Ho Vinh Thinh (2562630)
Prabal Agarwal (2561419)
Nurzat Rakhmanberdieva (2561148)

Problem 1

To prove:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If $A_1, A_2, ..., A_i, ...A_n$ is a set of mutually exclusive events, then $P(A_i \cap A_j) = 0$ and $P(A_1 \cup A_2 \cup ...A_i \cup ... \cup A_n) = P(A_1) + P(A_2) + .. + P(A_i) + .. + P(A_n)$

Considering the following disjoint set and using the axiom defined above:

$$\begin{aligned} A \cup (B \cap A^c) &= (A \cup B) \cap (A \cup A^c) \\ &= A \cup B \end{aligned}$$

where $A^c = R - A$

Therefore,

$$\begin{aligned} P(A \cup B) &= P(A \cup (B \cap A^c)) && (1) \\ &= P(A) + P(B \cap A^c) && (2) \end{aligned}$$

Considering another disjoint set and using the axiom defined above:

$$\begin{aligned} B &= B \cap (A \cup A^c) \\ &= (B \cap A) \cup (B \cap A^c) \end{aligned}$$

Therefore:

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap A^c) && (3) \\ P(B \cap A^c) &= P(B) - P(B \cap A) && (4) \end{aligned}$$

Using eq. 2 and 4, we get:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (5)$$

Ho Vinh Thinh (2562630)
Prabal Agarwal (2561419)
Nurzat Rakhmanberdieva (2561148)

Problem 2

Given the following data:

	$P(y=0)$	$P(y=1)$
$P(X=0)$	0.32	0.08
$P(X=1)$	0.48	0.12

Marginal probability distribution is given by:

$$P(X=x) = \sum_{y \in V} P(X=x, Y=y)$$
$$P(Y=y) = \sum_{x \in V} P(X=x, Y=y)$$

Marginal distribution probabilities:

$P(X=0) = P(X=0, Y=0) + P(X=0, Y=1) = 0.40$
$P(X=1) = P(X=1, Y=0) + P(X=1, Y=1) = 0.60$
$P(Y=0) = P(X=0, Y=0) + P(X=1, Y=0) = 0.80$
$P(Y=1) = P(X=0, Y=1) + P(X=1, Y=1) = 0.20$

Conditional probability distribution is given by:

$$P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y)}$$
$$P(Y=y|X=x) = \frac{P(X=x, Y=y)}{P(X)}$$

Table 1: Conditional prob. dist. $P(X=x|Y=y)$

	$P(y=0)$	$P(y=1)$
$P(X=0)$	0.40	0.40
$P(X=1)$	0.60	0.60

Table 2: Conditional prob. dist. $P(Y=y|X=x)$

	$P(y=0)$	$P(y=1)$
$P(X=0)$	0.80	0.20
$P(X=1)$	0.80	0.20

Condition for independence:
$P(A \cap B) = P(A) * P(B)$

Table 3: $P(X=x) * P(Y=y)$

	$P(y=0)$	$P(y=1)$
$P(X=0)$	0.32	0.08
$P(X=1)$	0.48	0.12

Hence as $P(X \cap Y) = P(X) * P(Y)$, $P(X)$ and $P(Y)$ are independent events.

Ho Vinh Thinh (2562630)
Prabal Agarwal (2561419)
Nurzat Rakhmanberdieva (2561148)

Problem 3

Given:
average length of each sentence = 5
Number of words in vocabulary = 30k

As, $P(w_1.w_2.w_3.w_4.w_5) = P(w_5|w_1.w_2.w_3.w_4).P(w_4|w_1.w_2.w_3).P(w_3|w_1.w_2).P(w_2|w_1).P(w_1)$

Number of probability estimated needed for $P(w_1) = 30k$
Number of probability estimated needed for $P(w_2|w_1) = (30k)^2$
Similarly, number of estimates needed for $P(w_5|w_1.w_2.w_3.w_4) = (30k)^5$

$$\begin{aligned} Total probability estimates &= (30k) + (30k)^2 + (30k)^3 + (30k)^4 + (30k)^5 \\ &= \frac{30k((30k)^5 - 1)}{30k - 1} \\ &\approx \frac{(30k)^6 - 30k}{30k} \\ &= (30k)^5 - 1 \\ &\approx (30k)^5 \end{aligned}$$

Hence, approximately $2.43 * 10^{22}$ parameters have to be estimated.
The above results are for a 5-gram model.

