

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

Math 1710 Class 18

Sampling Distributions, CLT
Dr. Back

Oct. 7, 2009

Quantitative Case of the CLT

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

- Suppose X is a random variable with mean μ and standard deviation σ .
- Let X_1, X_2, \dots, X_n be n independent copies of X .

- Set

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

- Then

- 1 The mean of \bar{X} is μ .
- 2 The standard deviation of \bar{X} is $\frac{\sigma}{\sqrt{n}}$.
- 3 \bar{X} is approximately normal as n gets large.

Quantitative CLT - Why?

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

- 1 The mean of \bar{X} is μ .
- 2 The standard deviation of \bar{X} is $\frac{\sigma}{\sqrt{n}}$.
- 3 \bar{X} is approximately normal as n gets large.

Quantitative CLT - Why?

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

- 1 The mean of \bar{X} is μ .
- 2 The standard deviation of \bar{X} is $\frac{\sigma}{\sqrt{n}}$.
- 3 \bar{X} is approximately normal as n gets large.

(1) is immediate from

$$E(X_1 + X_2 + \dots + X_n) = nE(X)$$

so

$$E(\bar{X}) = \frac{nE(X)}{n} = E(X).$$

Quantitative CLT - Why?

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

- 1 The mean of \bar{X} is μ .
- 2 The standard deviation of \bar{X} is $\frac{\sigma}{\sqrt{n}}$.
- 3 \bar{X} is approximately normal as n gets large.

(2) is immediate from

$$\text{Var}(X_1 + X_2 + \dots + X_n) = n\text{Var}(X)$$

so

$$\text{Var}(\bar{X}) = \frac{n\text{Var}(X)}{n^2} = \frac{\text{Var}(X)}{n}.$$

And so the std. dev of \bar{X} is $\frac{\sigma}{\sqrt{n}}$.

Quantitative CLT - Why?

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

- 1 The mean of \bar{X} is μ .
- 2 The standard deviation of \bar{X} is $\frac{\sigma}{\sqrt{n}}$.
- 3 \bar{X} is approximately normal as n gets large.

(3) is much deeper! We will illustrate, but it's only barely possible to prove in even an advanced undergraduate course. Your text uses the notation

$$SD(\bar{x}) = \frac{\sigma}{n}$$

for the exact standard deviation of the sampling distribution of \bar{x} and $SE(\bar{x}) = \frac{\hat{\sigma}}{n}$ for its approximation, the standard error of \bar{x} .

Underweight Babies?

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(2) At birth, babies average 7.8 pounds with a standard deviation of 2.1 pounds.

34 Babies born near a large possibly polluting factory average 7.2 pounds.

Underweight Babies?

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(2) At birth, babies average 7.8 pounds with a standard deviation of 2.1 pounds.

34 Babies born near a large possibly polluting factory average 7.2 pounds.

Is that unusually low?

Underweight Babies?

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(2) At birth, babies average 7.8 pounds with a standard deviation of 2.1 pounds.

34 Babies born near a large possibly polluting factory average 7.2 pounds.

Is that unusually low?

Solution: In fact 34 is not large enough to have confidence that an average of 34 copies will be approx. normal. Let's work the problem under the assumption that it is in this case.

(The usual rules of thumb have a cutoff of 40 or 50.)

Underweight Babies?

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(2) At birth, babies average 7.8 pounds with a standard deviation of 2.1 pounds.

34 Babies born near a large possibly polluting factory average 7.2 pounds.

Is that unusually low?

If so, then an RV X describing the avg of 34 observations would follow:

$$\bar{X} \sim N\left(7.8, \frac{2.1}{\sqrt{34}}\right) = N(7.8, .36).$$

Underweight Babies?

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(2) At birth, babies average 7.8 pounds with a standard deviation of 2.1 pounds.

34 Babies born near a large possibly polluting factory average 7.2 pounds.

Is that unusually low?

$$\bar{X} \sim N\left(7.8, \frac{2.1}{\sqrt{34}}\right) = N(7.8, .36).$$

The z score of 7.2 is $\frac{-0.6}{.36} = -1.67$ and $P(Z < -1.67) = .0485$.

Underweight Babies?

