Bi(Multi)variate Transformation of Variables
(continued)

Daniel B. Rowe, Ph.D.

Professor
Department of Mathematical and Statistical Sciences

MARQUETTE UNIVERSITY
Be The Difference.

Outline

- **Distribution of Mean of RVs**
  Normal and Uniform

- **Central Limit Theorem (CLT)**
  Uniform vs. Normal

Bivariate Change of Variable - Average

Talked about x_1 with PDF $f_{X_1}(x_1 \mid \theta_1)$, and x_2 with PDF $f_{X_2}(x_2 \mid \theta_2)$, then the PDF of $y_1 = x_1 + x_2$ can be found via the bivariate change of variable technique

$$f_{Y_1,Y_2}(y_1, y_2 \mid \theta) = f_{X_1,X_2}(x_1(y_1, y_2), x_2(y_1, y_2) \mid \theta) \times |J(x_1, x_2 \rightarrow y_1, y_2)|$$

with marginalization $f_{Y_1}(y_1 \mid \theta) = \int_{y_2} f_{Y_1,Y_2}(y_1, y_2 \mid \theta) dy_2$.

Bivariate Change of Variable - Average

Example: Normal

Let x_1~normal(μ_1,σ_1^2) and x_2~normal(μ_2,σ_2^2). x_1 & x_2 independent. The joint PDF of (x_1,x_2) is

$$f(x_1,x_2\,|\,\mu_1,\sigma_1^2,\mu_2,\sigma_2^2)=\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2}\frac{1}{\sigma_2\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}.$$

With $x_1=y_1-y_2$, $x_2=y_2$, $|J(x_1,x_2\rightarrow y_1,y_2)|=1$,

$$f_{Y_1,Y_2}(y_1,y_2\,|\,\theta)=\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y_1-y_2-\mu_1}{\sigma_1}\right)^2}\frac{1}{\sigma_2\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y_2-\mu_2}{\sigma_2}\right)^2}\times 1$$

We found that $y_1\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

Bivariate Change of Variable - Average

This change of variable technique can be repeated.

If x_3~normal(μ_3, σ_3^2) , then if we let $y_3 = x_3 + y_1$
(don't forget $y_1 = x_1 + x_2$),

then we can find that $y_3 \sim N(\mu_1 + \mu_2 + \mu_3, \sigma_1^2 + \sigma_2^2 + \sigma_3^2)$

we can repeat the procedure to get $y_n \sim N\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2\right)$.

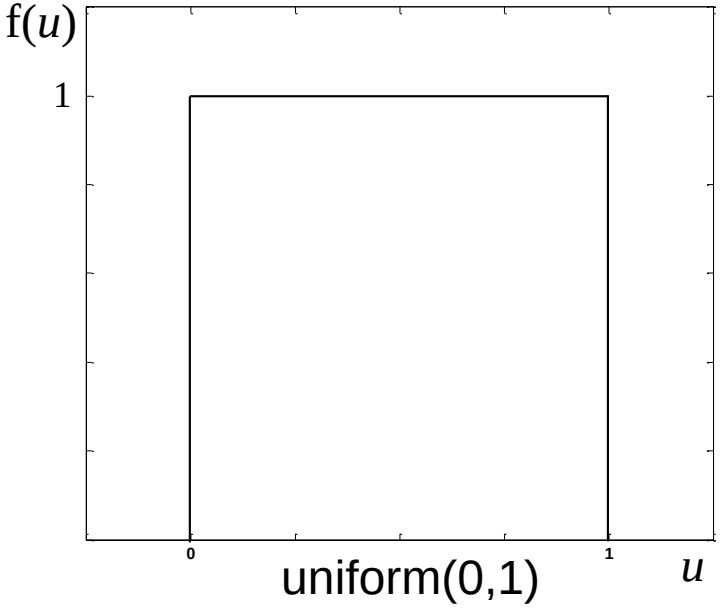
We can also find that $y = \frac{1}{n}\sum_{i=1}^{n} x_i \sim N\left(\frac{1}{n}\sum_{i=1}^{n}\mu_i, \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\right)$.

Bivariate Change of Variable - Average

The change of variable technique can be used to find the distribution of the sum or average of two uniform RVs.

