Goodness-of-fit Tests for the Normal Distribution
Project 1

Jeremy Morris

September 29, 2005

1 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov Test (KS test) is based on the cumulative distribution function of the underlying distribution. The graphic in Figure (1) shows the CDF of the Normal Distribution and an observed CDF or Empirical CDF of normally distributed random numbers. The KS test is based on the maximum difference between these two CDF's. To

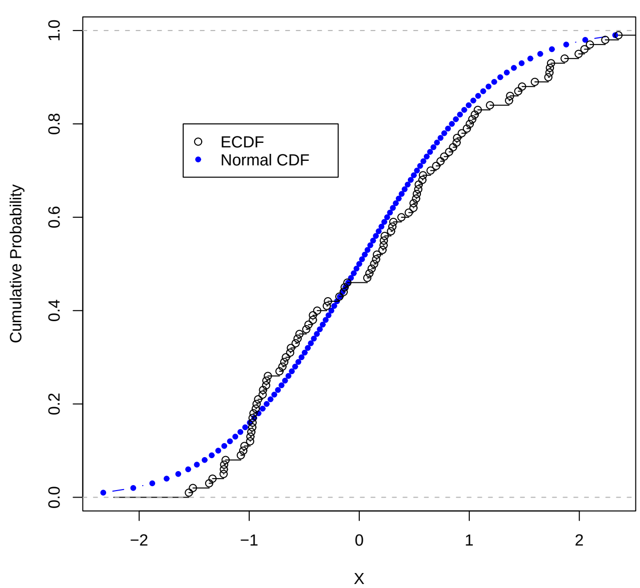

Figure 1: Comparison of ECDF and CDF of Normal Distribution

test the hypothesis $H_0 : X \sim F(x)$, where F is a fully specified CDF, we use the statistics

$$\begin{aligned} D^+ &= \max_i (i/n - F(x_{i:n})) && (1) \\ D^- &= \min_i (F(x_{i:n}) - (i-1)/n) && (2) \\ D &= \max(D^+, D^-) && (3) \end{aligned}$$

Where $x_{i:n}$ denotes the i^{th} order statistic of the random sample and $F(x_{i:n}) = P\{X < i\}$ for the distribution that is being fit (in our case, the standard normal distribution).

The distribution of (1) - (3) will not change if location-scale parameters are estimated, only the form of F will change in this case. Critical values for the D statistic have been calculated by Stephens[1].

1.1 Calculations / Simulations

R includes a function to run the KS test. It is called `ks.test`. The `ks.test` function takes two parameters, a data set and a character string naming which distribution we are testing. In our case, we will call `ks.test` with the parameter `'pnorm'`. `pnorm` is the normal density function provided by R.

We will examine some of the properties of the power function. This will be done by considering some data with the form

$$Z(\delta) = X + \delta Y \quad (4)$$

Where $X \sim N(0,1)$, $Y \sim U(-1,1)$ and δ is a value we use to skew the data. The first thing we look at is how the KS test reacts to changes in δ. Figure (2) shows a plot of the observed α level against δ for a few chosen δ values.

We see that at $\delta = 0$, we get a significance level of $\alpha = 0.05$ as to be expected. As δ increases, the observed significance level increases very rapidly. This demonstrates that the KS test recognizes small alterations to our data, and rejects more often. This also indicates that the KS test is fairly conservative, since it detects small changes in the data set.

Next, we fix a small number of δ values and change the sample size n in order to gauge what the power function does when the sample size is changed. In this case we take $N = 2500$ samples of size $n = 10, 20, \ldots, 100$ and plot the observed rejection rate against n. The results can be seen in Figure (3).

The test in this case is set up such that

$$H_0 : X \sim N(0,1) \qquad\qquad H_a : X \sim N(0,1) + \delta U(-1,1) \quad (5)$$

And we are concerned with the proportion of times we reject H_0 given that we have engineered it to be false. In this case we want this proportion to be greater than or equal to 0.95 or $1 - \alpha = 0.95$ and we would like to know what sample size is needed to acheive these requirements.

Figure (3) shows the percentage of times the test rejected as a function of sample size for various values of δ. It also shows a black line where our cut off is. This graphic shows that for very contaminated data, we need a very small sample size. This leads to the conclusion that the KS test has good power.