(2) At birth, babies average 7.8 pounds with a standard deviation of 2.1 pounds.

34 Babies born near a large possibly polluting factory average 7.2 pounds.

Is that unusually low?

$$\bar{X} \sim N\left(7.8, \frac{2.1}{\sqrt{34}}\right) = N(7.8, .36).$$

The z score of 7.2 is $\frac{-6}{.36} = -1.67$ and $P(Z < -1.67) = .0485$.

So (under our additional assumption that averaging 34 is enough for approximate normality), about 5% of the time a sample of 34 babies would average as low as 7.2 pounds or lower just by chance.

The evidence is not conclusive.

SAT Scores of Group Averages

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(3) Suppose SAT's have a mean of 500, standard deviation of 100 and are approximately normal.
Form means of samples of 20 students.

SAT Scores of Group Averages

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(3) Suppose SAT's have a mean of 500, standard deviation of 100 and are approximately normal.

Form means of samples of 20 students.

What distribution do these follow? What is the chance of a sample averaging over 600? Over 550?

SAT Scores of Group Averages

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(3) Suppose SAT's have a mean of 500, standard deviation of 100 and are approximately normal.

Form means of samples of 20 students.

What distribution do these follow? What is the chance of a sample averaging over 600? Over 550?

Solution:

$n = 20$ is too small to assume an average is automatically normal. But since the sum of independent normal RV's is normal, in this case we do have approximate normality.

SAT Scores of Group Averages

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(3) Suppose SAT's have a mean of 500, standard deviation of 100 and are approximately normal.

Form means of samples of 20 students.

What distribution do these follow? What is the chance of a sample averaging over 600? Over 550?

Solution:

Thus the RV \bar{X} describing means of 20 students satisfies

$$X \sim N\left(500, \frac{100}{\sqrt{20}}\right) = N(500, 22.36).$$

SAT Scores of Group Averages

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(3) Suppose SAT's have a mean of 500, standard deviation of 100 and are approximately normal.

Form means of samples of 20 students.

What distribution do these follow? What is the chance of a sample averaging over 600? Over 550?

Solution:

Thus the RV \bar{X} describing means of 20 students satisfies

$$X \sim N\left(500, \frac{100}{\sqrt{20}}\right) = N(500, 22.36).$$

The z score of 600 is $\frac{100}{22.36} = 4.47$.

$$P(\bar{X} > 600) = P(Z > 4.47) < .0001.$$

SAT Scores of Group Averages

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(3) Suppose SAT's have a mean of 500, standard deviation of 100 and are approximately normal.

Form means of samples of 20 students.

What distribution do these follow? What is the chance of a sample averaging over 600? Over 550?

Solution:

The z score of 600 is $\frac{100}{22.36} = 4.47$.

$$P(\bar{X} > 600) = P(Z > 4.47) < .0001.$$

Very tiny.

An individual has a good chance of getting over 600 but an average of 20 randomly chosen people does not.

SAT Scores of Group Averages

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(3) Suppose SAT's have a mean of 500, standard deviation of 100 and are approximately normal.

Form means of samples of 20 students.

What distribution do these follow? What is the chance of a sample averaging over 600? Over 550?

Solution:

The z score of 550 is $\frac{50}{22.36} = 2.23$.

$$P(\bar{X} > 550) = P(Z > 2.23) = P(Z < -2.23) = .0129.$$

SAT Scores of Group Averages

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(3) Suppose SAT's have a mean of 500, standard deviation of 100 and are approximately normal.

Form means of samples of 20 students.

What distribution do these follow? What is the chance of a sample averaging over 600? Over 550?

Good intuition why the standard deviation of an average \bar{X} is less than that of the RV X describing one individual:

SAT Scores of Group Averages

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(3) Suppose SAT's have a mean of 500, standard deviation of 100 and are approximately normal.

Form means of samples of 20 students.

What distribution do these follow? What is the chance of a sample averaging over 600? Over 550?

Good intuition why the standard deviation of an average \bar{X} is less than that of the RV X describing one individual:

Since 600 is one standard deviation above the mean, there is about a 16% chance for one individual to be above 600.

SAT Scores of Group Averages

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(3) Suppose SAT's have a mean of 500, standard deviation of 100 and are approximately normal.