Let u_1~uniform(0,1) and u_2~uniform(0,1).
The joint PDF of (u_1, u_2) is

$$f(u_1,u_2) = \begin{cases} 1 & \text{if} \quad u_1 \in [0,1] \text{ and } u_2 \in [0,1] \\ 0 & \text{if} \quad u_1 \notin [0,1] \text{ or } u_2 \notin [0,1] \end{cases}$$



Bivariate Change of Variable - Average

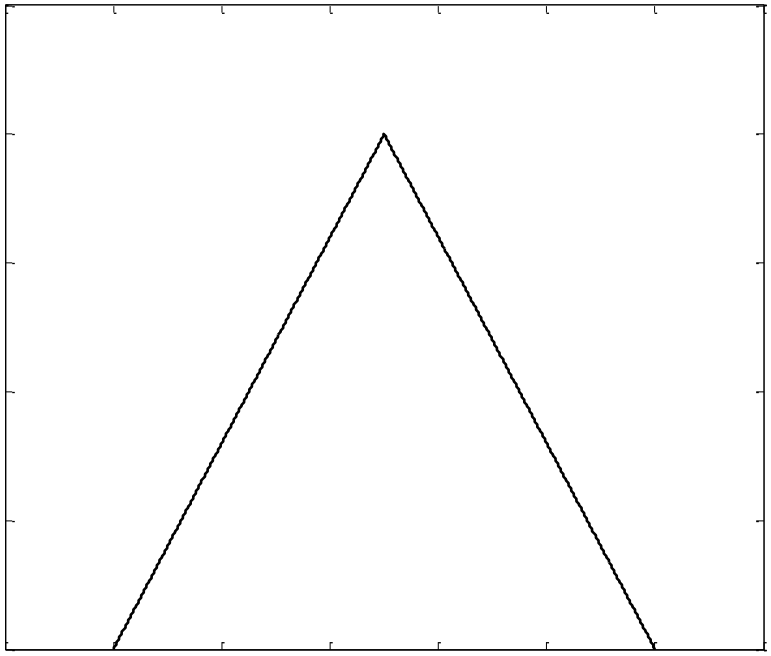
Using the change of variable technique

$$f_{Y_1,Y_2}(y_1,y_2\,|\,\theta) = f_{U_1,U_2}(u_1(y_1,y_2),u_2(y_1,y_2)\,|\,\theta)\times|\,J(u_1,u_2 \rightarrow y_1,y_2)\,|$$

$$f_{Y_1}(y_1\,|\,\theta) = \int_{y_2} f_{Y_1,Y_2}(y_1,y_2\,|\,\theta)dy_2 \ ,$$

the distribution of $y_1=u_1+u_2$ is

$$f(y_1) = \begin{cases} 0 & \text{if} \quad y_1 < 0 \\ y_1 & \text{if} \quad 0 \le y_1 < 1 \\ 2-y_1 & \text{if} \quad 1 \le y_1 \le 2 \\ 0 & \text{if} \quad y_1 > 2 \end{cases} .$$



triangular(0,2)

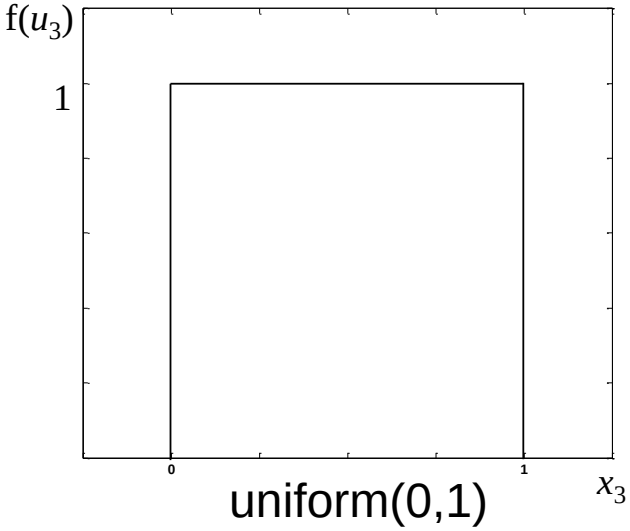
Bivariate Change of Variable - Average

This change of variable technique can be applied again, u_3~uniform(0,1), and the distribution of $y_3=y_1+u_3$ found

$$f_{Y_3,Y_4}(y_3,y_4\,|\,\theta)=f_{Y_3,Y_4}(y_1(y_3,y_4),u_3(y_3,y_4)\,|\,\theta)\times|\,J(y_1,u_3\rightarrow y_3,y_4)\,|$$

