Assignment3_Part 2 - Code Breaking

Emily Zhu

4/1/2021

```
library('tinytex')
```

```
english.letters <- c('a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j', 'k',
                     'l', 'm', 'n', 'o', 'p', 'q', 'r', 's', 't', 'u', 'v',
                     'w', 'x', 'y', 'z')
```

```
apply.cipher.to.string <- function(string, cipher)
{
  output <- ''

  for (i in 1:nchar(string))
  {
    output <- paste(output, cipher[[substr(string, i, i)]], sep = '')
  }

  return(output)
}

apply.cipher.to.text <- function(text, cipher)
{
  output <- c()

  for (string in text)
  {
    output <- c(output, apply.cipher.to.string(string, cipher))
  }

  return(output)
}
```

```
generate.random.cipher <- function()
{
  cipher <- list()

  inputs <- english.letters
  outputs <- english.letters[sample(1:length(english.letters),
  length(english.letters))]

  for (index in 1:length(english.letters))
  {
    cipher[[inputs[index]]] <- outputs[index]
```

```
  }

  return(cipher)
}

modify.cipher <- function(cipher, input, output)
{
  new.cipher <- cipher
  new.cipher[[input]] <- output
  old.output <- cipher[[input]]
  collateral.input <- names(which(sapply(names(cipher),
                      function (key) {cipher[[key]]}) == output))
  new.cipher[[collateral.input]] <- old.output
  return(new.cipher)
}

propose.modified.cipher <- function(cipher)
{
  input <- sample(names(cipher), 1)
  output <- sample(english.letters, 1)
  return(modify.cipher(cipher, input, output))
}
```

calculate the probability of one word

```
one.gram.probability <- function(decrypted.string, lexical.database)
{
  string_split <- strsplit(decrypted.string, "")[[1]]
  word.probability <- 0.0

  for (i in seq_along(string_split) - 1) {
    if(length(string_split[i]) > 0) {
      if(match(string_split[i+1],english.letters, nomatch = 0) != 0) {
        word.probability <- word.probability + lexical.database[match(string_split[i],english.letters),
      }
    }
    else { #applies to one-letter words. assign a very small probability close to 0, like with pairwise
        word.probability <- word.probability + 0.1
    }
  }
  return(word.probability)
}
```