[1] "EDF Statistics for Goodness of Fit and Some Comparisons" *J. Amer. Statist. Assoc.* **69**, p.730

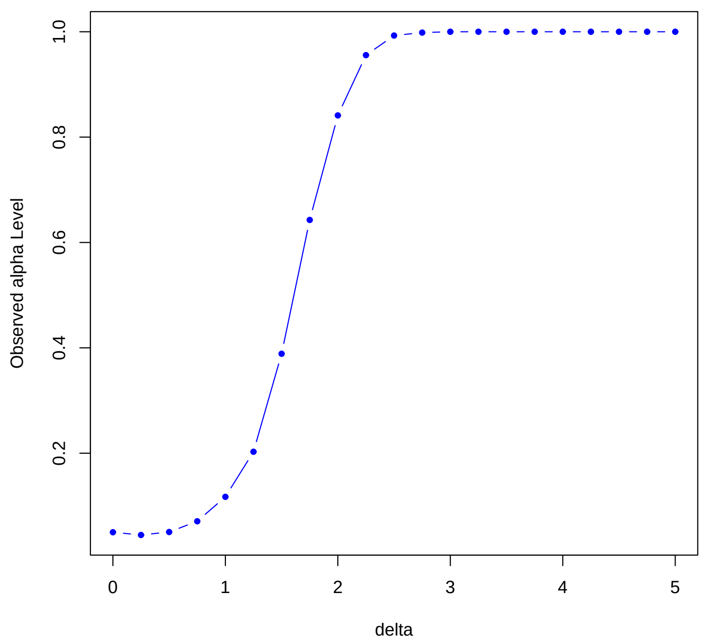

Figure 2: Power as a function of δ (KS test)

2 Cramer-von Mises Test

The Cramer-von Mises (CVM) test is also based on the CDF of the hypothesized distribution. In the case of the CVM test we use the statistic

$$CM = \frac{1}{12n} + \sum_{i=1}^{n} \left(F(x_{i:n}) - \frac{i - 0.5}{n} \right)^2 \tag{6}$$

The CM statistic takes into account all of the ordered data points and uses a half point correction. It can also be said that under H_0, the distribution of F is uniform on $(0, 1)$. This fact could be exploited to make calculations easier.The critical values for CM have been calculated by Pettitt and Stephans (1976)[2].

2.1 Calculations

`R` provides a function called `cvm.test` in the `nortest` package. This function performs the CVM test and tests for normality. It takes a dataset as its argument and returns a p-value. We will reject for p-values $< \alpha = 0.05$.

Again, our first simulations involve using the definition found in (4) to skew the data in order to simulate the power function. As in Figure (2) we will plot the observed rejection

[2] "Modified Cramer-von Mises Statistics for Censored Data" `Biometrika`. **63**, p. 291

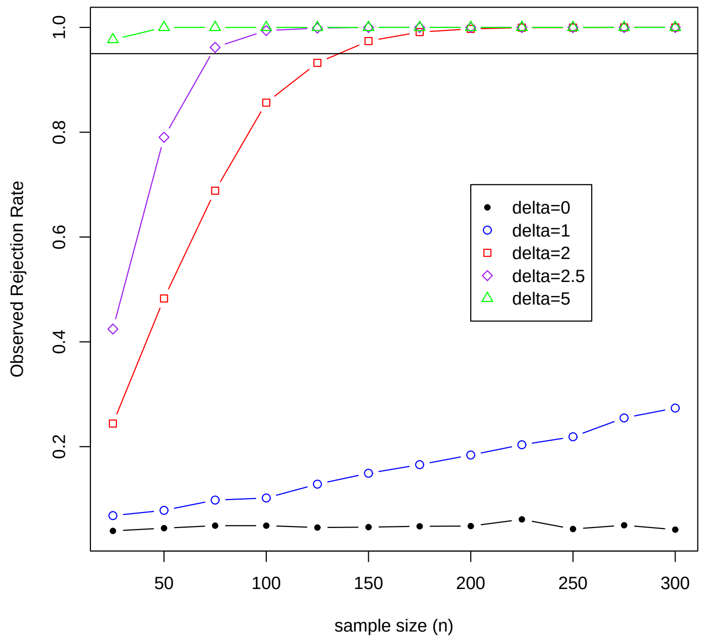

Figure 3: Power as a function of sample size (KS test)

rate as a function of δ.

Figure (4) shows that the CVM test does not recognize that the data has been contaminated very well. When we have $\delta = 5$ the observed rejection rate is only at 0.5, this does not compare well to the KS test.

When we look at the power analysis based on sample size, Figure (5), we see that in order to get good results we need a sample size of 150 or more. This also indicates that the CVM test does not perform as well as the KS test.