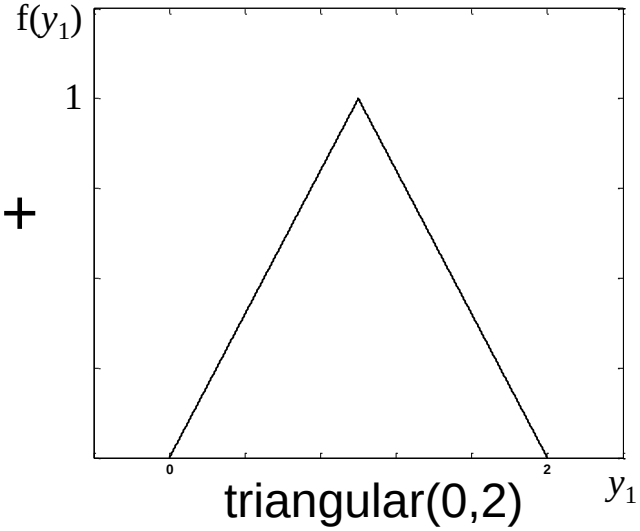
$$f_{Y_3}(y_3\,|\,\theta)=\int_{y_4} f_{Y_3,Y_4}(y_3,y_4\,|\,\theta)dy_4$$

y_4 is another variable not of interest for the bivariate change of variable



uniform(0,1)

RV +



triangular(0,2)

RV = ? RV

Homework.

Bivariate Change of Variable - Average

This change of variable technique can be repeated many times to determine the distribution of $y=(u_1+u_2+u_3+\ldots+u_n)/n$ where each u_i is uniform(a,b).

For large n, the mean becomes normally distributed with

$$\mu_{\bar{x}} = \mu, \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad .$$

$$\mu = \frac{b-a}{2}$$

$$\bar{x} \sim N(\mu = \mu_{\bar{x}}, \sigma^2_{\bar{x}} = \sigma^2/n)$$

$$\sigma^2 = \frac{(b-a)^2}{12}$$

The Central Limit Theorem (CLT)

Assume that we have a population with mean μ and standard deviation σ.

If we take random samples of size n, for large n, the distribution of the sample means is approximately normally distributed with

$\mu_{\bar{x}} = \mu, \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. Often $n \geq 30$ is considered large.

Often don't know the distribution of the individual observations, but know by CLT that the mean is approximately normal,

$\bar{x} \sim N(\mu = \mu_{\bar{x}}, \sigma^2_{\bar{x}} = \sigma^2 / n)$.

The Central Limit Theorem (CLT)

Two continuous data examples, $\mu = 100$ and $\sigma = 57.73$.

Generate data $x_1, ..., x_n$, for n=1, 2, 3, 4, 5, 15, 30, and 50. Calculate $\bar{x}$, and repeat one million times.

Uniform distribution, [a=0,b=200]. μ=100,σ=57.7

$$f(x) = \frac{1}{b-a} \qquad a \le x \le b$$

Normal distribution, [μ=100,σ=57.7].

$$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \qquad -\infty \le x \le \infty$$

The Central Limit Theorem (CLT)

According to CLT, if we had a sample $x_1,...,x_n$ of size n=1, 2, 3, 4, 5, 15, 30, and 50 from Uniform distribution, [a=0,b=200].

Sample Size, n	Mean, $\mu_{\bar{x}}$	SD, $\sigma_{\bar{x}}$
1	100	57.7350
2	100	40.8248
3	100	33.3333
4	100	28.8675
5	100	25.8199
15	100	25.8199
30	100	14.9071
50	100	8.1650

$$f(x) = \frac{1}{b-a}$$

$$a \le x \le b$$

$$\int_a^b x f(x) dx = \frac{b-a}{2}$$

$$\int_a^b (x-\mu)^2 f(x) dx = \frac{(b-a)^2}{12}$$

The Central Limit Theorem (CLT)

According to CLT, if we had a sample $x_1,...,x_n$ of size n=1, 2, 3, 4, 5, 15, 30, and 50 from Normal distribution, [μ=100,σ=(200-0)/$\sqrt{12}$]