compute the log probability of the text

```
log.probability.of.text <- function(text, cipher, lexical.database)
{
  log.probability.text <- 0.0

  for (string in text)
  {
    decrypted.string <- apply.cipher.to.string(string, cipher)
    log.probability.text <- log.probability.text +
```

```
    log(one.gram.probability(decrypted.string, lexical.database))
  }


  return(log.probability.text)
}
```

```
metropolis.step <- function(text, cipher, lexical.database = list())
{
  proposed.cipher <- propose.modified.cipher(cipher)

  lp1 <- log.probability
  lp2 <- log.probability.of.text(text, proposed.cipher, lexical.database)

  if (lp2 > lp1)
  {
    return(proposed.cipher)
  }
  else
  {
    a <- exp(lp2 - lp1)
    x <- runif(1)
    if (x < a)
    {
      return(proposed.cipher)
    }
    else
    {
      return(cipher)
    }
  }
}
```

```
encrypted.text <- tolower(unlist(strsplit(readLines("/Users/ezhu/pm520_repos/assignment3/assignment3-zh
lexical.database <-read.table("/Users/ezhu/pm520_repos/assignment3/assignment3-zhuemi/LetterPairFreqFro
```

```
set.seed(1)
cipher <- generate.random.cipher()

results <- data.frame()

number.of.iterations <- 100

for (iteration in 1:number.of.iterations)
{
  log.probability <- log.probability.of.text(encrypted.text, cipher, lexical.database)
  current.decrypted.text <- paste(apply.cipher.to.text(encrypted.text, cipher),
                                  collapse = ' ')
  results <- rbind(results,
                   data.frame(Iteration = iteration,
                              LogProbability = log.probability,
                              CurrentDecryptedText = current.decrypted.text
                              ))
```

```
  cipher <- metropolis.step(encrypted.text, cipher, lexical.database)


}

write.table(results, file = "", row.names = FALSE, sep = '\t')
```

```
## "Iteration"  "LogProbability"    "CurrentDecryptedText"
## 1    897.99453768926 "nhi ogtkn ngei g urgs ifik pb nittf uibbpd hi lrk stmbq gb r tpuuktpfwi kguzit
## 2    897.99453768926 "nhi ogtkn ngei g urgs ifik pb nittf uibbpd hi lrk stmbq gb r tpuuktpfwi kguzit
## 3    916.103084312753    "nhi ogtkn ngei g urgs ifik cb nittf uibbcd hi lrk stmbq gb r tcuuktcfwi kgu
## 4    916.103084312753    "nhi ogtkn ngei g urgs ifik cb nittf uibbcd hi lrk stmbq gb r tcuuktcfwi kgu
## 5    959.144923937967    "nhi kgton ngei g urgs ifio cb nittf uibbcd hi lro stmbq gb r tcuuotcfwi ogu
## 6    998.788501263898    "nhi kgcon ngei g urgs ifio tb niccf uibbtd hi lro scmbq gb r ctuuoctfwi ogu
## 7    998.788501263898    "nhi kgcon ngei g urgs ifio tb niccf uibbtd hi lro scmbq gb r ctuuoctfwi ogu
## 8    998.788501263898    "nhi kgcon ngei g urgs ifio tb niccf uibbtd hi lro scmbq gb r ctuuoctfwi ogu
## 9    998.788501263898    "nhi kgcon ngei g urgs ifio tb niccf uibbtd hi lro scmbq gb r ctuuoctfwi ogu
## 10   998.788501263898    "nhi kgcon ngei g urgs ifio tb niccf uibbtd hi lro scmbq gb r ctuuoctfwi ogu
## 11   999.245832445957    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scmbq hb r ctuuoctfwi ohu
## 12   999.245832445957    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scmbq hb r ctuuoctfwi ohu
## 13   999.245832445957    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scmbq hb r ctuuoctfwi ohu
## 14   999.245832445957    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scmbq hb r ctuuoctfwi ohu
## 15   999.245832445957    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scmbq hb r ctuuoctfwi ohu
## 16   999.245832445957    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scmbq hb r ctuuoctfwi ohu
## 17   999.245832445957    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scmbq hb r ctuuoctfwi ohu
## 18   999.245832445957    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scmbq hb r ctuuoctfwi ohu
## 19   999.245832445957    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scmbq hb r ctuuoctfwi ohu
## 20   1029.3615380486 "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scabq hb r ctuuoctfwi ohuzic
## 21   1029.3615380486 "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scabq hb r ctuuoctfwi ohuzic
## 22   