# ECE 514 Mini-project Fall 2019

## 1. Background

In this mini-project, I am about to estimate the bit-error-rate (BER) or probability of bit error for a simple digital communication system with the use of confidence intervals in terms of known variance and unknown variance. In theory, after appropriate filtering and sampling, the received sample for a particular bit period can be modeled as follows:

$$R = \sqrt{E_b}B + \sqrt{N_0/2}N \,, \tag{1}$$

where $E_b$ is the energy-per-bit, $N_0$ is the noise power spectral density, $B$ is the transmitted bit value (+1 or –1) and $N$ is normalized Gaussian noise with mean zero and variance 1. Thus, the signal has energy (mean-square value) $E_b$ and the noise has energy $N_0/2$, giving a signal-to-noise ratio (SNR) of $2E_b/N_0$, where $E_b/N_0$ is a conventional, normalized SNR used in the digital communication literature.

In order to accomplish project, I will use MATLAB to simulate received values and estimate BER when knowing the true variance or not. As bit values are equi-likely, it's safe for us to use detection threshold 0 so that the BER is the same for +1 and -1 bit values. While using MATLAB to do the simulation, I am going to assume -1 values are sent and define a bit error as the event $R > 0 \mid B = -1$. Define the random variable $X$ to be 0 if no error occurred and 1 if an error occurred.

From analysis, we know that the true BER is given by

$$BER = 0.5erfc(\sqrt{E_b/N_0}) \,, \tag{2}$$

where **erfc** is the complementary error function. Later I will use this to evaluate our confidence intervals. In this project, I will set

$$E_b/N_0 = -3dB(0.5) \,. \tag{3}$$

## 2. Experimental setup

I will use MATLAB to simulate m trials. For each trial, I follow the steps below.

**Step1:** Initialize the random number generator utilizing the built-in function **rng** (seed, 'twister'). My name is Yao Fu so that the sum of the set numbers corresponding to the vowels present is 1152 (16'a'+512'o'+1024'u'). Therefore, I will set seed as 1152.

**Step2:** After initializing the random number generator, I will use the **randn** function in MATLAB to generate n received values in each trial. It's not hard to get that

$$R\sqrt{\frac{2}{N_0}} = \sqrt{\frac{2E_b}{N_0}}B + N = B + N \,. \tag{4}$$

Recall that $B$ is -1 and I let those n received random values represent $N$. Hence, the number of values of $B+N$ is n. In other words, I have received n samples. If the value of $B+N$ is greater than zero, the value of this sample will be set as 1, or will be set as 0 if not. Now we have a vector of $X$, which has n components whose value is either 0 or 1. In this situation, I could use the sample mean of $X$ to estimate BER.

**Step3:** Determine the 68.3% confidence interval using the known variance or unknown variance of $X$. As a result, I will get lower and upper confidence interval boundaries. From the data of m trials, I will determine the fraction of trials for which the true BER falls within the confidence interval in the subsection 3.1 and plot six plots in the subsection 3.2. One important thing in this step is that I have to consider two different circumstances: one is that we know the true variance of sample $X$, and the other is that we have to estimate the variance of $X$ in order to calculate the lower and upper confidence interval boundaries.

**(i) know the true variance of $X$:**

From equation (2), we know the true BER. Also, we know that the $X$ vector has n components whose value is either 0 or 1. Therefore, it's easy to reach the true variance of $X$

$$\sigma_X^2 = E(x^2) - (E(x))^2 = BER - BER^2 \tag{5}$$

The sample mean

$$M = mean(X) \tag{6}$$

The lower confidence interval boundary is

$$BER - \sigma_X y_{0.5\alpha} \big/ \sqrt{n} \tag{7}$$

and the upper confidence interval boundary is

$$BER + \sigma_X y_{0.5\alpha} \big/ \sqrt{n} \tag{8}$$

Where

$$1 - \alpha = 68.3\%$$

**(ii) Estimate the variance of $X$:**

From equation (2), we know the true BER. Also, we know that the $X$ vector has n components whose value is either 0 or 1. Therefore, it's easy to reach the estimated variance of $X$

$$V = [(\sum_{i=1}^{n} X_i^2) - nM^2] \Big/ (n-1) = [nM - nM^2]/(n-1) \tag{9}$$

The sample mean

$$M = mean(X) \tag{10}$$

The lower confidence interval boundary is

$$BER - \sqrt{V} y_{0.5\alpha} \big/ \sqrt{n} \tag{11}$$

and the upper confidence interval boundary is

$$BER + \sqrt{V} y_{0.5\alpha} \big/ \sqrt{n} \tag{12}$$

where

$$1 - \alpha = 68.3\%$$

## 3. Results

Now, I use m=100 and produce results for n=10, 100, and 1000 respectively.