Sample Size, n	Mean, $\mu_{\bar{x}}$	SD, $\sigma_{\bar{x}}$
1	100	57.7350
2	100	40.8248
3	100	33.3333
4	100	28.8675
5	100	25.8199
15	100	25.8199
30	100	14.9071
50	100	8.1650

$$f(x)=\frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$$

$$-\infty \le x \le \infty$$

$$\int_{-\infty}^{\infty} xf(x)dx = \mu$$

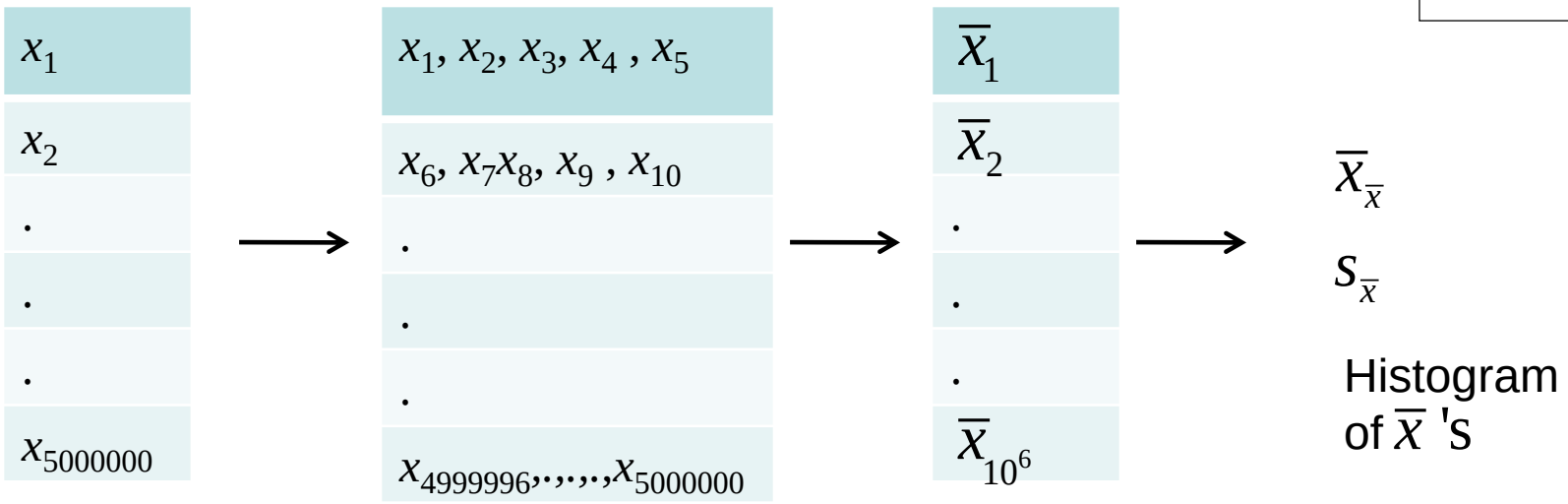
$$\int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \sigma^2$$

The Central Limit Theorem (CLT)

Using Matlab to generate n million observations $x_1,\ldots,x_{n\times10^6}$ from the Uniform [a=0,b=200] and Normal [μ=100,σ=57.7] distributions, for n=1, 2, 3, 4, 5, 15, 30, and 50.

