Math 461 Fall 2020

Renming Song

University of Illinois at Urbana-Champaign

October 05, 2020
Outline

1. General Info

2. 5.4 Normal Random Variables
Test 1 is on Friday. There is no homework due on Friday. Topics covered in Test 1 include everything we covered in the first 4 Chapters, except Section 4.9 (which will be covered by Test 2 and the Final).

I will do a brief review on Wed and spend most of the lecture time Wed answering questions.
Test 1 is on Friday. There is no homework due on Friday. Topics covered in Test 1 include everything we covered in the first 4 Chapters, except Section 4.9 (which will be covered by Test 2 and the Final).

I will do a brief review on Wed and spend most of the lecture time Wed answering questions.
Important! Login to the scheduler at https://cbtf.engr.illinois.edu/sched - You should automatically be registered for the CBTF course the first time you sign in. Sign in early, even if there is nothing there for you to do yet because it gets you into the system. That lets you get the email notifications when there are exams to reserve or other reminders. Quite a few of you have not logged in yet.

Important! Review the student instructions at https://cbtf.engr.illinois.edu/cbtf-online/students This includes information about making reservations, setting up your phone for Zoom (required), getting help (there are daily CBTF student office hours on Zoom), and also setting your DRES accommodations so you get their extended time during the exams. Make sure you make a reservation for test 1. 27 of you have logged in to CBTF, but only 25 of you have reserved for test 1.
**Important!** Login to the scheduler at https://cbtf.engr.illinois.edu/sched - You should automatically be registered for the CBTF course the first time you sign in. Sign in early, even if there is nothing there for you to do yet because it gets you into the system. That lets you get the email notifications when there are exams to reserve or other reminders. Quite a few of you have not logged in yet.

**Important!** Review the student instructions at https://cbtf.engr.illinois.edu/cbtf-online/students This includes information about making reservations, setting up your phone for Zoom (required), getting help (there are daily CBTF student office hours on Zoom), and also setting your DRES accommodations so you get their extended time during the exams. Make sure you make a reservation for test 1. 27 of you have logged in to CBTF, but only 25 of you have reserved for test 1.
An important result in probability theory, known as the DeMoivre-Laplace central limit theorem, states that, when $n$ is large, a binomial random variable with parameters $(n, p)$ will have approximately the same distribution as a normal random variable with the same mean and variance.

**DeMoivre-Laplace central limit theorem**

If $S_n$ denotes the number of successes that occur when $n$ independent trials, each resulting in a success with probability $p$, are performed, then, for any $a < b$,

$$
\lim_{n \to \infty} P \left( a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \right) = \Phi(b) - \Phi(a).
$$
An important result in probability theory, known as the DeMoivre-Laplace central limit theorem, state that, when $n$ is large, a binomial random variable with parameters $(n, p)$ will have approximately the same distribution as a normal random variable with the same mean and variance.

**DeMoivre-Laplace central limit theorem**

If $S_n$ denotes the number of successes that occur when $n$ independent trials, each resulting in a success with probability $p$, are performed, then, for any $a < b$,

$$
\lim_{n \to \infty} P \left( a \leq \frac{S_n - np}{\sqrt{np(1 - p)}} \leq b \right) = \Phi(b) - \Phi(a).
$$
I will not prove this theorem now. I will give a proof of a more general result in Chapter 8.

The theorem above says that when $n$ is large enough, the distribution of

$$\frac{S_n - np}{\sqrt{np(1 - p)}}$$

is approximately standard normal. But how large is large enough?

In general, the normal approximation will very good for values of $n$ satisfying $np(1 - p) \geq 10$. 
I will not prove this theorem now. I will give a proof of a more general result in Chapter 8.

The theorem above says that when $n$ is large enough, the distribution of

$$\frac{S_n - np}{\sqrt{np(1 - p)}}$$

is approximately standard normal. But how large is large enough?

In general, the normal approximation will very good for values of $n$ satisfying $np(1 - p) \geq 10$. 
I will not prove this theorem now. I will give a proof of a more general result in Chapter 8.

The theorem above says that when $n$ is large enough, the distribution of

$$\frac{S_n - np}{\sqrt{np(1 - p)}}$$

is approximately standard normal. But how large is large enough?

In general, the normal approximation will very good for values of $n$ satisfying $np(1 - p) \geq 10$. 
Example 3

Let $X$ be the number of times that a fair coin, flipped 40 times, lands Heads. Find the probability that $X = 20$. Use normal approximation and then compare it with the exact value.

$$P(X = 20) = \binom{40}{20} \left(\frac{1}{2}\right)^{40} \approx 0.1254.$$

Normal approximation ($np(1-p) = 10$)

$$P(X = 20) = P\left(\frac{X - 20}{\sqrt{10}} = \frac{20 - 20}{\sqrt{10}}\right) = 0.$$

What is the problem?