Form means of samples of 20 students.

What distribution do these follow? What is the chance of a sample averaging over 600? Over 550?

Good intuition why the standard deviation of an average \bar{X} is less than that of the RV X describing one individual:

Since 600 is one standard deviation above the mean, there is about a 16% chance for one individual to be above 600.

But for an average of 20 to be over 600, most of the 20 will likely have to achieve this 16% challenge.

SAT Scores of Group Averages

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(3) Suppose SAT's have a mean of 500, standard deviation of 100 and are approximately normal.

Form means of samples of 20 students.

What distribution do these follow? What is the chance of a sample averaging over 600? Over 550?

Good intuition why the standard deviation of an average \bar{X} is less than that of the RV X describing one individual:

Since 600 is one standard deviation above the mean, there is about a 16% chance for one individual to be above 600.

But for an average of 20 to be over 600, most of the 20 will likely have to achieve this 16% challenge.

The multiplication rule says this is quite unlikely.

SAT Scores of Group Averages

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

(3) Suppose SAT's have a mean of 500, standard deviation of 100 and are approximately normal.

Form means of samples of 20 students.

What distribution do these follow? What is the chance of a sample averaging over 600? Over 550?

Good intuition why the standard deviation of an average \bar{X} is less than that of the RV X describing one individual:

Since 600 is one standard deviation above the mean, there is about a 16% chance for one individual to be above 600.

But for an average of 20 to be over 600, most of the 20 will likely have to achieve this 16% challenge.

The multiplication rule says this is quite unlikely.

Our RV calculation

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

makes this intuition precise.

CLT Simulation Notes

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

Start by realizing 500 trials from a uniform $[0,1]$ distribution.
(Mean=.5, Standard Deviation= $\sqrt{1/12}$ =.289)

CLT Simulation Notes

Math 1710
Class 18

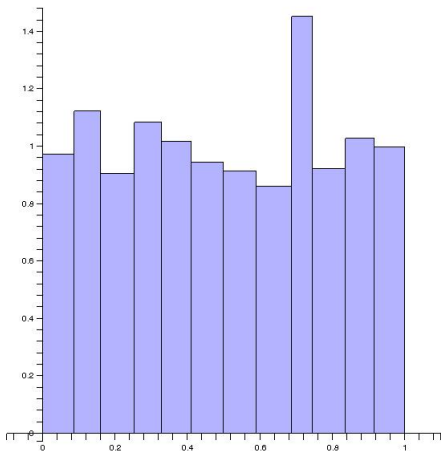
V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

Uniform Distribution



CLT Simulation Notes

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

Probability Histograms - area of blue rectangle is the fraction of the results falling within the range at the base.

CLT Simulation Notes

Math 1710
Class 18

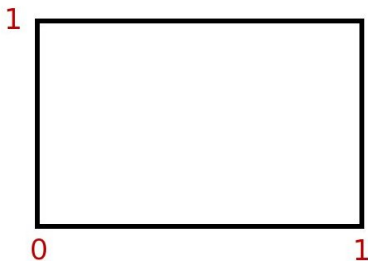
V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

A Theoretical Uniform Distribution



CLT Simulation Notes

Math 1710
Class 18

V2u

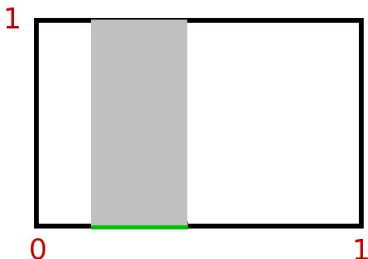
Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

Total area 1 and shaded area (gray) equal to prob. of a range of values (green) along the x-axis.

A Theoretical Uniform Distribution



CLT Simulation Notes

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

Now square each value to get the simulation of a new distribution. (Mean=.333, Standard Deviation=.298)

CLT Simulation Notes

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

Now square each value to get the simulation of a new distribution. (Mean=.333, Standard Deviation=.298)
Compare with a normal distribution $N(.333,.298)$.