8 data sets of Uniform and Normal

i.e. n=5, 5×10^6 Groups of n=5 Mean of groups



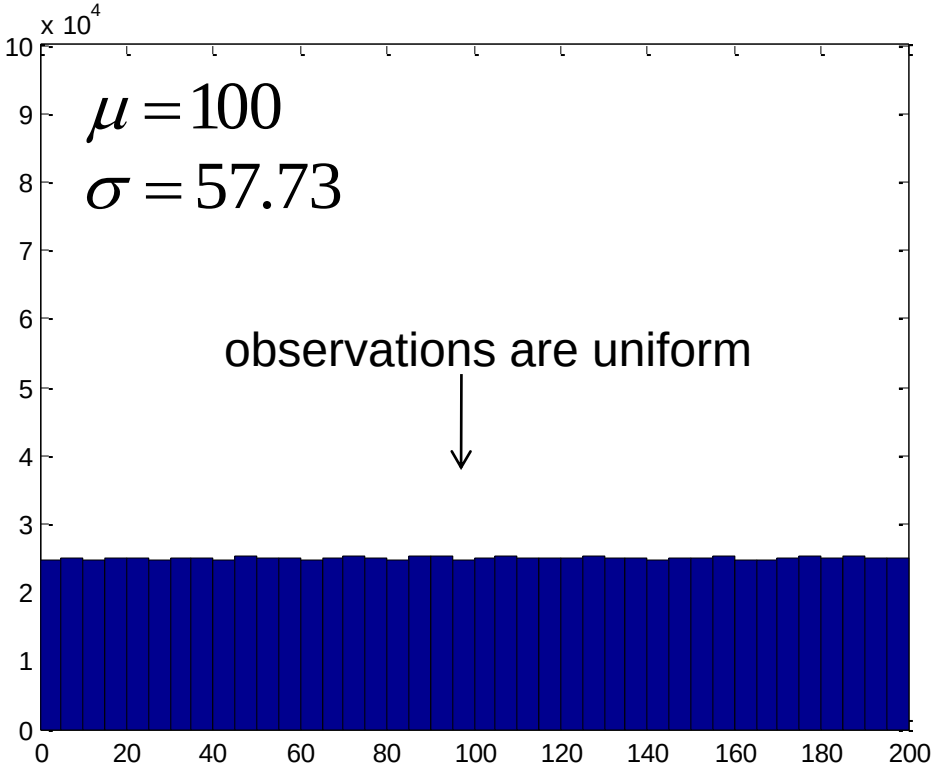
The Central Limit Theorem (CLT)

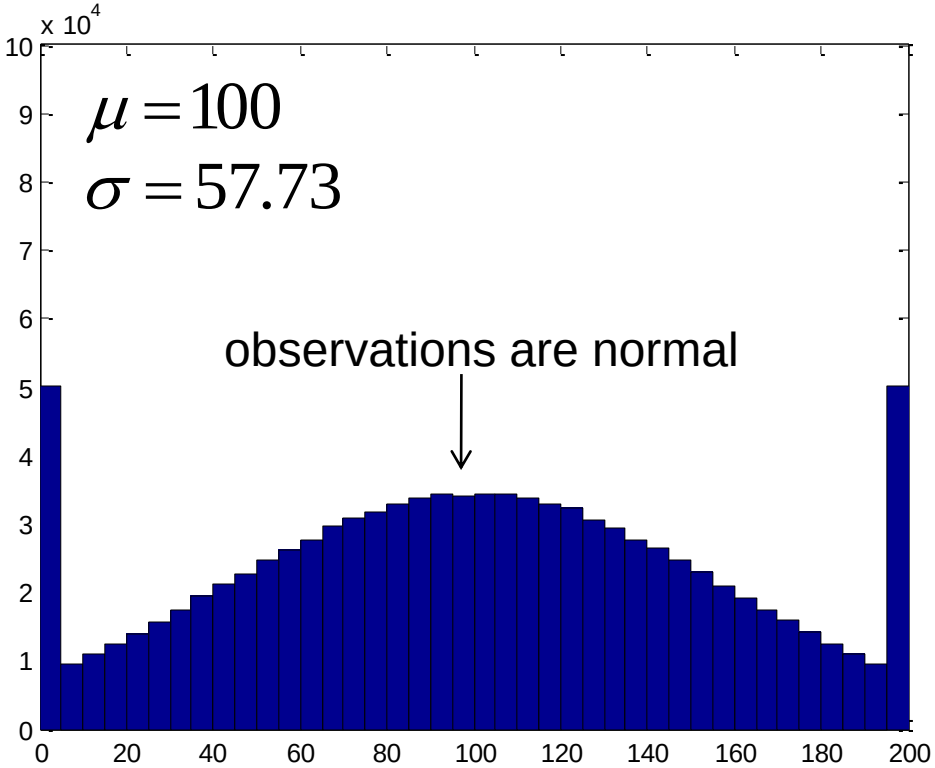
I wrote a program to make one million means $\bar{x}_1, ..., \bar{x}_{10^6}$ for n=1, 2, 3, 4, 5, 15, 30, and 50.