1031.32888221379    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scabq hb r ctuuoctfpi ohu
## 23   1031.32888221379    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scabq hb r ctuuoctfpi ohu
## 24   1035.14702015073    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scabv hb r ctuuoctfpi ohu
## 25   1035.14702015073    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scabv hb r ctuuoctfpi ohu
## 26   1035.14702015073    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scabv hb r ctuuoctfpi ohu
## 27   1035.14702015073    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scabv hb r ctuuoctfpi ohu
## 28   1035.14702015073    "ngi khcon nhei h urhs ifio tb niccf uibbtd gi lro scabv hb r ctuuoctfpi ohu
## 29   1048.86479280077    "ngi khcon nhei h uahs ifio tb niccf uibbtd gi lao scrbv hb a ctuuoctfpi ohu
## 30   1048.86479280077    "ngi khcon nhei h uahs ifio tb niccf uibbtd gi lao scrbv hb a ctuuoctfpi ohu
## 31   1048.86479280077    "ngi khcon nhei h uahs ifio tb niccf uibbtd gi lao scrbv hb a ctuuoctfpi ohu
## 32   1069.56435117746    "ngi khcon nhei h uahb ifio ts niccf uisstd gi lao bcrsv hs a ctuuoctfpi ohu
## 33   1069.56435117746    "ngi khcon nhei h uahb ifio ts niccf uisstd gi lao bcrsv hs a ctuuoctfpi ohu
## 34   1069.56435117746    "ngi khcon nhei h uahb ifio ts niccf uisstd gi lao bcrsv hs a ctuuoctfpi ohu
## 35   1070.24080438323    "ngi khcon nhei h uahb ifio ts niccf uisstd gi lao bcrsv hs a ctuuoctfpi ohu
## 36   1092.65463166871    "ngi khcon nhui h eahb ifio ts niccf eisstd gi lao bcrsv hs a cteeoctfpi ohe
## 37   1092.65463166871    "ngi khcon nhui h eahb ifio ts niccf eisstd gi lao bcrsv hs a cteeoctfpi ohe
## 38   1092.65463166871    "ngi khcon nhui h eahb ifio ts niccf eisstd gi lao bcrsv hs a cteeoctfpi ohe
## 39   1092.65463166871    "ngi khcon nhui h eahb ifio ts niccf eisstd gi lao bcrsv hs a cteeoctfpi ohe
## 40   1103.98807444099    "ngi hkcon nkui k eakb ifio ts niccf eisstd gi lao bcrsv ks a cteeoctfpi oke
## 41   1103.98807444099    "ngi hkcon nkui k eakb ifio ts niccf eisstd gi lao bcrsv ks a cteeoctfpi oke
## 42   1103.98807444099    "ngi hkcon nkui k eakb ifio ts niccf eisstd gi lao bcrsv ks a cteeoctfpi oke
## 43   1103.98807444099    "ngi hkcon nkui k eakb ifio ts niccf eisstd gi lao bcrsv ks a cteeoctfpi oke
## 44   1103.98807444099    "ngi hkcon nkui k eakb ifio ts niccf eisstd gi lao bcrsv ks a cteeoctfpi oke
## 45   1103.98807444099    "ngi hkcon nkui k eakb ifio ts niccf eisstd gi lao bcrsv ks a cteeoctfpi oke
```

```
## 46   1108.70612287664    "ngi hkcon nkui k eakl ifio ts niccf eisstd gi bao lcrsv ks a cteeoctfpi ok
## 47   1125.25518466762    "nge hkcon nkue k iakl efeo ts neccf iesstd ge bao lcrsv ks a ctiioctfpe ok
## 48   1125.25518466762    "nge hkcon nkue k iakl efeo ts neccf iesstd ge bao lcrsv ks a ctiioctfpe ok
## 49   1144.46619162635    "nge hkcon nkue k rakl efeo ts neccf resstd ge bao lcisv ks a ctrroctfpe ok
## 50   1144.46619162635    "nge hkcon nkue k rakl efeo ts neccf resstd ge bao lcisv ks a ctrroctfpe ok
## 51   1144.46619162635    "nge hkcon nkue k rakl efeo ts neccf resstd ge bao lcisv ks a ctrroctfpe ok
## 52   1144.46619162635    "nge hkcon nkue k rakl efeo ts neccf resstd ge bao lcisv ks a ctrroctfpe ok
## 53   1144.46619162635    "nge hkcon nkue k rakl efeo ts neccf resstd ge bao lcisv ks a ctrroctfpe ok
## 54   1155.01335182728    "nge hkcon nkue k rakl efeo ts neccf resstd ge mao lcisv ks a ctrroctfpe ok
## 55   1155.01335182728    "nge hkcon nkue k rakl efeo ts neccf resstd ge mao lcisv ks a ctrroctfpe ok
## 56   1155.01335182728    "nge hkcon nkue k rakl efeo ts neccf resstd ge mao lcisv ks a ctrroctfpe ok
## 57   1155.01335182728    "nge hkcon nkue k rakl efeo ts neccf resstd ge mao lcisv ks a ctrroctfpe ok
## 58   1159.74286645959    "nge hkcon nkue k rakl efeo ts neccf resstb ge mao lcisv ks a ctrroctfpe ok
## 59   1159.74286645959    "nge hkcon nkue k rakl efeo ts neccf resstb ge mao lcisv ks a ctrroctfpe ok
## 60   1159.74286645959    "nge hkcon nkue k rakl efeo ts neccf resstb ge mao lcisv ks a ctrroctfpe ok
## 61   1159.74286645959    "nge hkcon