We are using a continuous random variable to approximate an integer-valued random variable. We need “round” things up correctly!
Example 3

Let $X$ be the number of times that a fair coin, flipped 40 times, lands Heads. Find the probability that $X = 20$. Use normal approximation and then compare it with the exact value.

$$P(X = 20) = \binom{40}{20} \left(\frac{1}{2}\right)^{40} \approx 0.1254.$$  

Normal approximation ($np(1-p) = 10$)

$$P(X = 20) = P\left(\frac{X - 20}{\sqrt{10}} = \frac{20 - 20}{\sqrt{10}}\right) = 0.$$  

What is the problem?

We are using a continuous random variable to approximate an integer-valued random variable. We need “round” things up correctly!
Example 3

Let $X$ be the number of times that a fair coin, flipped 40 times, lands Heads. Find the probability that $X = 20$. Use normal approximation and then compare it with the exact value.

$$P(X = 20) = \binom{40}{20} \left(\frac{1}{2}\right)^{40} \approx 0.1254.$$ 

Normal approximation ($np(1 - p) = 10$)

$$P(X = 20) = P\left(\frac{X - 20}{\sqrt{10}} = \frac{20 - 20}{\sqrt{10}}\right) = 0.$$ 

What is the problem?

We are using a continuous random variable to approximate an integer-valued random variable. We need “round” things up correctly!
\[ P(X = 20) = P(19.5 \leq X < 20.5) \]
\[ = P\left(\frac{19.5 - 20}{\sqrt{10}} \leq \frac{X - 20}{\sqrt{10}} < \frac{20.5 - 20}{\sqrt{10}}\right) \]
\[ \approx P(-0.16 \leq \frac{X - 20}{\sqrt{10}} \leq 0.16) = \Phi(0.16) - \Phi(-0.16) = 0.1272. \]

The approximation is pretty good!

**Example 4**

The idea size of a first-year class in particular college is 150 students. Past experience shows that, on average, 30% of those accepted for admission will eventually attend the college. The college uses a policy of accepting 450 students. Find the probability that more than 150 first-year students will attend the college.
\[ P(X = 20) = P(19.5 \leq X < 20.5) = P\left(\frac{19.5 - 20}{\sqrt{10}} \leq \frac{X - 20}{\sqrt{10}} < \frac{20.5 - 20}{\sqrt{10}}\right) \approx P(-0.16 \leq \frac{X - 20}{\sqrt{10}} \leq 0.16) = \Phi(0.16) - \Phi(-0.16) = 0.1272. \]

The approximation is pretty good!

**Example 4**

The idea size of a first-year class in particular college is 150 students. Past experience shows that, on average, 30% of those accepted for admission will eventually attend the college. The college uses a policy of accepting 450 students. Find the probability that more than 150 first-year students will attend the college.
\[ P(X = 20) = P(19.5 \leq X < 20.5) = P\left(\frac{19.5 - 20}{\sqrt{10}} \leq \frac{X - 20}{\sqrt{10}} < \frac{20.5 - 20}{\sqrt{10}}\right) \]

\[ \approx P(-0.16 \leq \frac{X - 20}{\sqrt{10}} \leq 0.16) = \Phi(0.16) - \Phi(-0.16) = 0.1272. \]

The approximation is pretty good!

**Example 4**

The idea size of a first-year class in particular college is 150 students. Past experience shows that, on average, 30% of those accepted for admission will eventually attend the college. The college uses a policy of accepting 450 students. Find the probability that more than 150 first-year students will attend the college.
Let $X$ be the number of first-year students attending the college. Then $X$ is a binomial random variable with parameters $(450, 0.3)$. Thus

$$P(X > 150) = P(X \geq 150.5)$$

$$= P\left( \frac{X - 450 \cdot 0.3}{\sqrt{450 \cdot 0.3 \cdot 0.7}} \geq \frac{150.5 - 450 \cdot 0.3}{\sqrt{450 \cdot 0.3 \cdot 0.7}} \right)$$

$$\approx P\left( \frac{X - 135}{\sqrt{450 \cdot 0.3 \cdot 0.7}} \geq 1.59 \right)$$

$$= 1 - \Phi(1.59) \approx 0.0559.$$}

Now I am going to give an application of the normal approximation to polling.
Let $X$ be the number of first-year students attending the college. Then $X$ is a binomial random variable with parameters $(450, 0.3)$. Thus

$$P(X > 150) = P(X \geq 150.5)$$

$$= P\left( \frac{X - 450 \cdot 0.3}{\sqrt{450 \cdot 0.3 \cdot 0.7}} \geq \frac{150.5 - 450 \cdot 0.3}{\sqrt{450 \cdot 0.3 \cdot 0.7}} \right)$$

$$\approx P\left( \frac{X - 135}{\sqrt{450 \cdot 0.3 \cdot 0.7}} \geq 1.59 \right)$$

$$= 1 - \Phi(1.59) \approx 0.0559.$$
Example 5

A sample of size $n$ is taken to determine the percentage of the population planning to vote for a certain candidate in an upcoming election. Let $X_k = 1$ if the $k$-th person sampled plans to vote for the candidate and $X_k = 0$ otherwise. We assume that $X_1, \ldots, X_k$ are independently and identically distributed with

$$P(X_1 = 1) = p, \quad P(X_1 = 0) = 1 - p.$$ 

Assume that the election is not lopsided so that $\sqrt{p(1 - p)}$ is close to $1/2$. (If $p \in (0.3, 0.7)$, then $\sqrt{p(1 - p)} \geq 0.458$.)