Sample Size n	Mean $\mu_{\bar{x}}$	Mean U $\bar{x}_{\bar{x}}$	Mean N $\bar{x}_{\bar{x}}$	SD $\sigma_{\bar{x}}$	SD U $s_{\bar{x}}$	SD N $s_{\bar{x}}$
1	100	100.0642	100.0077	57.7350	57.7071	57.7888
2	100	99.9828	100.0037	40.8248	40.8418	40.8206
3	100	99.9909	99.9627	33.3333	33.3418	33.2984
4	100	99.9559	100.0642	28.8675	28.8946	28.8126
5	100	100.0074	100.0320	25.8199	25.7865	25.8397
15	100	100.0134	99.9517	14.9071	14.9035	14.8918
30	100	99.9934	99.9836	10.5409	10.5335	10.5352
50	100	99.9918	99.9890	8.1650	8.1605	8.1709

The Central Limit Theorem (CLT)

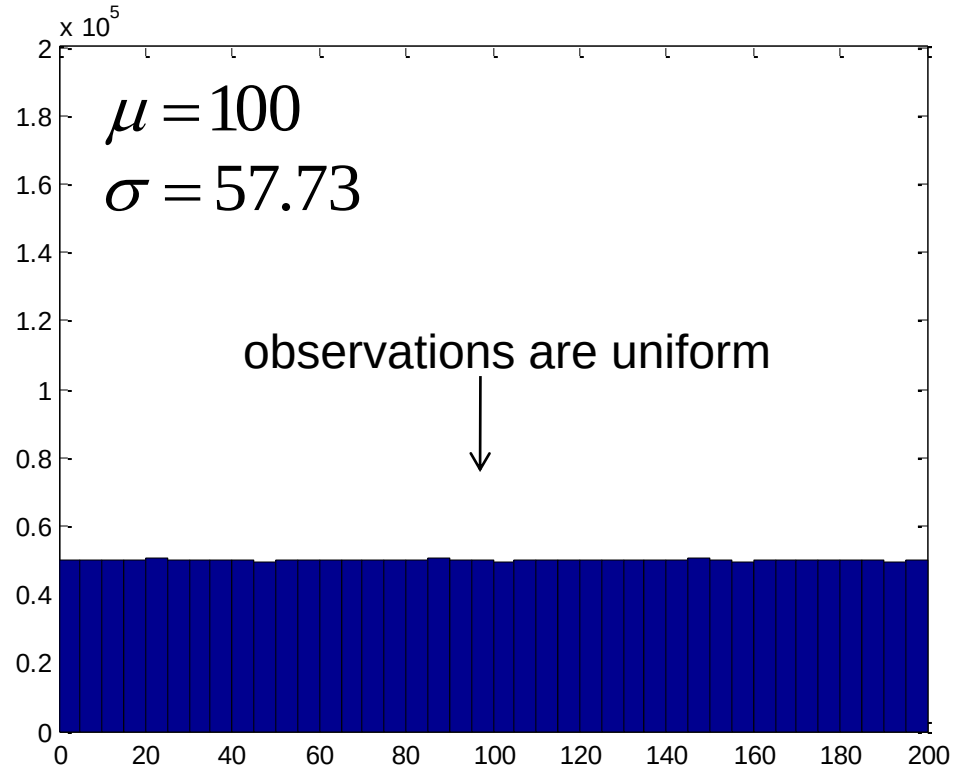
n=1 1×10^6 observations

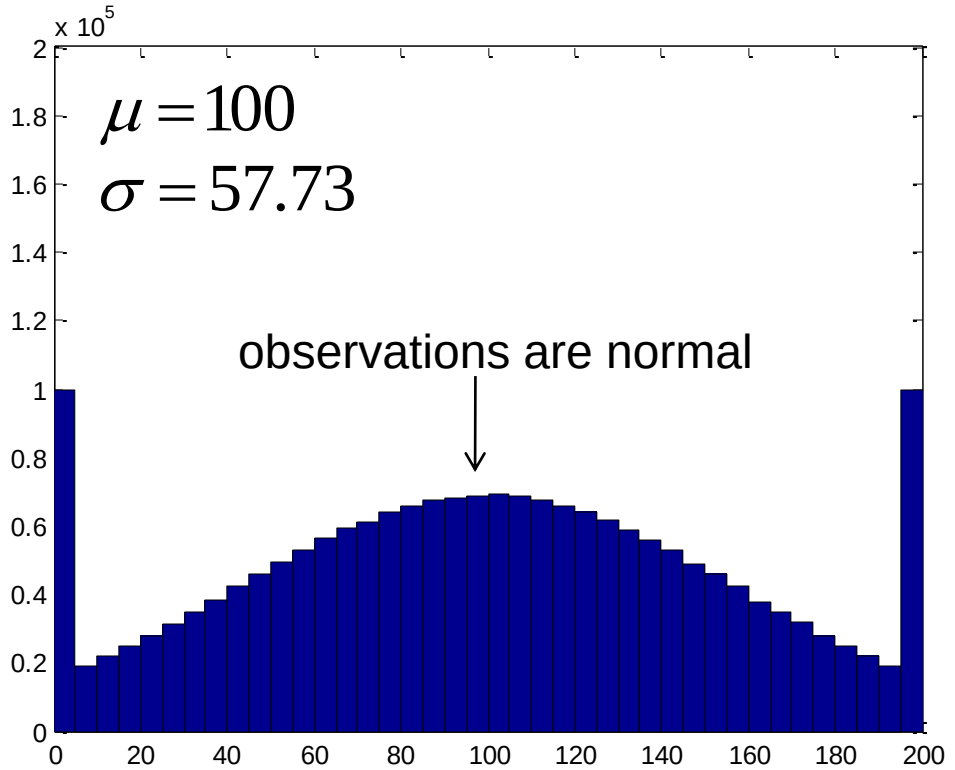




The Central Limit Theorem (CLT)