### 3.1. The resulting table

A table with the fraction of trials for which the true mean falls within the confidence interval, which is shown in the TABLE1. A row for each value of n and columns for known and estimated variance. Provide a paragraph discussing the results of the table.

TABLE1: the fraction of the true mean falling within the confidence interval

| | **Known variance** | **Estimated variance** |
|---|---|---|
| **n=10** | 0.63 | 0.75 |
| **n=100** | 0.63 | 0.60 |
| **n=1000** | 0.68 | 0.68 |

**The discussion of results:**

As is clearly seen from the TABLE1, the bigger of the value of n is, the closer the fraction is to the 0.683. This phenomenon corresponds to the fundamental rule of statistics that the more samples we have, the more accurate the results will be. Moreover, I could find out that a confidence interval is a random set, and the confidence level is the probability that the random set contains the unknown objective parameter.

### 3.2. Six plots

In this subsection, I will plot 6 figures (3 n values x known vs. estimated variance). In each figure, I will use the information of the first 10 trials:

a) the sample mean of $X$,

b) the lower confidence interval boundary,

c) the upper confidence interval boundary,

d) the true BER.

The 6 figures are shown in Fig.1, Fig.2, Fig.3, Fig.4, Fig.5, and Fig.6 respectively, where the red dot denotes the sample mean, the upper green triangle and the lower green triangle denote the upper confidence interval boundary and the lower confidence interval boundary accordingly, and the blue dashed line will represent the true BER.

**The discussion of results:**

As is clearly seen from the six figures, roughly speaking, more than half of the trials for which the true BER does fall within the two confidence intervals. From this angle, I could safely acknowledge the validity of the relationship between the confidence interval and the confidence level.

However, from the Fig.4, which is the plot with unknown variance when n=10, there is something wrong with the 5$^{th}$, 7$^{th}$, and 9$^{th}$ trials. The sample mean and two confidence boundaries are all zero! With hint of the professor, I look through the textbook. Finally, I discover that if the probability of the event is very small, we are likely to have all samples equal to zero, which forces both the sample mean and the sample variance to be zero too, thus making my figure look seemingly unreasonable! In practice, there are usually two methods to resolve this problem: taking more measurements and utilizing more sophisticated strategies such as importance sampling.

