Alexander Powell

DecisionQ Coding Exercise

02.04.2017

1. Web Scraper Task
   1. The scraper was written using the BeautifulSoup and urllib2 libraries. BeautifulSoup was very easy to use to parse the html and extract the desired string data. Because the assignment was to only scrape 6 years’ worth of movie data I added all the URLs directly into a list where each was scraped individually. One potential improvement would be to make this step more automatic for cases where hundreds of webpages need to be processed.

I also ran into a problem where one movie, “Waiting for ‘*Superman*’”, contained <i></i> tags. At first this tripped up the parser so an exception handler was added for this one movie. If these kind of movie titles were more common the scraper should probably be adjusted to handle this better.

Finally, most of the pages contained exactly 100 movies except for the final pages for every year (since the number of movies made in a year was never divisible by 100). The easiest way to account for this was to just check if the text in a row started with the words “Summary of” since that’s the message listed at the bottom of every movie table. This works fine for our purposes but could cause trouble if a movie was to be made whose title started with “Summary of”.

1. Numerical Algorithms Task
   1. See task2.py. Written in Python2.7 and uses the math library.
2. Unit Test Task
   1. Expected Value

The python function E() does not seem to satisfy the conditions to be a linear operator. In fact, none of the three conditions were satisfied. The imported ndarrays red and blue were used for testing purposes as well as random integers for constants like *a* and *c*. The function can be used as a generic moment generating function because E(X), E(X2), E(X3), etc. are all moments and E() is written so that X can be raised to any desired power.

Note: since E did not satisfy the linearity conditions I set the unit tests to assertFalse() to prevent all the error messages.

* 1. Variance

The python function var() does not satisfy the given theorem. Random integers were used to test the theorem.

Because the constant value, *b*, does not appear on the right side of the theorem, an example can easily be constructed to show that the property does not hold when *a* equals zero and *b* is a value large enough so that the variance of an ndarray does not equal zero.