nkue k rakl efeo ts neccf resstb ge mao lcisv ks a ctrroctfpe ok
## 62   1159.74286645959    "nge hkcon nkue k rakl efeo ts neccf resstb ge mao lcisv ks a ctrroctfpe ok
## 63   1159.74286645959    "nge hkcon nkue k rakl efeo ts neccf resstb ge mao lcisv ks a ctrroctfpe ok
## 64   1159.74286645959    "nge hkcon nkue k rakl efeo ts neccf resstb ge mao lcisv ks a ctrroctfpe ok
## 65   1159.74286645959    "nge hkcon nkue k rakl efeo ts neccf resstb ge mao lcisv ks a ctrroctfpe ok
## 66   1159.74286645959    "nge hkcon nkue k rakl efeo ts neccf resstb ge mao lcisv ks a ctrroctfpe ok
## 67   1159.74286645959    "nge hkcon nkue k rakl efeo ts neccf resstb ge mao lcisv ks a ctrroctfpe ok
## 68   1159.74286645959    "nge hkcon nkue k rakl efeo ts neccf resstb ge mao lcisv ks a ctrroctfpe ok
## 69   1159.74286645959    "nge hkcon nkue k rakl efeo ts neccf resstb ge mao lcisv ks a ctrroctfpe ok
## 70   1159.87057065472    "nge hkyon nkue k rakl efeo ts neyyf resstb ge mao lyisv ks a ytrroytfpe ok
## 71   1159.87057065472    "nge hkyon nkue k rakl efeo ts neyyf resstb ge mao lyisv ks a ytrroytfpe ok
## 72   1159.87057065472    "nge hkyon nkue k rakl efeo ts neyyf resstb ge mao lyisv ks a ytrroytfpe ok
## 73   1159.87057065472    "nge hkyon nkue k rakl efeo ts neyyf resstb ge mao lyisv ks a ytrroytfpe ok
## 74   1160.33365902979    "nge hkuon nkye k rakl efeo ts neuuf resstb ge mao luisv ks a utrroutfpe ok
## 75   1160.33365902979    "nge hkuon nkye k rakl efeo ts neuuf resstb ge mao luisv ks a utrroutfpe ok
## 76   1160.18168512492    "nge hkpon nkye k rakl efeo ts neppf resstb ge mao lpisv ks a ptrroptfue ok
## 77   1159.72607795779    "nge hkpon nkye k rakl efeo ts neppf resstb ge mao lpisv ks a ptrroptfce ok
## 78   1159.72607795779    "nge hkpon nkye k rakl efeo ts neppf resstb ge mao lpisv ks a ptrroptfce ok
## 79   1159.72607795779    "nge hkpon nkye k rakl efeo ts neppf resstb ge mao lpisv ks a ptrroptfce ok
## 80   1159.72607795779    "nge hkpon nkye k rakl efeo ts neppf resstb ge mao lpisv ks a ptrroptfce ok
## 81   1158.72268962563    "nge hkmon nkye k rakl efeo ts nemmf resstb ge pao lmisv ks a mtrromtfce ok
## 82   1158.72268962563    "nge hkmon nkye k rakl efeo ts nemmf resstb ge pao lmisv ks a mtrromtfce ok
## 83   1158.72268962563    "nge hkmon nkye k rakl efeo ts nemmf resstb ge pao lmisv ks a mtrromtfce ok
## 84   1158.72268962563    "nge hkmon nkye k rakl efeo ts nemmf resstb ge pao lmisv ks a mtrromtfce ok
## 85   1158.72268962563    "nge hkmon nkye k rakl efeo ts nemmf resstb ge pao lmisv ks a mtrromtfce ok
## 86   1158.72268962563    "nge hkmon nkye k rakl efeo ts nemmf resstb ge pao lmisv ks a mtrromtfce ok
## 87   1158.72268962563    "nge hkmon nkye k rakl efeo ts nemmf resstb ge pao lmisv ks a mtrromtfce ok
## 88   1158.72268962563    "nge hkmon nkye k rakl efeo ts nemmf resstb ge pao lmisv ks a mtrromtfce ok
## 89   1158.85995049766    "nge hkmon nkye k rakl eweo ts nemmw resstb ge pao lmisv ks a mtrromtwce ok
## 90   1165.66141745881    "nge hkmdn nkye k rakl ewed ts nemmw resstb ge pad lmisv ks a mtrrdmtwce dk
## 91   1165.66141745881    "nge hkmdn nkye k rakl ewed ts nemmw resstb ge pad lmisv ks a mtrrdmtwce dk
## 92   1165.66141745881    "nge hkmdn nkye k rakl ewed ts nemmw resstb ge pad lmisv ks a mtrrdmtwce dk
## 93   1165.66141745881    "nge hkmdn nkye k rakl ewed ts nemmw resstb ge pad lmisv ks a mtrrdmtwce dk
## 94   1165.66141745881    "nge hkmdn nkye k rakl ewed ts nemmw resstb ge pad lmisv ks a mtrrdmtwce dk
## 95   1165.66141745881    "nge hkmdn nkye k rakl ewed ts nemmw resstb ge pad lmisv ks a mtrrdmtwce dk
## 96   1165.66141745881    "nge hkmdn nkye k rakl ewed ts nemmw resstb ge pad lmisv ks a mtrrdmtwce dk
## 97   1165.66141745881    "nge hkmdn nkye k rakl ewed ts nemmw resstb ge pad lmisv ks a mtrrdmtwce dk
## 98   1165.66141745881    "nge hkmdn nkye k rakl ewed ts nemmw resstb ge pad lmisv ks a mtrrdmtwce dk
## 99   1165.67392593242    "nge hkmdn nkye k rakl ewed ts nemmw resstz ge pad lmisv ks a mtrrdmtwce dk
```

```
## 100  1165.67392593242    "nge hkmdn nkye k rakl ewed ts nemmw resstz ge pad lmisv ks a mtrrdmtwce dk
```

