Assignment 2

Sainivedhitha Arunajatesan

Exercise 1

Consider random numbers X_i that are independent and all have the uniform distribution on the interval (0,1). We study the sample mean $\overline{X} = (X_1 + X_2 + ... + X_n)/n$.

$$Mean(\mu) = \frac{a+b}{2} = 0.5$$
$$\sigma^2 = \frac{(b-a)^2}{12} = 0.083$$
$$StandardDeviation(\sigma) = \sqrt{0.083} = 0.288$$

1. For the sample sizes n=100, n=1000, n=10 000 and n=100 000 approximate the probability $P(\overline{X} \leq 0.507)$ using the central limit theorem (CLT).

   Also,

   $$Z = \frac{n*\overline{X} - n*\mu}{\sigma * \sqrt{n}}$$

   As uniform distribution is approximated using normal distribution, the value of $\overline{X} \leq 0.507$ is replaced by $\overline{X} \leq 0.557$

   For n=100,

   $$P(\overline{X} \leq 0.557) = P(Z \leq \frac{5.7}{2.88}) = P(Z \leq 1.979)$$
   $$P(Z \leq 1.979) = 0.9756 = 97.56\%$$

   For n=1000,

   $$P(\overline{X} \leq 0.557) = P(Z \leq \frac{57}{9.10}) = P(Z \leq 6.263)$$
   $$P(Z \leq 6.263) = 1 = 100\%$$

   For n=10 000,

   $$P(\overline{X} \leq 0.557) = P(Z \leq \frac{570}{28.8}) = P(Z \leq 19.791)$$
   $$P(Z \leq 19.791) = 1 = 100\%$$

   For n=100 000,

   $$P(\overline{X} \leq 0.557) = P(Z \leq \frac{5700}{91}) = P(Z \leq 62.637)$$
   $$P(Z \leq 62.637) = 1 = 100\%$$

2. The probability $P(\overline{X} \leq 0.507)$ is an increasing function of n. Check whether your calculations are in agreement with this statement and explain why this statement is true.

   The statement that the probability $P(\overline{X} \leq 0.507)$ is an increasing function of n is true until a certain value of n. At a point, the value of $P(\overline{X} \leq 0.507)$ reaches a maximum of 100% above which there will be no change. It can be witnessed in the above calculation of $P(\overline{X} \leq 0.507)$ for increasing n values. For n=1000, the value reached maximum of 100% and for any value of n above n=1000 it is 100% and hence proved.

Exercise 2

The probability of successful first serve (p) and the total number of serves (n) is noted from the question.

p=0.8 ; n=120

The mean and variance is calculated by using the formula of the binomial distribution.

Mean = 96; Variance = 4.381

Using normal approximation of the binomial distribution to compute the probabilities, the value of Z is formulated as:

$$Z = \frac{X-96}{4.381}$$

The values of Z are calculated from the above equation while the probabilities are found using the normal distribution table.

1. $P(X \geq 100)$

$$Z = \frac{99.5 - 96}{4.381} = 0.798$$
$$P(Z \geq 0.798) = 1 - 0.7852 = 0.2148 = 21.48\%$$

2. $P(X < 90)$

$$Z = \frac{90.5 - 96}{4.381} = -1.25$$
$$P(Z \leq -1.25) = 1 - 0.8944 = 0.1056 = 10.56\%$$

3. $P(95 < X \leq 105)$

$$Z1 = \frac{94.5 - 96}{4.381} = -0.342$$
$$Z2 = \frac{105.5 - 96}{4.381} = 2.168$$

$$P(-0.342 < Z \leq 2.168) = 0.9846 - (1 - 0.6331) = 0.6177 = 61.77\%$$

Exercise 3

The waiting time X of patients in some hospital has an exponential distribution with (long run) average μ=12 (unit is minute).

1. Calculate the probability P(X >20), the probability that an arbitrary patient has to wait more than 20 minutes.

The mean of exponential distribution is calculated using the formula:

$$\begin{aligned}
\mu &= 1/\lambda \\
12 &= 1/\lambda \\
\lambda &= 0.083 \\
P(X > 20) &= -e^{-\lambda * x} \\
&= e^{-0.083*20} = -e^{-1.66} \\
&= 0.19
\end{aligned}$$

2. Consider the total waiting time $T = X_1 + X_2 + ... + X_{50}$ 50 patients and assume that the 50 waiting times are independent and exponentially distributed with expectation 12. Approximate the probability that the total waiting time exceeds 660 minutes (11 hours).