CLT Simulation Notes

Math 1710
Class 18

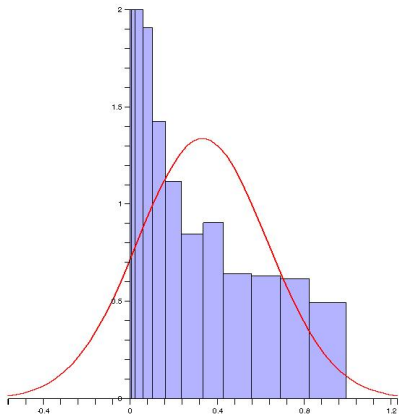
V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

Square of a Uniform Distribution



CLT Simulation Notes

Math 1710
Class 18

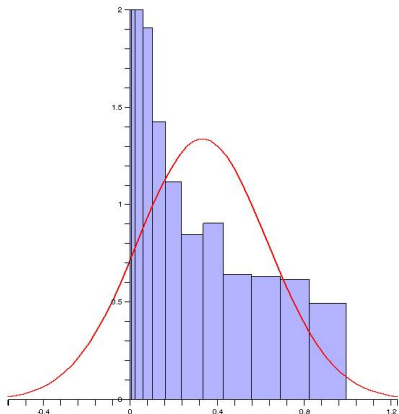
V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

Square of a Uniform Distribution



A poor match!

CLT Pictures

Math 1710
Class 18

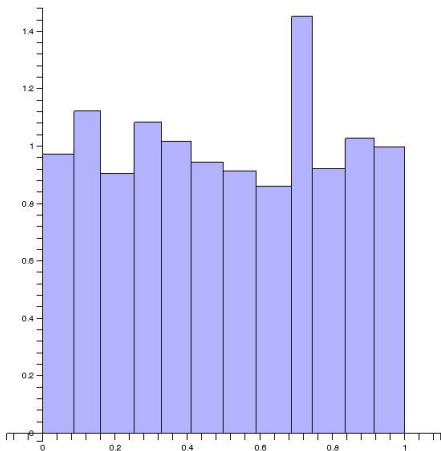
V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

Simulation of a Uniform Distribution Uniform Distribution



CLT Pictures

Math 1710
Class 18

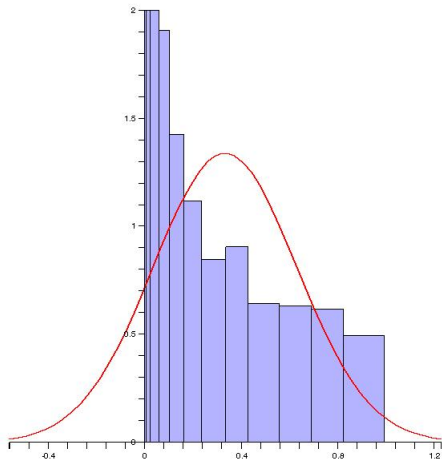
V2u

Simulation of Square of a Uniform Distribution Square of a Uniform Distribution

Last Time

Quant. CLT
Problems

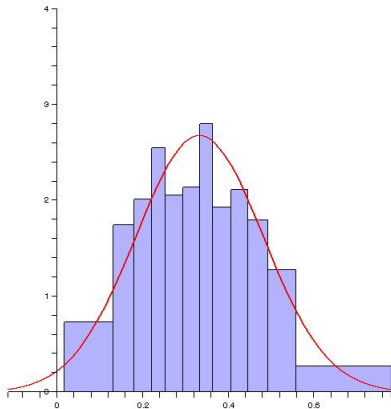
Normal
Arising in
Averages



CLT Pictures

Now do the above four times and compare the averages of each of 500 sets of four values with $N(.333, .298/2)$.

Average of Four



Math 1710
Class 18

V2u

Last Time

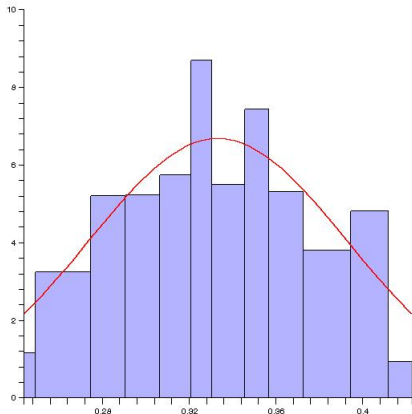
Quant. CLT
Problems

Normal
Arising in
Averages

CLT Pictures

Then 25 times and compare the averages of each of 500 sets of 25 values with $N(.333, .298/5)$.

