0.65)(0.35)/n}} \geq 1.96.$$ 

(b) Given that 414 out of a sample of 600 favor this proposal, find the $p$-value. (3 points)

**Ans.**

$$z = \frac{414/600 - 0.65}{\sqrt{(0.65)(0.35)/600}} = 2.054,$$

and the $p$-value $\approx P(Z \geq 2.054) = 0.0200$.

(c) Should we reject or accept $H_0$? (2 points)

**Ans.** Since $z > 1.96$ and the $p$-value $< 0.0250$, we reject $H_0$ at an $\alpha = 0.025$ significance level.

8. Chi-Square Test: (8 points)

The teacher claims that $1/4$ of the students will receive A grade, $1/4$ will receive B and $1/2$ will receive C grade. If among the 40 students, 6 receive A, 7 receive B and 27 receive C. Would the claim be rejected at $\alpha = 0.05$ significance level?

**Ans.** $q_2 = 4.95 < 5.991 = \chi^2_{0.05}(2)$, so we do not reject $H_0$ at $\alpha = 0.05$.

9. Mutual Information: (7 points)

A six-sided fair die is rolled. What is the mutual information between the topside and the front face (the side most facing you)?

**Hint:** The sum of two opposite sides is always 7.

**Ans.** Note that having observed a side $F$ of the cube facing us, there are four equally probable possibilities for the top $T$. Thus,

$$I(T; F) = H(T) - H(T|F)$$

$$= \log 6 - \log 4$$

$$= \log 3 - 1,$$

since $T$ has a uniform distribution on $\{1, 2, \ldots, 6\}$.

10. Bayesian Network and Association Rule:

Half of the Taiwanese students in the class get high score, and $2/3$ of the students in the class are Taiwanese. Only $1/10$ of the non-Taiwanese students get high score.

(a) Define the random variables and draw the Bayesian network (with conditional probability table) for this statement. (2 points)

**Ans.**

```
   T
  /|
 / | \\
H P(T) = 2/3
  |
  | P(H| T) = 1/2
  |
  P(H| T) = 1/10
  ```
(b) What is the probability that a randomly chosen student is a Taiwanese who gets high score? (3 points)

**Ans.** $P(T, H) = P(T)P(H|T) = (2/3)(1/2) = 1/3$.

(c) Given an association rule that says “Japanese = true” $\Rightarrow$ “score = high,” please provide a pair of “reasonable” min-support and min-confidence that make this rule true. (5 points)

**Ans.** The answers have to satisfy the following conditions:
- min-support < $P(score = high)$,
- $P(score = high)$ < min-confidence.


11. Bayesian Network Inference: (10 points)

Given the following Bayesian network,

```
    H
   / \  
  P(H) = 1/2  
  F
  / \  
 S  P(F|H) = 0
  |  
 W
  |
 P(W|S) = 1/2  
  |
 \
 P(W|\neg S) = 1
```

please calculate $P(W, F)$.

**Ans.** $P(W, F) = P(W, F, H) + P(W, F, \neg H)$, but $P(W, F, H) = 0$. Therefore it suffices to compute

\[
P(W, F, \neg H) = P(W, F|\neg H)P(\neg H) = P(W|\neg H)P(F|\neg H)P(\neg H) = (P(W|S)P(S|\neg H) + P(W|\neg S)P(\neg S|\neg H)) \frac{1}{4}
\]

(Given $S$, $W$ is independent of $H$)

\[
= (\frac{3}{2} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} \cdot \frac{1}{4}) = \frac{3}{16}.
\]

12. TFIDF

Corpus C consists of only three documents:

- $D_1$: “new york times”
- $D_2$: “new york post”
- $D_3$: “los angeles times”
(a) Please use the vector-space model to represent these three documents, assuming the weights are all binary and the words in the vectors are ordered alphabetically. (2 points)

\[ \begin{array}{cccccc}
\text{angeles} & \text{los} & \text{new} & \text{post} & \text{times} & \text{york} \\
\text{new york times} & 0 & 0 & 1 & 0 & 1 \\
\text{new york post} & 0 & 0 & 1 & 1 & 0 & 1 \\
\text{los angeles times} & 1 & 1 & 0 & 0 & 1 & 0 \\
\end{array} \]