   For exponential distribution, mean = standard deviation

$$\mu = \sigma = 12; n = 50$$

   Approximating the probability using CLT,

$$Z = \frac{(T - n\mu)}{\sigma * \sqrt{n}}$$

$$\begin{aligned}
Z &= \frac{(659.5 - 600)}{84.852} = 0.7012 \\
P(Z > 0.7012) &= 1 - 0.7580 = 0.242 = 24.2\%
\end{aligned}$$

Exercise 4

Let us return to the binomial test of the lecture notes. We observe X= the number of cured patients by the new medicine which has the binomial distribution with n=200 and unknown p, we test H_0:p=0.60 against H_1: p>0.60 and we reject the null hypothesis if $X \geq c$. In this exercise we choose α=10%.

1. Use Excel and the statistical function BINOM.DIST to determine c.

   Using the values of n=1 to n=200, the binomial distribution was calculated using BINOM.DIST in EXCEL. The values ranged between 10^{-78} to +1. These binomial distribution values were subtracted from 1 to obtain α. Between n=129 and n=130, the value of α = 0.1 or 10%.

$$\begin{aligned}
P(X \geq 129) &= 0.109415 \\
P(X \geq 130) &= 0.084398
\end{aligned}$$

   From these values, it is evident that $P(X \geq 129)$ is closer to 0.1 and therefore the value of c=129.

2. Calculate the probability $P(X \geq c)$ for p=0.65 using Excel and BINOM.DIST. This is the power of the test for p=0.65.

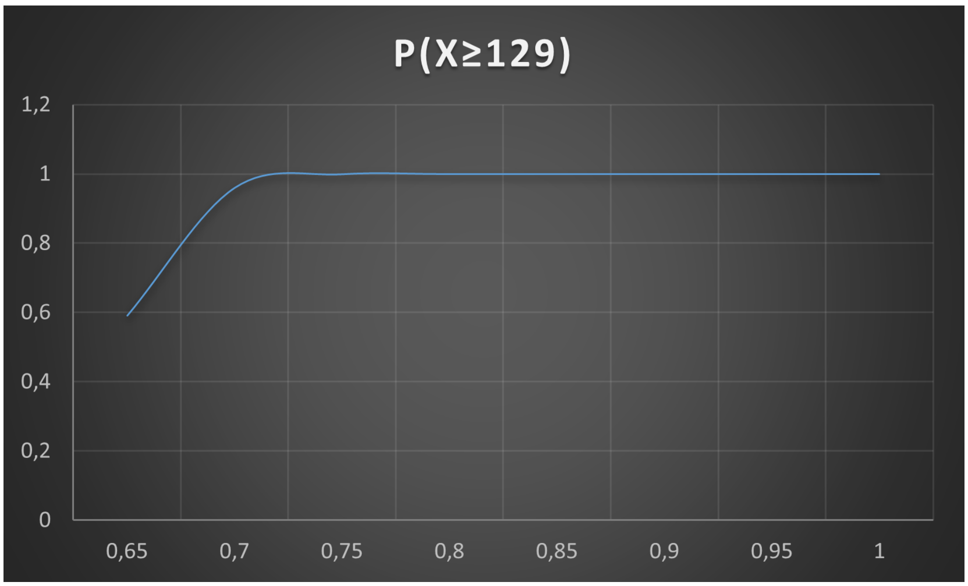
   The BINOM.DIST was used for n=129, trials=200 and p=0.65, the value of

$$P(X \geq 129) = 0.5906 = 59.1\%$$

3. Calculate the probability $P(X \geq c)$ for p=0.70, p=75, ... and sketch the graph of the probability $P(X \geq c)$, this is the power of the test as function of p>0.60. You should see a curve/function that increases.