Average of 25



Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

CLT Pictures

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

As we compare the different averages, note that the horizontal scale is decreasing.

CLT Pictures

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages

For averages of four values, the part of the graph shown is from below 0 to .78.

CLT Pictures

Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

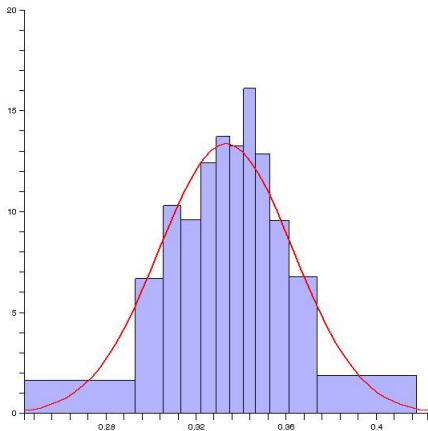
Normal
Arising in
Averages

For averages of 25 values, the part of the graph shown is from .25 to .43.

CLT Pictures

Now 100 times and compare the averages of each of 500 sets of 100 values with $N(.333, .298/10)$.

Average of 100



Math 1710
Class 18

V2u

Last Time

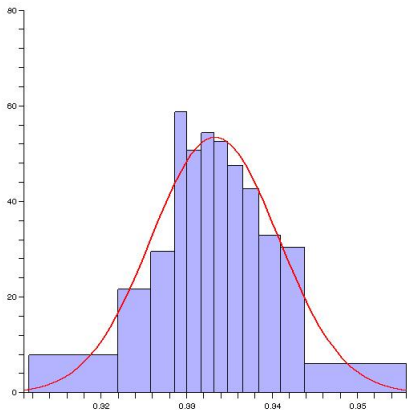
Quant. CLT
Problems

Normal
Arising in
Averages

CLT Pictures

1600 times.

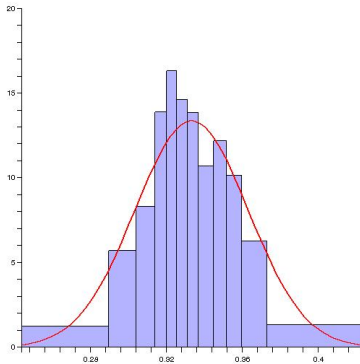
Average of 1600



CLT Pictures

A normal by comparison.

Average of 100 Normal



Math 1710
Class 18

V2u

Last Time

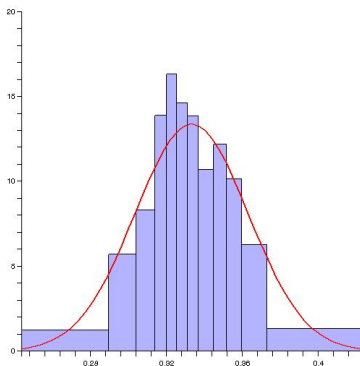
Quant. CLT
Problems

Normal
Arising in
Averages

CLT Pictures

A normal by comparison.

Average of 100 Normal

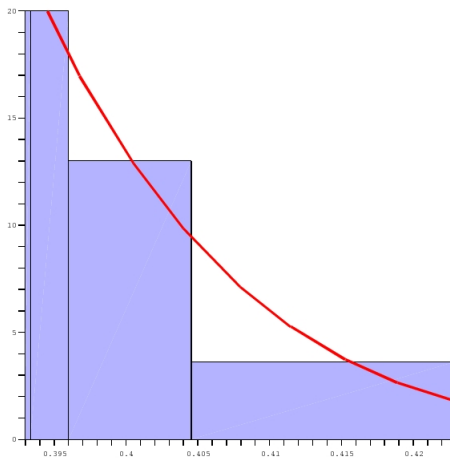


This suggests the parts we see not matching are more due to simulation than non-normality.

CLT Pictures

Fit near the ends.

Probability Histogram of Last41 Observations Scaled to Norm 1



Math 1710
Class 18

V2u

Last Time

Quant. CLT
Problems

Normal
Arising in
Averages