Math 3080 Project 2018

Change Point Detection : It's the economy, stupid

A time series is a sequence of data points indexed by time: X1,X2,...,XT . Suppose the data in your time series follows a normal distribution, but, at some time k, the underlying mean possibly changes: Xt = mu + epsilon_t, t = 1, ... , k mu* + epsilon_t, t = k+1, ... , T where epsilon_t is iid ~ N(0, sigma^2).

We can interpret this as a one factor ANOVA problem with two treatments: data sampled before the change and after the change, and can test H_0 : mu = mu* against H_A,k : mu != mu* and calculate a test statistic Fk ~H0 F(n,m), if we know k in advance. (This requires performing ANOVA on treatments with unequal sample sizes, which is not too difficult; consult section 10.3 of the text for reference.) In reality, we may not have a guess as to when the change occurred. So instead, we perform the test for all k = 2, ... , T-2 (why not k = 1,...,T?) and many resulting test statistics, fk. The larger the test statistic is, the more contradictory it is towards our H0, so to find the most contradictory one we take the maximum: lambda = T-2_max_k=2 Fk

However, the difficulty here lies in the fact that we do not know the distribution of lambda; we can't just look it up to obtain the critical values! Instead, we must use Monte Carlo simulation to estimate the value c so that P(lambda > c jH0) = alpha.

Generate a time series (random sample) of T N(0; 1) (you know H0 holds here!) random variables, calculate the lambda statistic, repeat this many times, and find out how large the largest alpha(100%) lambdas are to determine your critical values.

Apply this test to US 10 year treasury note yields from each month for the past 20 years to find any possible change points. (Find this data here.) Calculate a P-value. When does the changepoint occur? After identifying the first changepoint, you can split the data across it, and see if you can find further changepoints across the segmented pieces. Attempt to verify all necessary assumptions. (Hint: Look at increments the Xt+1 - Xt to check assumptions about the errors. Why should this (mostly) work?)