（1）均匀随机数

通过混合同余法得到的10000个在[0,1]区间均匀分布随机数的描点，连线，直方图正常，在各个区间均匀分布且和由系统函数rand所生成的随机数图像保持一致。均值0.49724，约为0.5，最大值0.99986，最小值7.6294e-06，与理论值非常接近，满足数字特征。而理论概率密度分布图不像预设的那样，只在（0,1）区间上有且为1,在其余的区间均为0，而是在（-0.2,1.2）。这是由于ksdensity函数在拟合曲线时会用一些数据平滑的方式（矩形拟合，三角拟合等）对样本点进行平滑，有一个从0至1渐变的过程，无法产生突变，所以出现了缓冲带。可以通过设置ksdensity的参数来优化概率密度曲线，比如将取值区间限制在[0,1]，以及改变样本的平滑方式，但只能做到优化，而不能完全解决。

至于问题为什么减6，是因为期望和为12\*0.5=6，选择12为一组，是因为sqrt（x/12），选择12方便计算。

（2）高斯随机数

使用自编的函数（中心极限法）产生的高斯随机数。预设参数，均值为：3,标准差为2。实际均值为：2.9808,标准差为：1.9838。相对误差分别为：-0.64052 %, 和：-0.81214 %

。误差很小，可以认为服从高斯分布。并且由画出的频度分布直方图可以进一步证明正确性。

变换抽样法则预设参数，均值为：3,标准差为：2，计算参数，均值为：3.013,标准差为：2.0212，相对误差分别为：0.43337 %, 和：1.0578 %。误差很小，可以认为服从高斯分布。

概率密度分布图正常，分布均以3为中心集中，向两边逐渐稀疏减少，与预设理想N（3,4）概率密度分布保持一致。

对于由交换法生成的两组高斯随机数的关系，理论上应该是相互独立但不正交。但事实上不正交且相关。说明matlab只能模拟，与实际有区别。两组高斯随机数的协方差为0，说明是相互独立的（高斯随机数线性无关和统计独立等效）。通过计算其量组随机数的互相关函数，发现其不为0，说明两组高斯随机数不正交。

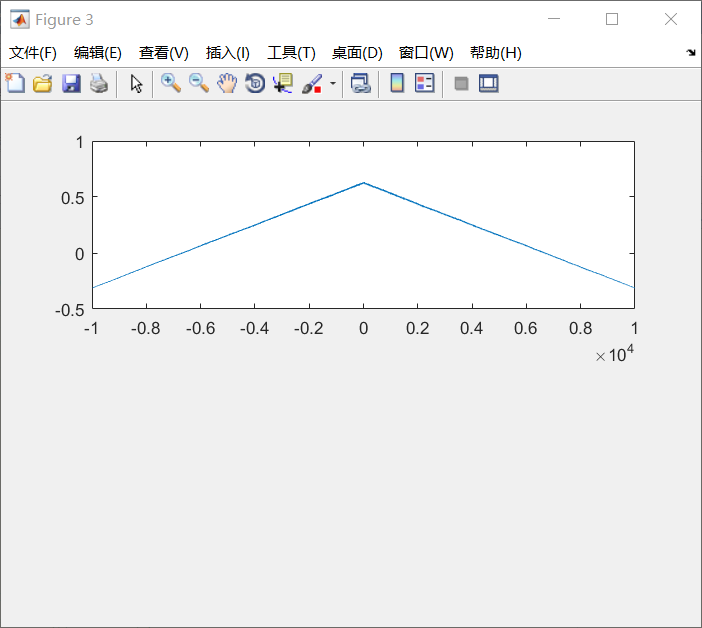
num=10000;

s=rand(1,num);

y=rand(1,num);

[zxg1,n1]=xcorr(s,y,'coeff');

figure,subplot(2,1,1),plot(n1,zxg1-0.25);



（3）高斯分布的自相关以及互相关函数

其中自相关函数的图像能够比较好的用理论知识进行解释，为偶函数并且在0处取得最大值。比较难解释的是互相关函数为什么呈偶对称，从理论上看，互相关函数不一定是偶对称的，但是我测了多组高斯分布的互相关函数，无一例外的图像都近似是偶对称，而数据却不是偶对称的。按我的理解，两个独立的高斯分布，是由高斯变换法中的ax+b生成的，x是生成的标准正态函数，所以他们俩的互相关函数实质上是x互相关函数和它们期望的线性和，而由于x具有宽平稳的性质，为常数，所以它们的互相关有着x的互相关函数的性质。