(b) Please use the vector-space model to represent these three documents, assuming the weights are TFIDF values and assuming that term frequencies are normalized by the maximum frequency in a document. (5 points)

Note: Please use the base-10 logarithm with the following table:

<table>
<thead>
<tr>
<th>( \log_{10} )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log_{10} 2 )</td>
<td>0.3010</td>
<td>0.4771</td>
<td>0.6021</td>
<td>0.6990</td>
<td>0.7782</td>
<td>0.8451</td>
<td>0.9031</td>
<td>0.9542</td>
</tr>
</tbody>
</table>

\[ \begin{array}{cccccc}
\text{angeles} & \text{los} & \text{new} & \text{post} & \text{times} & \text{york} \\
\text{new york times} & 0 & 0 & 0.1761 & 0 & 0.1761 & 0.1761 \\
\text{new york post} & 0 & 0 & 0.1761 & 0.4771 & 0 & 0.1761 \\
\text{los angeles times} & 0.4771 & 0.4771 & 0 & 0 & 0.1761 & 0 \\
\end{array} \]

Since \( \log_{10} 3 = 0.4771 \) and \( \log_{10} \frac{3}{2} = 0.1761 \), we have the following TFIDF weighted term vectors:

(c) Given the following query: “new new times,” calculate the corresponding TFIDF-based vector, and compute its distance with \( D_1 \) using the cosine similarity measure. Assume that term frequencies are normalized by the maximum frequency in a given query. (3 points)

\[ \begin{array}{cccccc}
\text{angeles} & \text{los} & \text{new} & \text{post} & \text{times} & \text{york} \\
\text{new york times} & 0 & 0 & \frac{3}{2} \log_{10} \frac{3}{2} = 0.1761 & 0 & 0.1761 & 0.1761 \\
\text{new york post} & 0 & 0 & 0.1761 & 0.4771 & 0 & 0.1761 \\
\text{los angeles times} & 0.4771 & 0.4771 & 0 & 0 & 0.1761 & 0 \\
\end{array} \]

The vector lengths of the query and \( D_1 \) are \( \sqrt{0.1761^2 + 0.0881^2} = 0.1969 \) and \( \sqrt{0.1761^2 + 0.1761^2 + 0.1761^2} = 0.3050 \), respectively. Hence, the cosine similarity measure between the query and \( D_1 \) is

\[ \frac{0.1761 \times 0.1761 + 0.0881 \times 0.1761}{0.1969 \times 0.3050} = 0.7747. \]
There are six paths of length 2:

Sue \overset{\text{calls}}{\rightarrow} N_1 \overset{\text{calls}}{\rightarrow} Jean

Jean \overset{\text{emails}}{\rightarrow} Sue \overset{\text{emails}}{\rightarrow} P_1

Jean \overset{\text{emails}}{\rightarrow} Sue \overset{\text{calls}}{\rightarrow} N_1

N_1 \overset{\text{calls}}{\rightarrow} Jean \overset{\text{calls}}{\rightarrow} C_1

N_1 \overset{\text{calls}}{\rightarrow} Jean \overset{\text{calls}}{\rightarrow} P_1

N_1 \overset{\text{calls}}{\rightarrow} Jean \overset{\text{emails}}{\rightarrow} Sue

If we perform a random experiment to pick a length-2 path randomly, and define two random variables $S$ and $P$:

$S$: the starting node of the path (e.g., “Sue”)

$P$: the link-combination of the path (e.g., \{calls, emails\}, \{calls, calls\})

(a) What is the size of $P$’s outcome space? (2 points)

Ans. 4.