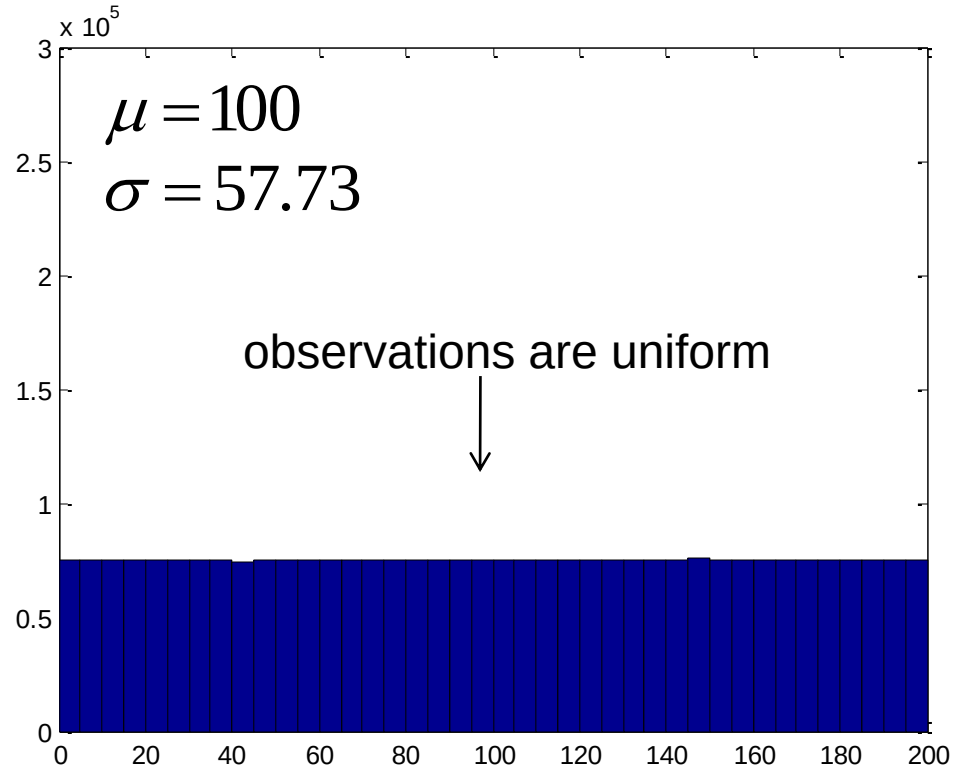
n=2 2×10^6 observations

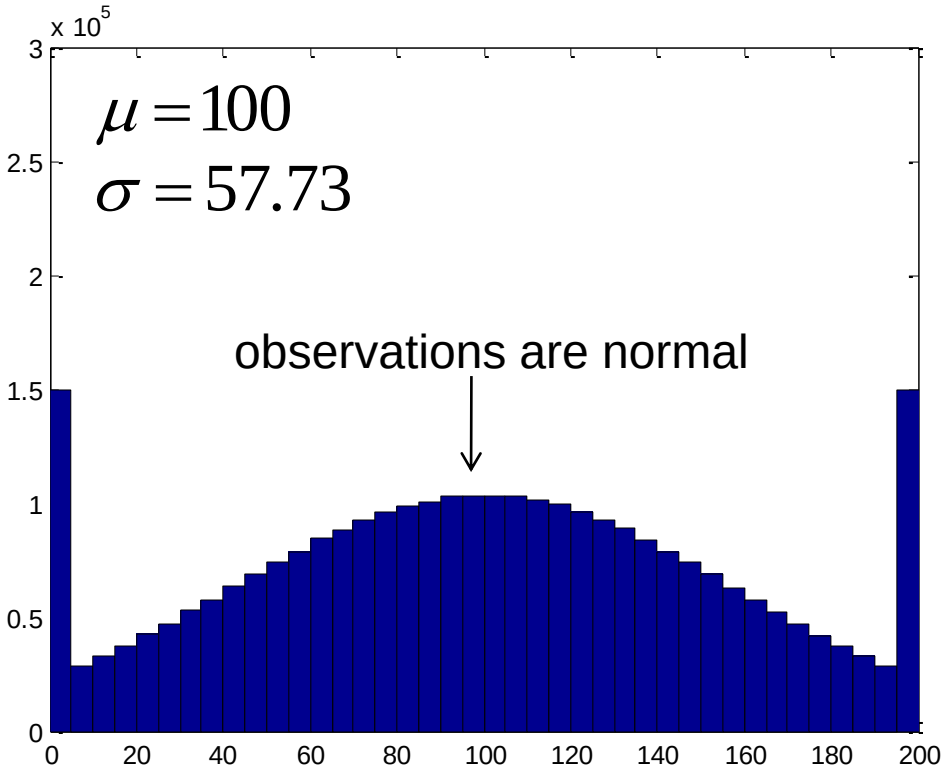