3 Fisher's Permutation Test

For the resampling portion of this project, I have chosen to use the Fisher permutation (FP) test. For the purposes of explination, assume we have two random samples with distributions F, G. The Fisher permutation test has the form

$$H_0 : F = G \qquad\qquad H_A : F \neq G \tag{7}$$

Assuming that we have n observations from F and m observations from G, the test uses the following algorithm

1. Choose a statistic and calculate on the original data, call it S.

2. Take a random sample (without replacement) of size n from all $n + m$ observations.

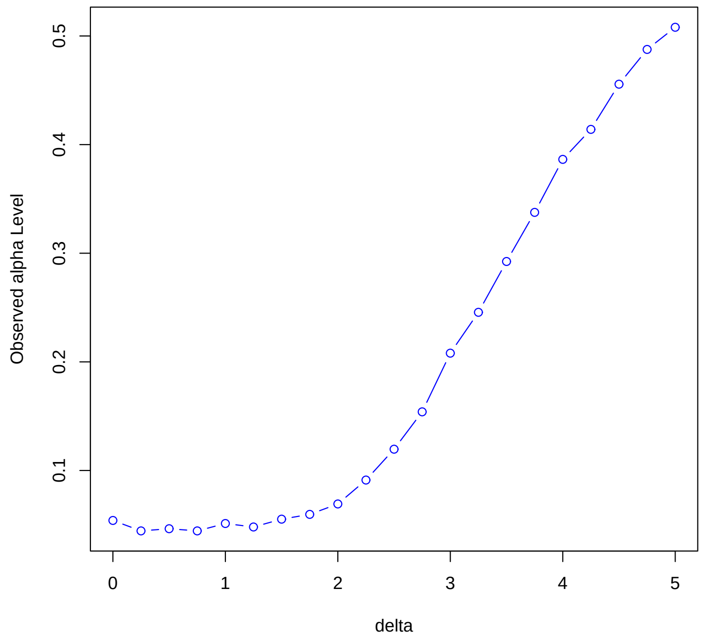

Figure 4: Power as a function of δ (CVM test)

3. Let the first n observations be in F^* and the remaining m observations G^*.
4. Calculate the statistic using F^* and G^*. Save this new statistic.
5. Repeat steps (2) - (4) many times.
6. Reject H_0 if S is in the 0.05 percentile.

The typical statistic to choose for this test is $S = \overline{x} - \overline{y}$, this is the statistic chosen in the following simulations.

3.1 Simulations

Code for this test can be found in Figure (6). The function to perform the FP test is called `test.permute` it takes a matrix containing the data that is to be tested for normality. The function `permute` performs the permutation tests. It follows the procedures outlined in the discussion of the FP test. Since we are testing for normality, we set y to evenly spaced values coming from the density of a standard normal random variable. These are computed using `R`'s `qnorm` function.

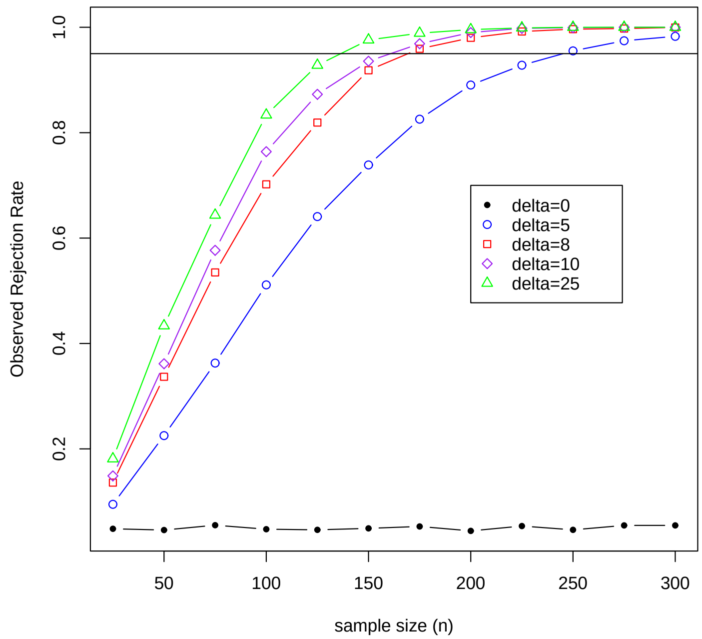

Figure 5: Power as a function of sample size (CVM test)