If we use a uni-gram model, the estimated probability is $P(w1)$. Hence the number of probability estimates required = $30k$
Similarly for a bi-gram model, number of estimates = $30k + (30k)^2$
For a tri-gram model, number of estimates = $30k + (30k)^2 + (30k)^3$
For a 4-gram model, number of estimates = $30k + (30k)^2 + (30k)^3 + (30k)^4$

Ho Vinh Thinh (2562630)
Prabal Agarwal (2561419)
Nurzat Rakhmanberdieva (2561148)

Problem 4

Figure 1:

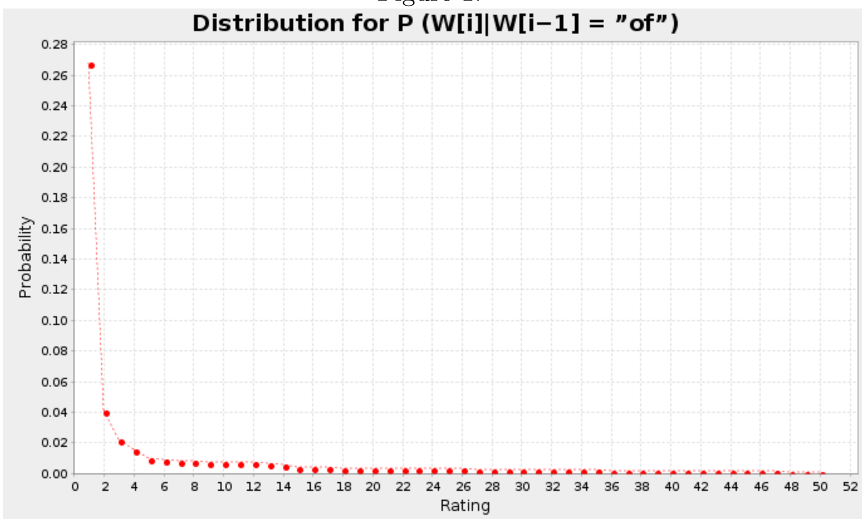
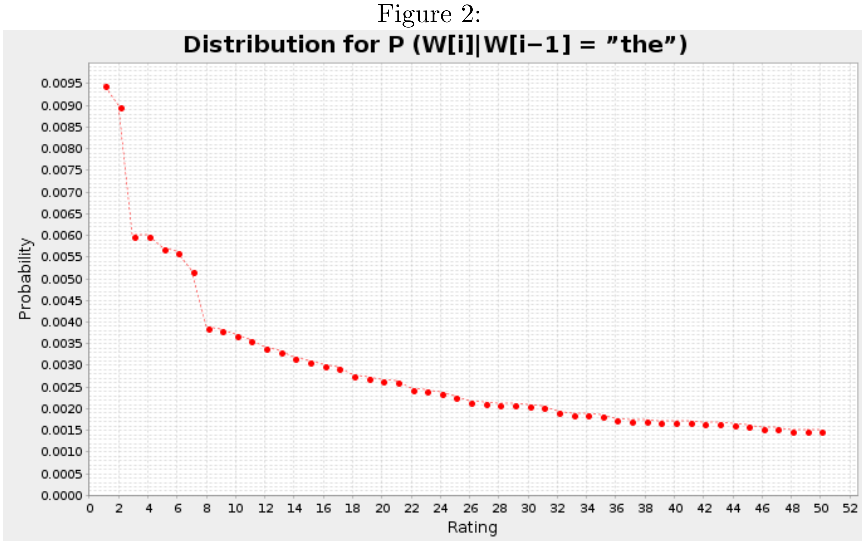


Figure 2:



The entropy is an indicator the uncertainty of the data. From these two graphs in Figure 1, we expect the first graph with term ”of” to have less entropy value. Because the difference in probability of the elements is very high, and we can expect that certain events happen with really large probability. On the other hand, the second graph displays more or less same probability for many variables, which indicates that its entropy will be high. It is more predictable than the first distribution.

The result from the computing the entropy for these given distributions
The value for distribution with the term ”of” is 8.6917
The value for distribution with the term ”the” is 12.0230
These values are calculated in logarithm base-2

Ho Vinh Thinh (2562630)
Prabal Agarwal (2561419)
Nurzat Rakhmanberdieva (2561148)

Instructions to run the the code for Linux:

```
tar -xzvf code.tar.gz
cd code
bash build
bash run [inputPath]
```

bash run takes 1 parameter:
[inputPath]: input file/folder contains the corpus.

To run the code with Brown corpus: `bash run ./brown`

The tool will print out the entropy for 2 distributions and then draw the 50 most frequent tokens for both distributions.

In addition, 2 file called "ofProb.txt" and "theProb.txt" is generated which contains all distribution of P(W[i], W[i-1]="of") and P(W[i], W[i-1]="the") with frequency and probability.