Two Sample Mood Test

Ryan Honea

10/11/2017

Calculating the M_n Statistic

The Two Sample Mood Test can be used to determine whether or not a variable is widely dispersed. So in this case,

$$X \quad X \quad X \quad Y \quad Y \quad Y \quad Y \quad X \quad X \quad X$$

it seems clear that X is far more dispersed than Y as opposed to the following case

$$X \quad Y \quad X \quad Y \quad X \quad Y \quad X \quad Y$$

where the observations seem to be of similar scale.

So, the two-sample Mood Test was devised to test whether or not X (or Y) has large scale compared to its counterpart.

The mood statistic is

$$M_n = \sum_{i=1}^{N} \left(i - \frac{N-1}{2} \right)^2 Z_i$$

where high values of M_n would lead us to reject the null hypothesis that the scale of X and Y are equal. The code below will do just that.

```
mood_statistic <- function(X, Y) {
  X <- cbind(X, rep(1, length(X)))
  Y <- cbind(Y, rep(0, length(Y)))
  seq <- rbind(X, Y)
  seq <- seq[order(seq[,1]),]
  N <- length(seq[,2])
  sum <- 0
  for (i in 1:N) {
    sum <- sum + (i-(N+1)/2)^2 * seq[i,2]
  }
  return(sum)
}
```

With increasing m and n observations, this test will clearly lose power which is evident by the increase of the expected value for increasing m and n.

Below is a plot of M_n for distributions where the null hypothesis should be rejected and for when it should not be. The distributions where it should not be rejected are simulations from two standard normals. The distribution where it should will be a normal distribution of mean 0 and standard deviation 4 versus the standard normal.

```
N <- seq(2,50, by = 2)
Mn <- (N/2)*((N/2)^2 - 1)/12
Mn0 <- c()
Mna <- c()
for (i in 1:length(N)) {
  sim_0 <- rnorm(N[i])
  sim_a <- rnorm(N[i],0,3)
  Mn0[i] <- mood_statistic(sim_0,sim_0)
  Mna[i] <- mood_statistic(sim_a, sim_0)
}
plot(N, Mn, main = "Expected Value of Mn versus N", type = "l")
lines(N, Mn0, col = "red")
lines(N, Mna, col = "blue")
```

Expected Value of Mn versus N
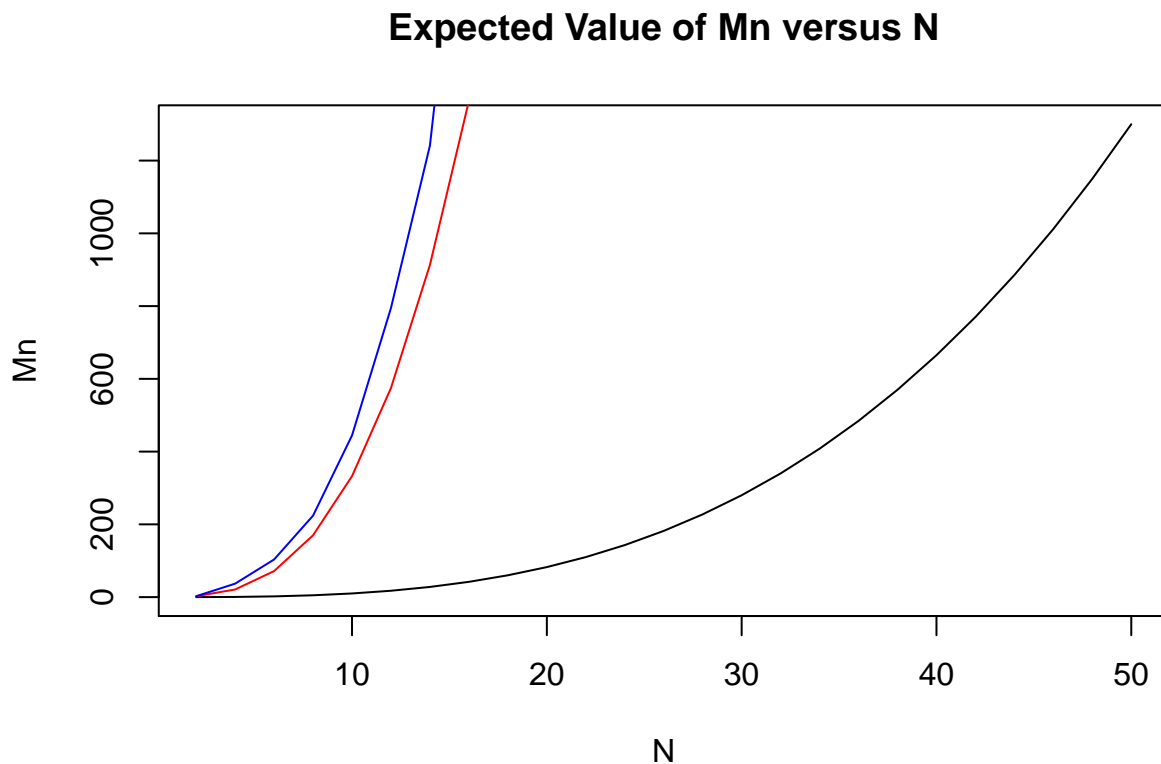
And so even though the different distributions do have higher values of M_n, a distriution where the scale is the same is still resulting in high values. Luckily, the `mood.test` command in R is able to do the heavy computation in p-value calculation.

```
sim_0 <- rnorm(25)
sim_a <- rnorm(25,0,2)
mood.test(sim_0, sim_0)
```

```
##
##  Mood two-sample test of scale
##
## data:  sim_0 and sim_0
## Z = 0, p-value = 1
## alternative hypothesis: two.sided
```

```
mood.test(sim_0, sim_a)
```

```
## 
##  Mood two-sample test of scale
## 
## data:  sim_0 and sim_a
## Z = -3.8867, p-value = 0.0001016
## alternative hypothesis: two.sided
```