By increasing the values of p from 0.65 to 1.0 at the interval of 0.5, there appears a graph in which the value of $P(X \geq 129)$ is increasing until 0.75 and after which it reaches the maximum probability of 1.
The values of the binomial distribution are given in the table:

p	$P(X \geq 129)$
0.65	0,59069
0.7	0,96037
0.75	0,9996
0.8	1
0.85	1
0.9	1
0.95	1
1.0	1

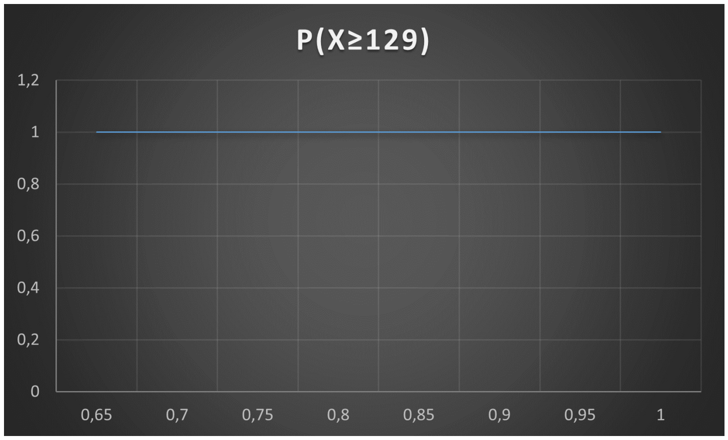


4. Imagine that we change the value of n. We take n=500 instead of n=200. The graph of the power will change as well. Indicate in which way the graph of the power will change and explain why.

   By changing the value of n=200 to n=500 and increase the value of p as in the previous case, we could see a graph which is a straight line at the probability of 1.This is due to the normal approximation of the binomial distribution that occurs as the value of n increases.

   The values of the binomial distribution are given in the table:

p	$P(X \geq 129)$
0.65	1
0.7	1
0.75	1
0.8	1
0.85	1
0.9	1
0.95	1
1.0	1



Exercise 5

$\alpha = 1\%$, $\overline{X} = 1.23$ and S =5.35

1. Determine the critical value c and determine whether we have to reject the null hypothesis in case of n=25, n=50 and n=100 .
   For n=25,

$$\begin{aligned} T &= \frac{\overline{X}}{S/\sqrt{n}} \\ &= \frac{1.23}{5.35/\sqrt{25}} \\ &= 1.14 \end{aligned}$$

   From the t-table, we get the value of c=2.492
   T<c, so we do not reject null hypothesis.

   For n=50,

$$\begin{aligned} T &= \frac{1.23}{5.35/\sqrt{50}} \\ &= 1.62 \end{aligned}$$

   From the t-table, we get the value of c=2.409
   T<c, so we do not reject null hypothesis.

   For n=100,

$$\begin{aligned} T &= \frac{1.23}{5.35/\sqrt{100}} \\ &= 2.29 \end{aligned}$$

   From the t-table, we get the value of c=2.365
   T<c, so we do not reject null hypothesis.

2. a. To verify that the null hypothesis is rejected for large n if T≥2.326,

$$P(-1.96 < Z < 1.96) = 0.95$$

If Z satisfies the above condition, the null hypothesis is not rejected.

$$\begin{aligned} Z &= \frac{T - \mu_T}{\sigma_T} \\ &= \frac{T - 0.5\sqrt{n}}{1} \\ &= 2.326 - 0.5\sqrt{n} \end{aligned}$$

The value of Z for increasing n values is reported in the table:

n	Z	null hypothesis
25	-0.174	Do not reject
50	-1.21	Do not reject
75	-2.004	Reject
100	-2.674	Reject

From the table, it is clear that we reject the null hypothesis for large n if $T \geq 2.326$

2. b. To determine the minimal value of n, Power = 95%, $\alpha = 1\%$, $\mu = 0.5\sqrt{n}$

$$\begin{aligned} Z &= \frac{T - \mu_T}{\sigma_T} \\ &= \frac{T - 0.5\sqrt{n}}{1} \\ &= 2.326 - 0.5\sqrt{n} \\ P(Z \geq 2.326 - 0.5\sqrt{n}) &= 0.95 \\ 1 - P(Z \leq 2.326 - 0.5\sqrt{n}) &= 0.95 \\ P(Z \leq 2.326 - 0.5\sqrt{n}) &= 0.05 \end{aligned}$$

Using the inverse standard normal distribution,

$$\begin{aligned} 2.326 - 0.5\sqrt{n} = 0 \\ n = 21.64 \end{aligned}$$

Therefore the minimal value of n for P(T≥2.326) = 0.95 and μ/σ=0.5 is 22.