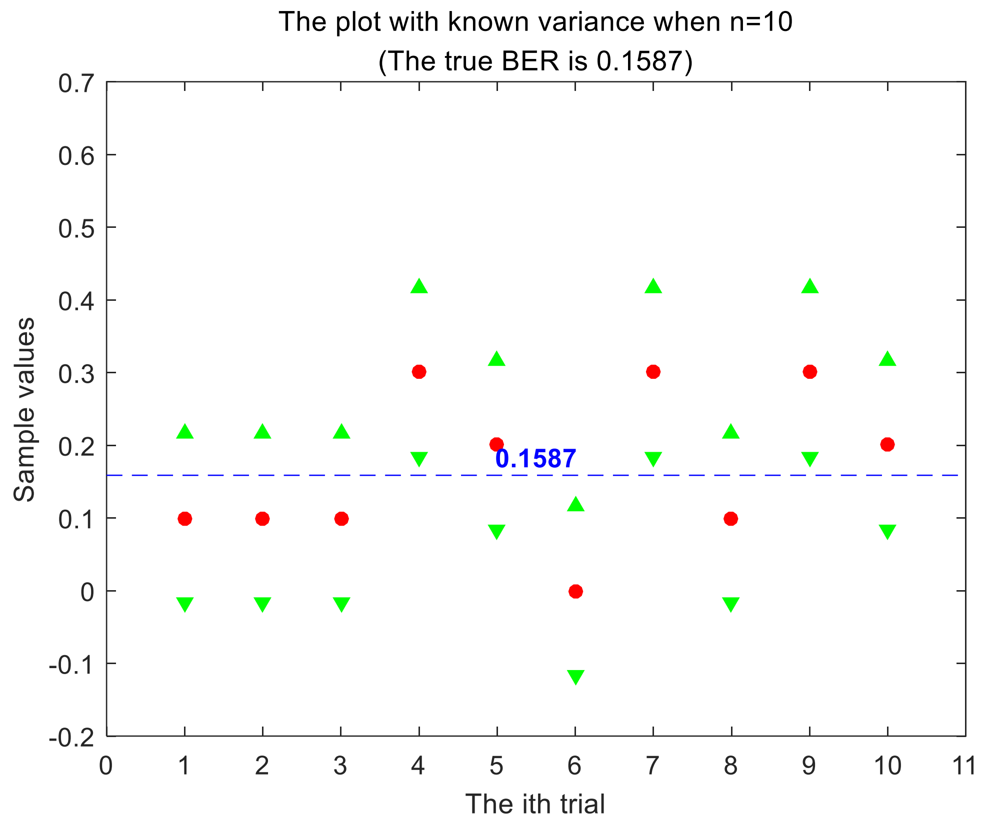


Fig. 1. The plot of first 10 trials with known variance when n=10

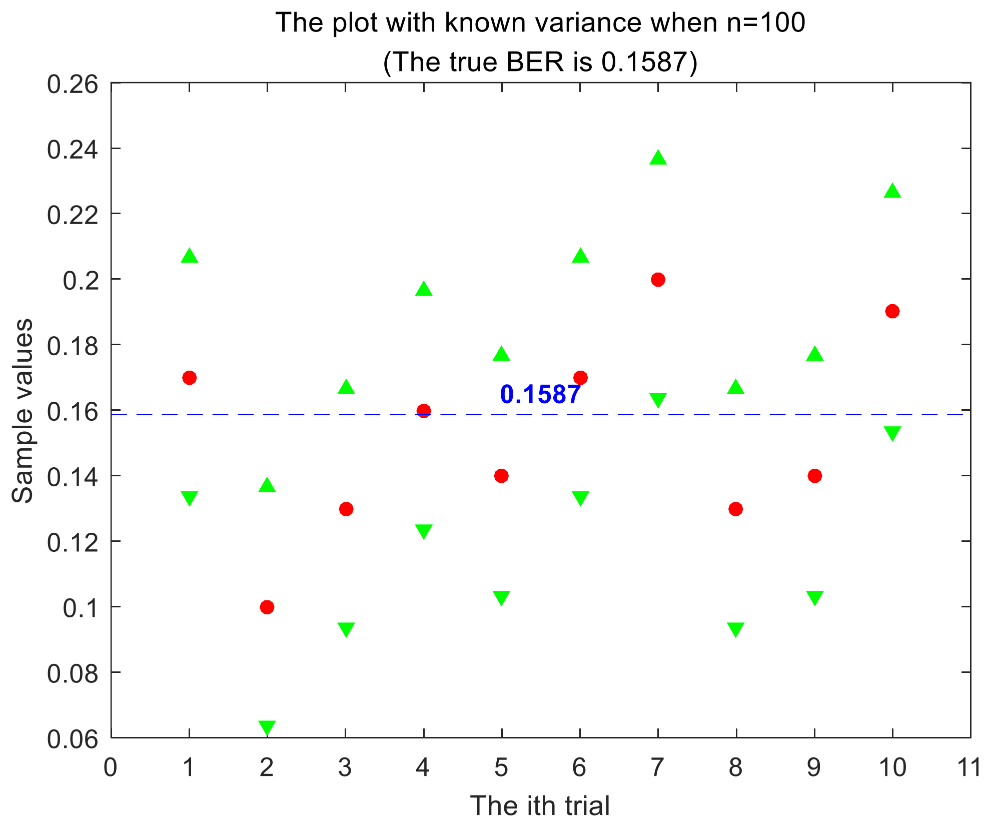


Fig. 2. The plot of first 10 trials with known variance when n=100

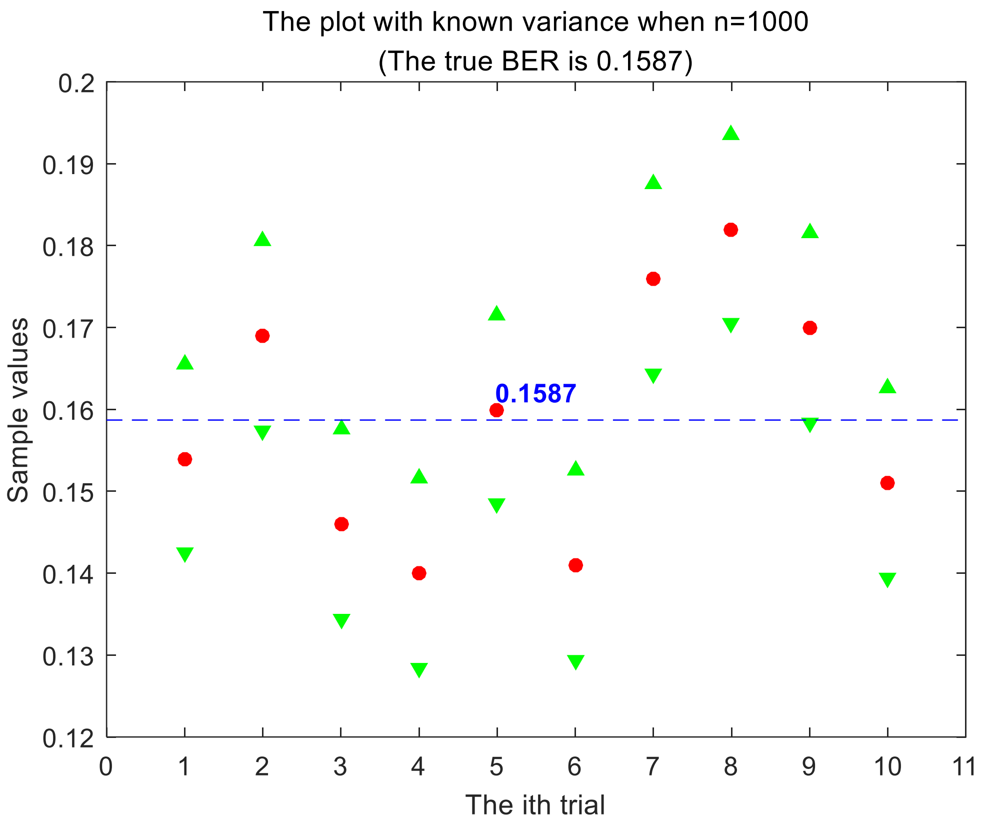


Fig. 3. The plot of first 10 trials with known variance when n=1000

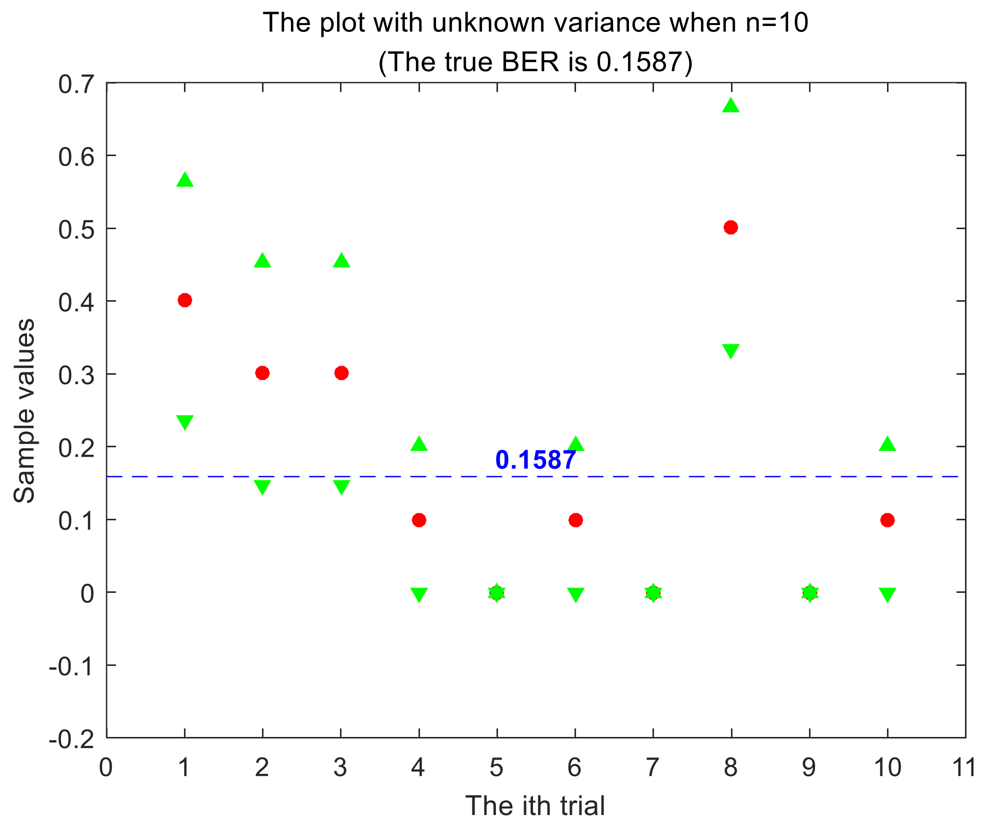


Fig. 4. The plot of first 10 trials with unknown variance when n=10

