MPCS58020 Assignment 2

Yecheng Li

Random permutations

1. E[X] = 50 * 1/50 = 1
   Var[X] = E[X^2] – E[X]^2 = 1
2. See attached file **Q1.m**

Exponential random variables

The probability is roughly 0.1065. For details, see attached file Q2.m

Poisson process

The arrival of fans is roughly 180. For details, see attached file Q3.m

PDF & CDF

h(x) is roughly twice as efficient as compare to g(x) in terms of average time and iterations per random number accepted. For details, see attached file Q4.m

(Problem 2.1)

[See attached file Q2_1.m]

1. In this linear regression model, β, α_1, α_2, α_3, and α_4 are constant regression coefficients. Parameter β models how Johnson & Johnson's earning per share changes with respect to time (in terms of quarter number); parameters α_1, α_2, α_3, and α_4 models if there is "drift", which is a constant and unrelated to which year it is, of the share price in quarter 1, 2, 3, or 4.
2. If an intercept term is included, that means a constant β_2 is added to the model. This is redundant, and will not affect the result. The fitted regression model will still be the same. This is demonstrated in the Q2_1.m
3. In the plot, it is observed that the plot of residuals fluctuate around y = 0 with no regular pattern. In other words, the residuals do not look white. Thus the model fits the data reasonably well.

(Problem 2.2)

[See attached file Q2_1_1.m and Q2_2_2.m]

1. According to the plot, it is observed that with or without adding P_{t-4} to the model does not make much difference. The plots of the two models have little difference that is barely observable, and the mean residuals of the two models are the same. Thus whether to include P_{t-4} in the model is insignificant.

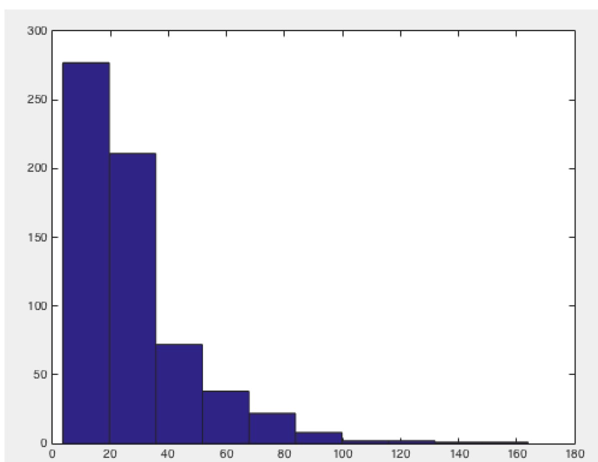
2. According to calculation, the correlation between M and P_{t-4} is greater than the correlation between M and P, though the difference is not significant. For details, see script Q2_2_2.m

(Problem 2.3)

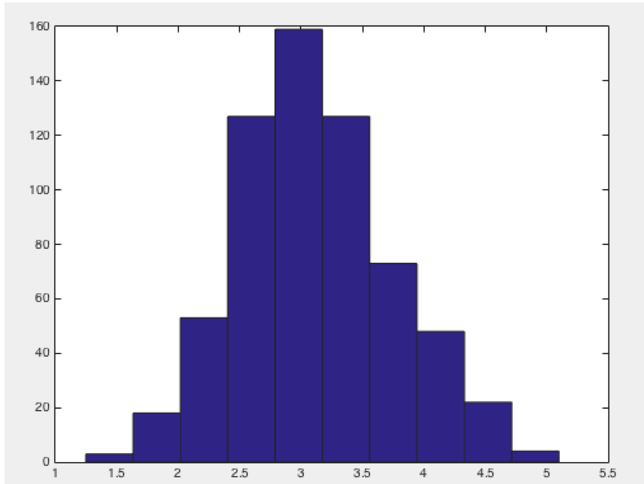
1. As the number of iterations increase, the average gradient of the fitted lines are getting close to the drift 0.01. That is, the mean function is tend to be u = 0.01t. For details, see Q2_3.m

(Problem 2.8)

1. Sample variance of the first half of the data is 133.45742, and the variance of the second half of the data is 594.49044. This indicates that the data exhibits hetero-scedasticity. According to the plot in Q2_8.m, the variance is visually more stable than the original data. This transformation helps stabilize the variance as when $\log(x_t)>\log(x_{t-1})$, it is likely that $x_t >> x_{t-1}$. That is, by taking log transformation, the variability of large values of data are reduced dramatically, thus reducing the overall data variability, and stabilized the data. As the calculation suggests, after log transformation, the variance of the first half of the data becomes 0.27071, and the variance of the second half of the data is 0.45137. This supports our observation that log transformation drastically stabilized the data.
2. It is observed that the histogram of x_t does not have any obvious pattern, while the histogram of y_t is close to normal distribution. The pattern of y_t makes sense as either very large or very small data are rare in the data sample. It is also observed that in plot 2, y's value is between 1 and 5, while x's range is between 0 and 160. This indicates some stability of variance.

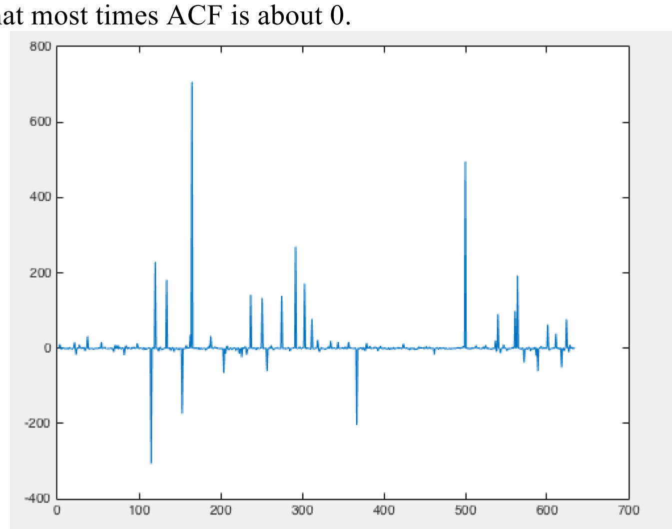


Plot 1. histogram of x



Plot 2. histogram of y (ie: log(x))

3. See Q2_8.m for detail. Sample ACF of y_t and also plot of the first difference of y_t. It is observed that most times ACF is about 0.



4. $u_t = \log(x_t) - \log(x_{t-1}) = \log(\frac{x_t}{x_{t-1}}) = \log(1+\frac{x_t - x_{t-1}}{x_{t-1}}) \approx \frac{x_t - x_{t-1}}{x_{t-1}}$. The result is a very small number (Taylor's Theorem). A practical interpretation of u_t could be the percentage change in x_t.