LING82100 - homework 2

(due 3/4)

1 Reporting a binomial test

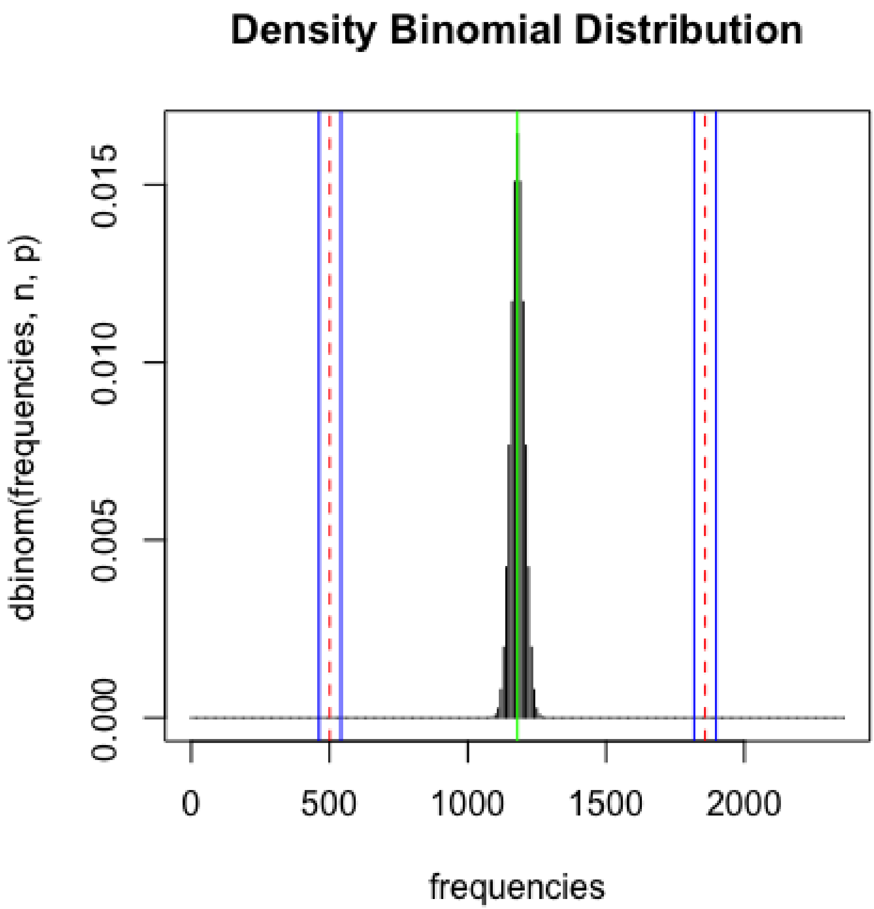
Background: As is well known, Mainstream English has two competing forms of the dative. A binomial test can be performed to determine the extent to which a speaker is equally likely to use one form over the other–that is, we will assume, as a null hypothesis that the two constructions are equiprobable. This test will be performed on a sample of data provided by Bresnan et al. (2007), who have collected the counts of these two constructions from the corpus of American English spontaneous phone conversations, where the count of prepositional datives is 501, and that of double objects is 1,859.

Results of the binomial test: The results of the binomial test reveal that the probability that a sample of data will yield a test statistic of either 501/(501+1859) or 1859/(501+1859) within a binomial distribution is 2.2E-16. The confidence intervals for both test statistics are [0.1959 to 0.2293] and [0.7706 to 0.8040], respectively. Given that the probability 2.2E-16 is much lower than $\alpha = .05$ and that neither confidence interval overlaps with the mean of the binomial distribution, the likelihood that the dative and double object constructions are equiprobable is not statistically significant.

Stretch Goal: The figure below shows that the confidence intervals of the two aforementioned test statistics do not overlap with the mean of this data's binomial distribution. The two red lines represent the test statistics for x=501 and x=1859, respectively, and the blue lines around them demarcate the confidence intervals. The mean of the distribution is in green. The R commands used to generate the table are as follows:

```
> n< −(501+1859)
> p< −.5
> frequencies < − seq(0, (501+1859), by=10)
> probabilities < − dbinom(frequencies, n, p)
> plot(frequencies, probabilities, type=”h”, xlab=”frequency of success”, ylab=”frequency of the mean of success”, main=””)
> abline(v=1859, lty=2, col=”red”)
> abline(v=501, lty=2, col=”red”)

% adds the confidence intervals for mean=501 and mean=1859, respectively
> abline(v=(n*0.2293504), lty=1, col=”blue”)
> abline(v=(n*0.1959431), lty=1, col=”blue”)
> abline(v=(n* 0.8040569), lty=1, col=”blue”)
> abline(v=(n* 0.7706496 ), lty=1, col=”blue”)
```



2 McNemar's Test

- The number of wins for the Stanford tagger is 943.
- The number of wins for the NLP4J tagger is 1016.
- Although the NLP4J tagger has slightly more wins than the Stanford tagger, it's not considered significantly better if we decide to set $\alpha = .05$. Reason being that it's probability of success at .5186 falls within the 95% confidence interval of the binomial distribution at .1038, which is greater than α.
- R commands:

  %loads the tsv file into R as a table.
  > df < − read.table("/Users/mariagarza/Desktop/Statistics/PTB.tsv", header=TRUE, comment.char="")

  % converts the values for Stanford and NLP4J POS tags to booleans.
  > Stanford.correct < − df$gold.tag == df$Stanford.tag
  > NLP4J.correct < − df$gold.tag == df$NLP4J.tag

  % computes the number of "wins" wins for Stanford over NLP4J and assigns it to x1, and vice-versa.
  > x1 < − sum(Stanford.correct & !NLP4J.correct)

```
[1] 943
> x2 <- sum(NLP4J.correct & !Stanford.correct)
[1] 1016

%runs McNemar's Binomial Test
> binom.test(min(x1, x2), (x1+x2), .5)
```

```
        Exact binomial test

data:  min(x1, x2) and (x1 + x2)
number of successes = 943, number of trials = 1959, p-value =
0.1038
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.459029 0.503763
sample estimates:
probability of success
              0.481368
```

.