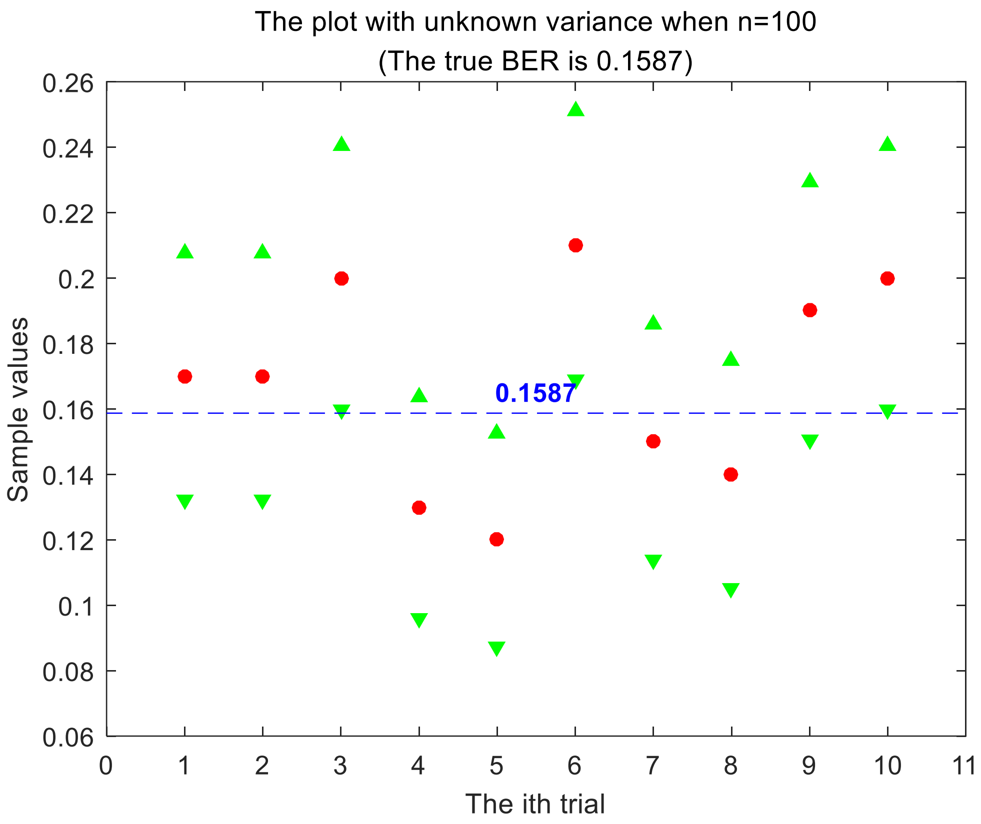


Fig. 5. The plot of first 10 trials with unknown variance when n=100

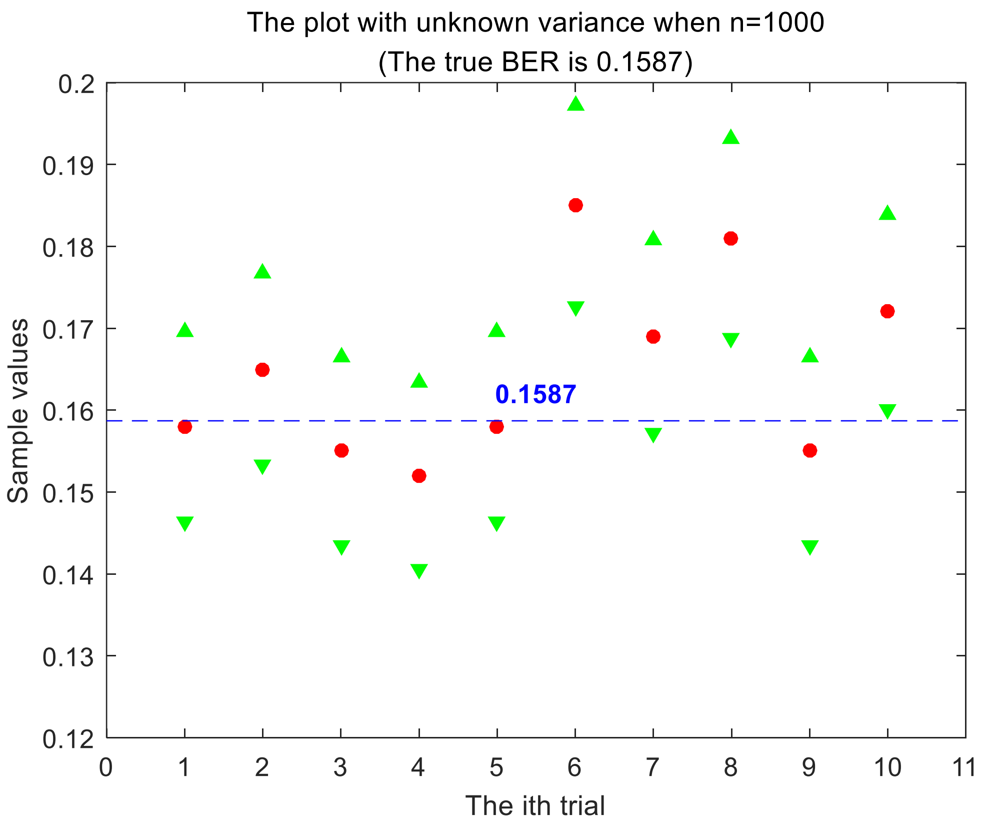


Fig. 6. The plot of first 10 trials with unknown variance when n=1000

## 4. Conclusion

During the course of finishing my project, I have learned a great deal either from understanding the fundamental statistical knowledge or from using MATLAB. First of all, in the past I just merely know the theoretical knowledge of confidence interval and confidence level. Now after accomplishing this project, I have a more impressing experience to know how we could say something about how close a sample mean is close to the true mean when we only have limited samples. In the future, I am likely to take some courses ranging from Machine Learning to Data Analysis, where I will confront more sophisticated knowledge of statistics. I hope this project can motivate me to learn more about statistics.

What's more, in this process, I also acquire more experience and improve my programming skills while using MATLAB. For example, I have learned how to generate a number of random numbers to conduct some research; and I have also known some new operations to use MATLAB to plot figures. In the future's study, I will definitely use MATLAB to do more research.