（4）探究性实验

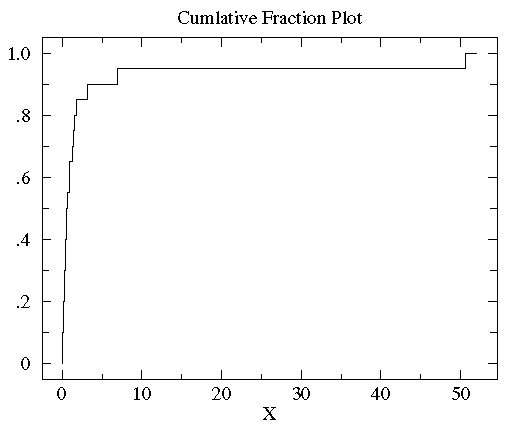
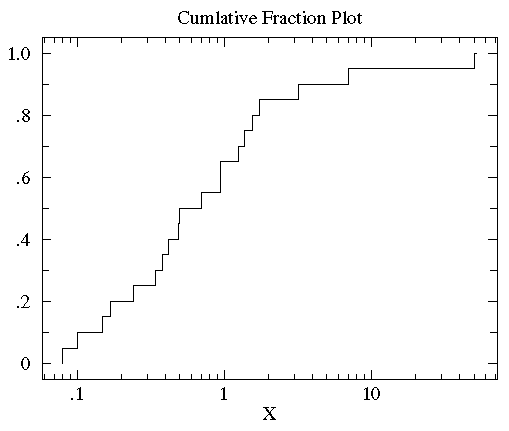
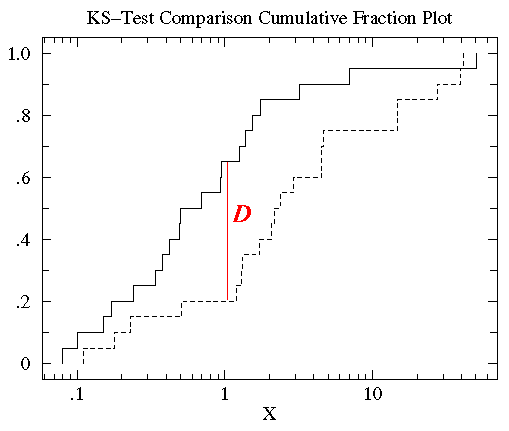
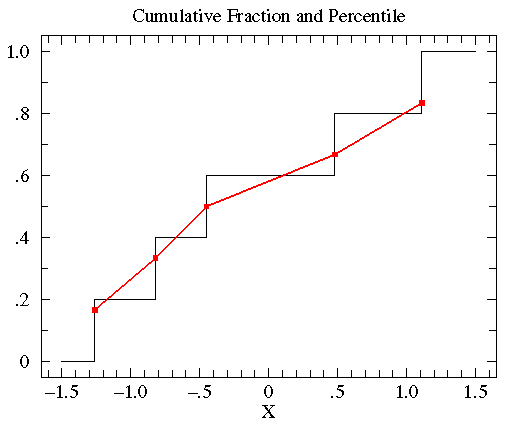
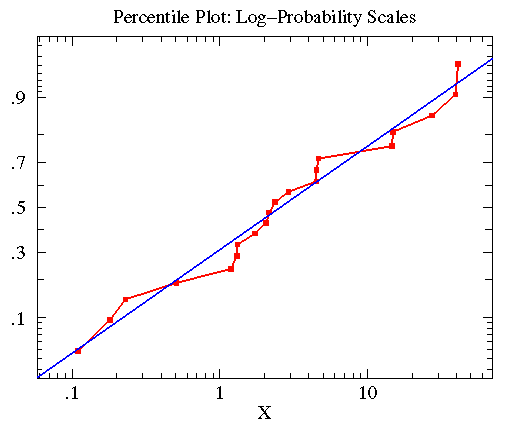
计算两个函数互相关函数，得到延迟点数为12000点，根据采样率算出延迟时间为1464.8438ms。接着画出概率密度图像，发现与高斯分布的图像相似。最后使用matlab自带的kstest函数进行置信度为95%的假设检验，得出结果为符合均值为0,方差为0.11的高斯分布，猜想正确。

kstest函数不需要知道数据原分布是什么类型，是一种非参数检验方法。通过normfit函数得到噪声信号在置信区间下的高斯均值和方差的估计值，然后由normcdf函数根据估计值得到噪声信号的高斯概率密度分布，再由kstest函数进行判断其是否在置信区间内满足估计概率密度分布与实际的一致。当实际观测值D>D(n,α)，在置信水平上不满足假设，则拒绝H0，认为两者分布不同，否则认为H0成立。还可以通过unifit和unifcdf函数检验其是否满足均匀分布，答案是不满足，还可以通过改变检验条件检验很多类型的分布，验证结束后通过画噪声信号的直方图和概率分布图以及对误差的计算进一步验证了其为高斯分布，且近似标准高斯分布。