(b) What is the mutual information $I(S; P)$? (5 points)

Ans. We need to average all possible combinations of PMI($S$, $P$):

$$\frac{1}{6} \log \frac{\frac{1}{6}}{\frac{1}{2}} + \frac{1}{6} \log \frac{\frac{1}{6}}{\frac{1}{2}} + \frac{1}{6} \log \frac{\frac{1}{6}}{\frac{1}{2}} + \frac{1}{6} \log \frac{\frac{1}{6}}{\frac{1}{2}} + \frac{1}{6} \log \frac{\frac{1}{6}}{\frac{1}{2}} = \log 2.$$

(c) Assume that min-support is 0.3 and min-confidence is 0.7, can we conclude an association rule $N_1 \rightarrow \{\text{calls, calls}\}$? Why? (3 points)

Ans. No. Although support = $2/6 = 1/3 >$ min-support, confidence = $2/3 <$ min-confidence.

(d) Assume the initial PageRank values for each node is 0.2. Which node(s) have the highest PageRank values after two iterations? (5 points)

Ans. Let “≈” denote normalization.
\[ T = 0 \quad T = 1 \quad T = 2 \]

<table>
<thead>
<tr>
<th></th>
<th>[ T = 0 ]</th>
<th>[ T = 1 ]</th>
<th>[ T = 2 ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sue</td>
<td>( \frac{1}{5} )</td>
<td>( \frac{1}{3} \cdot \frac{1}{3} \approx \frac{1}{9} )</td>
<td>( \frac{1}{3} \cdot \frac{1}{3} \approx \frac{2}{11} )</td>
</tr>
<tr>
<td>Jean</td>
<td>( \frac{1}{5} )</td>
<td>( \frac{1}{3} \cdot \frac{1}{3} \approx \frac{1}{9} )</td>
<td>( \frac{1}{3} \cdot \frac{1}{3} \approx \frac{2}{11} )</td>
</tr>
<tr>
<td>( N_1 )</td>
<td>( \frac{1}{5} )</td>
<td>( \frac{1}{3} \cdot \frac{1}{3} \approx \frac{1}{9} )</td>
<td>( \frac{1}{3} \cdot \frac{1}{3} \approx \frac{2}{11} )</td>
</tr>
<tr>
<td>( P_1 )</td>
<td>( \frac{1}{5} \cdot \frac{1}{2} \cdot \frac{1}{3} \approx \frac{1}{9} )</td>
<td>( \frac{1}{2} \cdot \frac{1}{3} \approx \frac{5}{18} )</td>
<td>( \frac{1}{2} \cdot \frac{1}{3} \approx \frac{3}{11} )</td>
</tr>
<tr>
<td>( C_1 )</td>
<td>( \frac{1}{5} )</td>
<td>( \frac{1}{3} \cdot \frac{1}{3} \approx \frac{1}{9} )</td>
<td>( \frac{1}{3} \cdot \frac{1}{3} \approx \frac{2}{11} )</td>
</tr>
</tbody>
</table>

Therefore, Jean and \( P_1 \) have the highest PageRank values after two iterations.

14. N-gram Language Model: (10 points)

The problem of “Chinese poetry segmentation” aims at breaking a Chinese poetry sentence into a section of terms, for example,

\[ \text{“夜半鐘聲到客船”} \rightarrow \text{“夜半 鐘聲 到 客船.”} \]

Can you carefully describe a way to use n-gram LM to do this job?

Hint: You need to determine not only where to put the breaks but also how many breaks there are.

**Ans.** A plausible answer is as the following. First calculate the bi-gram probability for each term (using the poetry corpus), for example

\[
\begin{array}{cccccc}
0.5 & 0.01 & 0.7 & 0.3 & 0.2 & 0.5 \\
\end{array}
\]

\text{夜 半 鐘 聲 到 客 船}

some break points. Assuming each break is a binary decision, then in this case we will have \( 2^6 \) different choices. For each case, we can calculate the corresponding probability. For example, in the following case the probability is \( (0.5 \times 0.7 \times 0.3 \times 0.5)/3 \) (we normalize the values by dividing the whole probability by the number of sections).

\[
\begin{array}{cccccc}
0.5 & 0.01 & 0.7 & 0.3 & 0.2 & 0.5 \\
\end{array}
\]

\text{夜 半 鐘 聲 到 客 船}