## 5. Appendix

```
%% Operate the data
rng(1552,'twister')
BER_true=0.5*erfc(sqrt(0.5));
alpha=1-0.683;
y_half_a=norminv(1-alpha/2);
%Use the known variance of X
lower_confi_boud11_k=zeros(1,100);
upper_confi_boud11_k=zeros(1,100);
lower_confi_boud12_k=zeros(1,100);
upper_confi_boud12_k=zeros(1,100);
lower_confi_boud13_k=zeros(1,100);
upper_confi_boud13_k=zeros(1,100);
trial_11=zeros(1,100);
trial_12=zeros(1,100);
trial_13=zeros(1,100);
M11=zeros(1,100);
M12=zeros(1,100);
M13=zeros(1,100);
for m=1:1:100
X11=randn(1,10)-ones(1,10);
X12=randn(1,100)-ones(1,100);
X13=randn(1,1000)-ones(1,1000);
for i=1:1:10
    if X11(i)>0
        X11(i)=1;
    else
        X11(i)=0;
    end
end
for i=1:1:100
```

```
    if X12(i)>0
        X12(i)=1;
    else
        X12(i)=0;
    end
end
for i=1:1:1000
    if X13(i)>0
        X13(i)=1;
    else
        X13(i)=0;
    end
end
M11(m)=mean(X11);                % The sample mean of X11
M12(m)=mean(X12);                % The sample mean of X12
M13(m)=mean(X13);                % The sample mean of X13
V=BER_true-BER_true^2;            % The true variance of X

lower_confi_boud11_k(m)=M11(m)-sqrt(V)*y_half_a/sqrt(10);
upper_confi_boud11_k(m)=M11(m)+sqrt(V)*y_half_a/sqrt(10);
lower_confi_boud12_k(m)=M12(m)-
sqrt(V)*y_half_a/sqrt(100);
upper_confi_boud12_k(m)=M12(m)+sqrt(V)*y_half_a/sqrt(100)
;
lower_confi_boud13_k(m)=M13(m)-
sqrt(V)*y_half_a/sqrt(1000);
upper_confi_boud13_k(m)=M13(m)+sqrt(V)*y_half_a/sqrt(1000
);
end
for m=1:1:100
    if
((lower_confi_boud11_k(m)<=BER_true)&&(BER_true<=upper_co
nfi_boud11_k(m)))
        trial_11(m)=1;
    end
    if
((lower_confi_boud12_k(m)<=BER_true)&&(BER_true<=upper_co
nfi_boud12_k(m)))
        trial_12(m)=1;
    end
    if
((lower_confi_boud13_k(m)<=BER_true)&&(BER_true<=upper_co
nfi_boud13_k(m)))
        trial_13(m)=1;
```

```
    end
end
mean(trial_11)
mean(trial_12)
mean(trial_13)
% Use the unknown variance of X
lower_confi_boud21_k=zeros(1,100);
upper_confi_boud21_k=zeros(1,100);
lower_confi_boud22_k=zeros(1,100);
upper_confi_boud22_k=zeros(1,100);
lower_confi_boud23_k=zeros(1,100);
upper_confi_boud23_k=zeros(1,100);
trial_21=zeros(1,100);
trial_22=zeros(1,100);
trial_23=zeros(1,100);
M21=zeros(1,100);
M22=zeros(1,100);
M23=zeros(1,100);
for m=1:1:100
X21=randn(1,10)-ones(1,10);
X22=randn(1,100)-ones(1,100);
X23=randn(1,1000)-ones(1,1000);
for i=1:1:10
    if X21(i)>0
        X21(i)=1;
    else
        X21(i)=0;
    end
end
for i=1:1:100
    if X22(i)>0
        X22(i)=1;
    else
        X22(i)=0;
    end
end
for i=1:1:1000
    if X23(i)>0
        X23(i)=1;
    else
        X23(i)=0;
    end
end
```

```
M21(m)=mean(X21);                          % The sample mean of
X21
M22(m)=mean(X22);                          % The sample mean of
X22
M23(m)=mean(X23);                          % The sample mean of
X23
V1=(10*M21(m)-10*(M21(m))^2)/9;              % The sample
variance of X21
V2=(100*M22(m)-100*(M22(m))^2)/99;           % The sample
variance of X22
V3=(1000*M23(m)-1000*(M23(m))^2)/999;    % The sample
variance of X23

lower_confi_boud21_k(m)=M21(m)-
sqrt(V1)*y_half_a/sqrt(10);
upper_confi_boud21_k(m)=M21(m)+sqrt(V1)*y_half_a/sqrt(10)
;
lower_confi_boud22_k(m)=M22(m)-
sqrt(V2)*y_half_a/sqrt(100);
upper_confi_boud22_k(m)=M22(m)+sqrt(V2)*y_half_a/sqrt(100
);
lower_confi_boud23_k(m)=M23(m)-
sqrt(V3)*y_half_a/sqrt(1000);
upper_confi_boud23_k(m)=M23(m)+sqrt(V3)*y_half_a/sqrt(100
0);
end
for m=1:1:100
    if
((lower_confi_boud21_k(m)<=BER_true)&&(BER_true<=upper_co
nfi_boud21_k(m)))
        trial_21(m)=1;
    end
    if
((lower_confi_boud22_k(m)<=BER_true)&&(BER_true<=upper_co
nfi_boud22_k(m)))
        trial_22(m)=1;
    end
    if
((lower_confi_boud23_k(m)<=BER_true)&&(BER_true<=upper_co
nfi_boud23_k(m)))