Let $S_n = X_1 + \cdots + X_n$. Then $S_n/n$ denotes the fraction of the people sampled plan to vote for the candidate and can be used to estimate the true but unknown probability $p$. 
Example 5

A sample of size $n$ is taken to determine the percentage of the population planning to vote for a certain candidate in an upcoming election. Let $X_k = 1$ if the $k$-th person sampled plans to vote for the candidate and $X_k = 0$ otherwise. We assume that $X_1, \ldots, X_k$ are independently and identically distributed with

$$P(X_1 = 1) = p, \quad P(X_1 = 0) = 1 - p.$$ 

Assume that the election is not lopsided so that $\sqrt{p(1 - p)}$ is close to $1/2$. (If $p \in (0.3, 0.7)$, then $\sqrt{p(1 - p)} \geq 0.458$.)

Let $S_n = X_1 + \cdots + X_n$. Then $S_n/n$ denotes the fraction of the people sampled plan to vote for the candidate and can be used to estimate the true but unknown probability $p$. 
(a) Suppose \( n = 900 \). Find \( P(\left| \frac{S_n}{n} - p \right| \geq 0.025) \). (b) Suppose \( n = 900 \). Find \( c \) so that \( P(\left| \frac{S_n}{n} - p \right| \geq c) = 0.01 \). (c) Find \( n \) such that \( P(\left| \frac{S_n}{n} - p \right| \geq 0.025) = 0.01 \).

\[
P(\left| \frac{S_n}{n} - p \right| \geq c) = P(S_n \leq np - cn) + P(S_n \geq np + cn) \]

\[
= P\left( \frac{S_n - np}{\sqrt{np(1-p)}} \leq -\frac{cn}{\sqrt{np(1-p)}} \right) + P\left( \frac{S_n - np}{\sqrt{np(1-p)}} \geq \frac{cn}{\sqrt{np(1-p)}} \right) \]

\[
\approx P(Z < -2c\sqrt{n}) + P(Z > 2c\sqrt{n}) = 2(1 - \Phi(2c\sqrt{n})).
\]
(a) Suppose $n = 900$. Find $P(\left| \frac{S_n}{n} - p \right| \geq 0.025)$. (b) Suppose $n = 900$. Find $c$ so that $P(\left( \left| \frac{S_n}{n} - p \right| \geq c \right) = 0.01)$. (c) Find $n$ such that $P(\left( \left| \frac{S_n}{n} - p \right| \geq 0.025 \right) = 0.01$.

\[
P(\left| \frac{S_n}{n} - p \right| \geq c) = P(S_n \leq np - cn) + P(S_n \geq np + cn)
\]

\[
= P\left( \frac{S_n - np}{\sqrt{np(1 - p)}} \leq - \frac{cn}{\sqrt{np(1 - p)}} \right) + P\left( \frac{S_n - np}{\sqrt{np(1 - p)}} \geq \frac{cn}{\sqrt{np(1 - p)}} \right)
\]

\[
\approx P(Z < -2c\sqrt{n}) + P(Z > 2c\sqrt{n})
\]

\[
= 2(1 - \Phi(2c\sqrt{n})).
\]
(a) \[ P\left( \left| \frac{S_{900}}{900} - p \right| \geq 0.025 \right) \approx 2(1 - \Phi(1.5)) \approx 0.134. \]

(b) Since

\[ P\left( \left| \frac{S_{900}}{900} - p \right| \geq c \right) \approx 2(1 - \Phi(60c)), \]

in order for

\[ P\left( \left| \frac{S_{900}}{900} - p \right| \geq c \right) = 0.01, \]

we must have

\[ 2(1 - \Phi(60c)) = 0.01. \]

That is

\[ \Phi(60c) = 0.995. \]

Thus \( 60c = 2.58 \) and hence \( c = 0.043. \)
(a) \[ P\left(\left| \frac{S_{900}}{900} - p \right| \geq 0.025 \right) \approx 2(1 - \Phi(1.5)) \approx 0.134. \]

(b) Since \[ P\left(\left| \frac{S_{900}}{900} - p \right| \geq c \right) \approx 2(1 - \Phi(60c)), \]
in order for \[ P\left(\left| \frac{S_{900}}{900} - p \right| \geq c \right) = 0.01, \]
we must have \[ 2(1 - \Phi(60c)) = 0.01. \]
That is \[ \Phi(60c) = 0.995. \]
Thus \( 60c = 2.58 \) and hence \( c = 0.043. \)
Since
\[ P\left( \left| \frac{S_n}{n} - p \right| \geq 0.025 \right) \approx 2(1 - \Phi(0.05\sqrt{n})), \]
in order for
\[ P\left( \left| \frac{S_n}{n} - p \right| \geq 0.025 \right) = 0.01, \]
we must have
\[ 2(1 - \Phi(0.05\sqrt{n})) = 0.01. \]
So
\[ 0.05\sqrt{n} = 2.58 \]
and
\[ n = 2663. \]