3.2 Results

We again perform the same power analysis as in the case of the KS and CVM tests. The graph in Figure (7) shows the power as a function of δ, and Figure (8) shows power as a function of sample size for several values of δ. Both figures seem to indicate that the FP test does not test very well for normality. To begin with, the test should look like some kind of increasing function as δ grows and the graph seems to indicate that this is not the case. Also, the test should always reject more as the sample size increases, Figure (8) indicates that the test acutally rejects less as n grows. Perhaps if a different statistic were chosen this test would have better power, the difficulty is that the FP test is very computationally intense and the choice of a different statistic would increase the processing time such that the test may not be as effective as a different method that relies on pre-calculated critical values.

```
#a permutation test
test.permute <- function(x,alpha=0.05,...){
    stats <- apply(x,2,resample,...)
    return( mean(stats <= alpha) )
}

resample <- function(x,s=10){
    n <- length( x )
    smpl <- numeric( n * s )
    y <- qnorm( (1:(199))/200 )
    S <- my.stat(x,y)
    z <- c( x, y )
    #now comes the resampling/permutation part
    for( i in seq( n * s ) ){
       rks <- sample( seq(along=z), n )
       x.star <- z[rks]
       y.star <- z[-rks]
       S.star <- my.stat(x.star,y.star)
       smpl[i] <- S.star
    }

    return( mean( smpl >= S ) )
}

my.stat <- function(x,y){
    return( mean(x) - mean(y) )
}
```

Figure 6: Code for Permutation Test

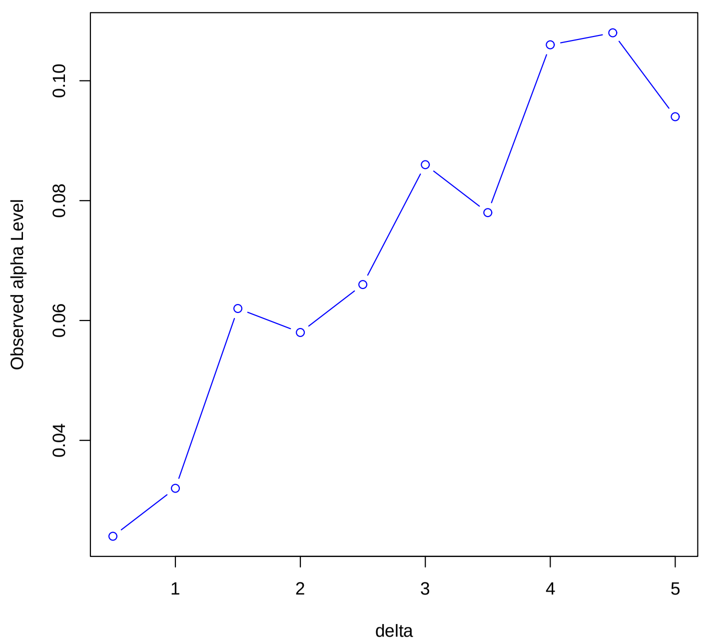

Figure 7: Power as a function of δ (FP test)

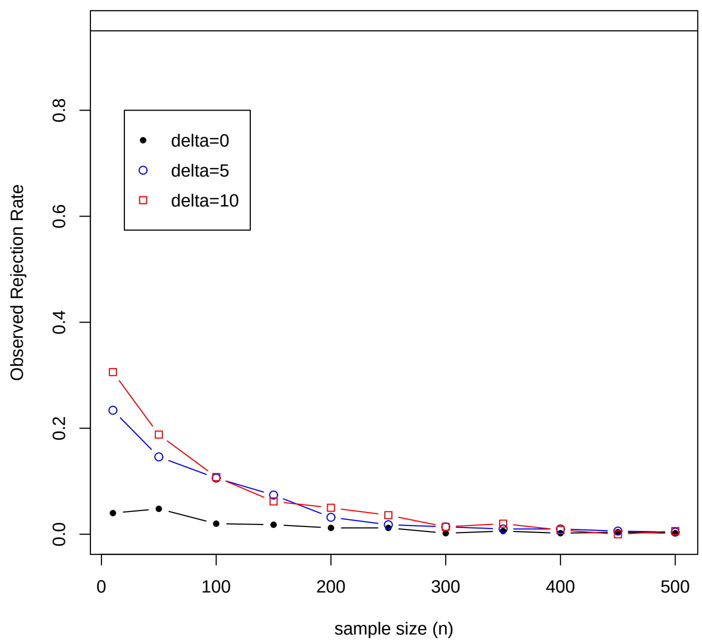

Figure 8: Power as a function of sample size (FP test)