        trial_23(m)=1;
    end
end
mean(trial_21)
```

```
mean(trial_22)
mean(trial_23)
%% Use the data to draw figures
%% The sample mean of X with known variance
figure(1)   % n=10
X=1:10;
plot(X,M11(1:10),'ro','MarkerSize',5,'MarkerFaceColor','r
');hold on;
plot(X,lower_confi_boud11_k(1:10),'gv','MarkerSize',5,'Ma
rkerFaceColor','g');hold on;
plot(X,upper_confi_boud11_k(1:10),'g^','MarkerSize',5,'Ma
rkerFaceColor','g');hold on;
xlim([0 11])
ylim([-0.2 0.7])
title({'The plot with known variance when n=10';'(The
true BER is 0.1587)'})
xlabel('The ith trial')
ylabel('Sample mean values')
line([0,11],[0.1587,0.1587],'linestyle','--
','color','b');
text(5.5,0.16,'0.1587','FontWeight','bold','Color','b','h
oriz','center','vert','bottom')

figure(2)   % n=100
X=1:10;
plot(X,M12(1:10),'ro','MarkerSize',5,'MarkerFaceColor','r
');hold on;
plot(X,lower_confi_boud12_k(1:10),'gv','MarkerSize',5,'Ma
rkerFaceColor','g');hold on;
plot(X,upper_confi_boud12_k(1:10),'g^','MarkerSize',5,'Ma
rkerFaceColor','g');hold on;
xlim([0 11])
ylim([0.06 0.26])
title({'The plot with known variance when n=100';'(The
true BER is 0.1587)'})
xlabel('The ith trial')
ylabel('Sample mean values')
line([0,11],[0.1587,0.1587],'linestyle','--
','color','b');
text(5.5,0.16,'0.1587','FontWeight','bold','Color','b','h
oriz','center','vert','bottom')

figure(3)   % n=1000
X=1:10;
```

```
plot(X,M13(1:10),'ro','MarkerSize',5,'MarkerFaceColor','r
');hold on;
plot(X,lower_confi_boud13_k(1:10),'gv','MarkerSize',5,'Ma
rkerFaceColor','g');hold on;
plot(X,upper_confi_boud13_k(1:10),'g^','MarkerSize',5,'Ma
rkerFaceColor','g');hold on;
xlim([0 11])
ylim([0.12 0.20])
title({'The plot with known variance when n=1000';'(The
true BER is 0.1587)'})
xlabel('The ith trial')
ylabel('Sample mean values')
line([0,11],[0.1587,0.1587],'linestyle','--
','color','b');
text(5.5,0.16,'0.1587','FontWeight','bold','Color','b','h
oriz','center','vert','bottom')
%% The sample mean of X with unknown variance
figure(4)   % n=10
X=1:10;
plot(X,M21(1:10),'ro','MarkerSize',5,'MarkerFaceColor','r
');hold on;
plot(X,lower_confi_boud21_k(1:10),'gv','MarkerSize',5,'Ma
rkerFaceColor','g');hold on;
plot(X,upper_confi_boud21_k(1:10),'g^','MarkerSize',5,'Ma
rkerFaceColor','g');hold on;
xlim([0 11])
ylim([-0.2 0.7])
title({'The plot with unknown variance when n=10';'(The
true BER is 0.1587)'})
xlabel('The ith trial')
ylabel('Sample mean values')
line([0,11],[0.1587,0.1587],'linestyle','--
','color','b');
text(5.5,0.16,'0.1587','FontWeight','bold','Color','b','h
oriz','center','vert','bottom')

figure(5)   % n=100
X=1:10;
plot(X,M22(1:10),'ro','MarkerSize',5,'MarkerFaceColor','r
');hold on;
plot(X,lower_confi_boud22_k(1:10),'gv','MarkerSize',5,'Ma
rkerFaceColor','g');hold on;
plot(X,upper_confi_boud22_k(1:10),'g^','MarkerSize',5,'Ma
rkerFaceColor','g');hold on;
```

```
xlim([0 11])
ylim([0.06 0.26])
title({'The plot with unknown variance when n=100';'(The
true BER is 0.1587)'})
xlabel('The ith trial')
ylabel('Sample mean values')
line([0,11],[0.1587,0.1587],'linestyle','--
','color','b');
text(5.5,0.16,'0.1587','FontWeight','bold','Color','b','h
oriz','center','vert','bottom')

figure(6)   % n=1000
X=1:10;
plot(X,M23(1:10),'ro','MarkerSize',5,'MarkerFaceColor','r
');hold on;
plot(X,lower_confi_boud23_k(1:10),'gv','MarkerSize',5,'Ma
rkerFaceColor','g');hold on;
plot(X,upper_confi_boud23_k(1:10),'g^','MarkerSize',5,'Ma
rkerFaceColor','g');hold on;
xlim([0 11])
ylim([0.12 0.20])
title({'The plot with unknown variance when n=1000';'(The
true BER is 0.1587)'})
xlabel('The ith trial')
ylabel('Sample mean values')
line([0,11],[0.1587,0.1587],'linestyle','--
','color','b');
text(5.5,0.16,'0.1587','FontWeight','bold','Color','b','h
oriz','center','vert','bottom')
```