
*M.Phil. in Mathematical Finance*

Numerical Methods in Finance 2020

Tutorial 3

---

## Problems

1. In this problem you will implement the chi-square goodness of fit test. Start by generating $n = 5,000$ random variables using the Matlab command `ncx2rnd(1,lambda,n,1);` with $\lambda = 100$. To ensure that the results are deterministic, issue the command `rng(0);` at the beginning of your code. Plot the histogram of these numbers.

   You will test the null hypothesis, $H_0$, that these numbers come from a normal distribution ($G$) with mean $\lambda + 1$ and standard deviation $\sqrt{2 + 4\lambda}$, using a two-sided chi-square test at the $\alpha = 5\%$ significance level. To do this, implement the following steps:

   (a) Divide the distribution $G$ into 10 intervals (i.e., $d = 9$), with each interval having equal probability. (Hint: use the Matlab commands `linspace` and `norminv`.)

   (b) Calculate the number of points expected in each interval ($np_i$, this should be a simple expression because of the equal probability intervals). Also, compute the observed number $n_i$ of points that occur in each interval. (Hint: use a logical expression in conjunction with the `sum` command.)

   (c) Compute the chi-square statistic

   $$Q_d = \sum_{i=1}^{d+1} \frac{(n_i - np_i)^2}{np_i}.$$

   (d) Use the function `chi2inv` to determine the critical values $\left(x_{\frac{\alpha}{2}} \text{ and } x_{1-\frac{\alpha}{2}}\right)$.

   (e) Compare the statistic $Q_d$ to the critical values, and determine whether to reject $H_0$ or fail to reject it.

   Now, perform the same steps again, but this time generate the 5,000 random variables using $\lambda = 1000$. Does your conclusion change?

   Note that in this problem you have generated non-central chi-square random variables with 1 degree of freedom and non-centrality parameter $\lambda$. As $\lambda$ gets larger, the limiting distribution becomes Gaussian.

2. Write a function called `runs` that takes a sequence of numbers and returns a vector of length 6 which contains the number of runs $(R_1, \ldots, R_5, \sum_{r=6}^{n} R_r)^T$ in the sequence. (Hint: This can be achieved with a for loop, a counter that represents the length of the current run and a carefully crafted if statement that determines whether the current element is the end of a run or not.) To ensure that your `runs` function is working correctly, compute the vector $\mathbf{X}$ on page 25 of the notes by issuing the command: `runs(LCG(2^16+3,0,2^31,2,5000))`, where `LCG` is the function you wrote in Tutorial 1.

   Now perform a one-sided chi-square test at the $\alpha = 1\%$ significance level to test the hypothesis, $H_0$, that the sequence (specified below) passes the runs test, by implementing the following steps:

   (a) Generate a sequence of 5000 numbers using your `LCG` function with the parameters $a = 2^{16} + 1$, $c = 2^8 + 3$ $m = 2^{32}$ and an initial seed value of $x_1 = 10$.

   (b) Plot a histogram of the sequence.

   (c) Compute the test statistic

   $$(\mathbf{X} - \mu)^T \hat{\Sigma}^{-1} (\mathbf{X} - \mu),$$

   where $\mathbf{X}$ is the output of the `runs` function on the sequence, and $\mu$ and $\hat{\Sigma}^{-1}$ are the expressions in the notes on pages 23 and 24.

   (d) Use the function `chi2inv` to determine the critical value.

   (e) Compare the statistic and the critical value, and determine whether to reject $H_0$ or fail to accept it.

   Repeat the process with different initial seeds. Would you trust the linear congruential generator with these parameter values?

3. (Bonus) Note that the runs test requires distinct random numbers. Convince yourself that the parameters above satisfy the conditions for a maximum period $m$ given by Theorem 1.1. in the notes. (Hint: use the Matlab command `factor`.)

4. (Bonus) Write code to compute the matrix $\Sigma$ from scratch, using the results of Theorem 1.8. You can then use this matrix to compute the statistics required for problem 2.

5. (Bonus) Produce the graph shown in Figure 1.4 in the notes. Perform the Kolmogorov-Smirnov test on the same sequence used in Example 1.7 in the notes and produce the graph in Figure 1.6. At what significance level does the test fail to accept the null hypothesis when you use a sequence of 10,000 numbers instead (all other parameters the same)?