**下面是ks检验的解释，感觉这个解释最通俗易懂：**

Kolmogorov-Smirnov是比较一个频率分布f(x)与理论分布g(x)或者两个观测值分布的检验方法。其原假设H0:两个数据分布一致或者数据符合理论分布。D=max| f(x)- g(x)|，当实际观测值D>D(n,α)则拒绝H0，否则则接受H0假设。  
KS检验与t-检验之类的其他方法不同是KS检验不需要知道数据的分布情况，可以算是一种非参数检验方法。当然这样方便的代价就是当检验的数据分布符合特定的分布事，KS检验的灵敏度没有相应的检验来的高。在样本量比较小的时候，KS检验最为非参数检验在分析两组数据之间是否不同时相当常用。  
**PS：**t-检验的假设是检验的数据满足正态分布，否则对于小样本不满足正态分布的数据用t-检验就会造成较大的偏差，虽然对于大样本不满足正态分布的数据而言t-检验还是相当精确有效的手段。

**KS检验是如何工作的？**

1. 首先观察下分析数据  
   对于以下两组数据：  
   controlB={1.26, 0.34, 0.70, 1.75, 50.57, 1.55, 0.08, 0.42, 0.50, 3.20, 0.15, 0.49, 0.95, 0.24, 1.37, 0.17, 6.98, 0.10, 0.94, 0.38}  
   treatmentB= {2.37, 2.16, 14.82, 1.73, 41.04, 0.23, 1.32, 2.91, 39.41, 0.11, 27.44, 4.51, 0.51, 4.50, 0.18, 14.68, 4.66, 1.30, 2.06, 1.19}  
   对于controlB，这些数据的统计描述如下：  
   Mean = 3.61  
   Median = 0.60  
   High = 50.6 Low = 0.08  
   Standard Deviation = 11.2  
   可以发现这组数据并不符合正态分布， 否则大约有15%的数据会小于均值-标准差（3.61-11.2），而数据中显然没有小于0的数。
2. 观察数据的累计分段函数（Cumulative Fraction Function）  
   对controlB数据从小到大进行排序：  
   sorted controlB={0.08, 0.10, 0.15, 0.17, 0.24, 0.34, 0.38, 0.42, 0.49, 0.50, 0.70, 0.94, 0.95, 1.26, 1.37, 1.55, 1.75, 3.20, 6.98, 50.57}。10%的数据（2/20）小于0.15，85%（17/20）的数据小于3。所以，对任何数x来说，其累计分段就是所有比x小的数在数据集中所占的比例。下图就是controlB数据集的累计分段图  
     
   可以看到大多数数据都几种在图片左侧（数据值比较小），这就是非正态分布的标志。为了更好的观测数据在x轴上的分布，可以对x轴的坐标进行非等分的划分。**在数据都为正的时候有一个很好的方法就是对x轴进行log转换。**下图就是上图做log转换以后的图：  
     
   将treatmentB的数据也做相同的图（如下），可以发现treatmentB和controlB的数据分布范围大致相同（0.1 - 50）。但是对于大部分x值，在controlB数据集中比x小的数据所占的比例比在treatmentB中要高，也就是说达到相同累计比例的值在treatment组中比control中要高。KS检验使用的是两条累计分布曲线之间的最大垂直差作为D值（statistic D）作为描述两组数据之间的差异。在此图中这个D值出现在x=1附近，而D值为0.45（0.65-0.25）。  
     
   值得注意的是虽然累计分布曲线的性状会随着对数据做转换处理而改变（如log转换），但是D值的大小是不会变的。
3. 百分比图（percentile plot）  
   估算分布函数肩形图（Estimated Distribution Function Ogive）是一种累计分段图的替代方式。其优势在于可以让你使用概率图纸作图（坐标轴经过特殊分段处理，y轴上的数值间隔符合正态分布），从而根据概率在y轴上的分布可以直观的判断数据到底有多符合正态分布，因为正态分布的数据在这种坐标上是呈一条直线。  
   那么这种图是如何画的呢？  
   假设我们有这5个数{-0.45, 1.11, 0.48, -0.82, -1.26}，从小到大对它们进行排序，{ -1.26, -0.82, -0.45, 0.48, 1.11 }。0.45是中位数，百分比为0.5，而0.45的累计分布函数中占了0.4到0.6的区间。根据数据x在数据集（N）中排位r可以计算x的百分数（percentile）为r/(N+1)。将上述数据与他们的百分数配对，得到{ (-1.26,.167), (-0.82,.333), (-0.45,.5), (0.48,.667), (1.11,.833) }。然后将各点之间用直线连接就是百分比图了。如下图中红线所示（另一条线为累计分段曲线）。  
     
   treatmentB的数据近似对数正态分布，其几何均值为2.563，标准差为6.795。该数据的百分图（红）与其近似的对数正态分布曲线（蓝）如下。  
     
   由于数据近似正态分布，所以对其采用t-检验是最佳的检验方法。