For this part of the assignment, the Metropolis method is implemented as a way to decode messages. Using the table of pairwise frequencies, we can calculate a score for each word by adding up all of the pairwise frequencies that occur, and ultimately doing so for the entire text. This is done so different decryption rules can be applied to the body of text and then compared to each other based on its score.

The log of the text's score/probability is taken in order to prevent numeric instability that would occur from summing the raw probabilities. The instability occurs from finite precision arithmetic in floating point numbers.

In order to prevent greedy optimization, the algorithm does not always accept the decryption rule on the basis of increasing the probability of the decrypted text. Even if the new proposed rule is not as "great" as the original, we will still accept the new rule regardless... sometimes. If the score of the new proposed rule is greater than that of the original, then we replace the old rule with the new rule. If the score of the new proposed rule is greater than that of the original, the new rule will still replace the old rule prob(new rule) / prob(old rule) percent of the time.

Simulated annealing would probably yield better results by making the algorithm greedier the longer it runs by accepting non-greedy proposals less often.

Here is the output after 4654 iterations: "the first time i laid eyes on terry lennoq he was drunk in a rollsroyce silver wraith outside the terrace of the dancers the parking lot attendant had brought the car out and he was still holding the door open because terry lennoqs left foot was still dangling outside as if he had forgotten he had one he had a younglooking face but his hair was bone white you could tell by his eyes that he was plastered to the hairline but otherwise he looked like any other nice young guy in a dinner jacket who had been spending too much money in a joint that eqists for that purpose and for no other there was a girl beside him her hair was a lovely shade of dark red and she had a distant smile on her lips and over her shoulders she had a blue mink that almost made the rollsroyce look like just another automobile it didnt zuite nothing can the attendant was the usual halftough character in a white coat with the name of the restaurant stitched across the front of it in red he was getting fed up look mister he said with an edge to his voice would you mind a whole lot pulling your leg into the car so i can kind of shut the door or should i open it all the way so you can fall out the girl gave him a look which ought to have stuck at least four inches out of his back it didnt bother him enough to give him the shakes at the dancers they get the sort of people that disillusion you about what a lot of golfing money can do for the personality"

The results may be better if more iterations were performed, but my computer probably can't handle too many without crashing.