The Central Limit Theorem (CLT)

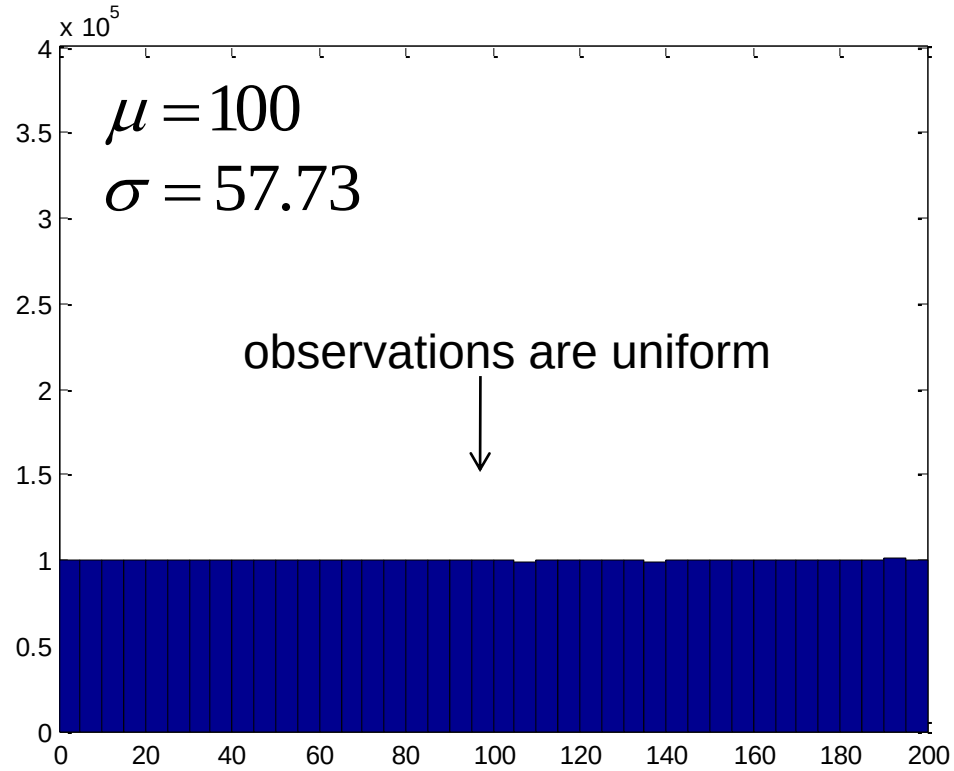
n=3 3×10^6 observations

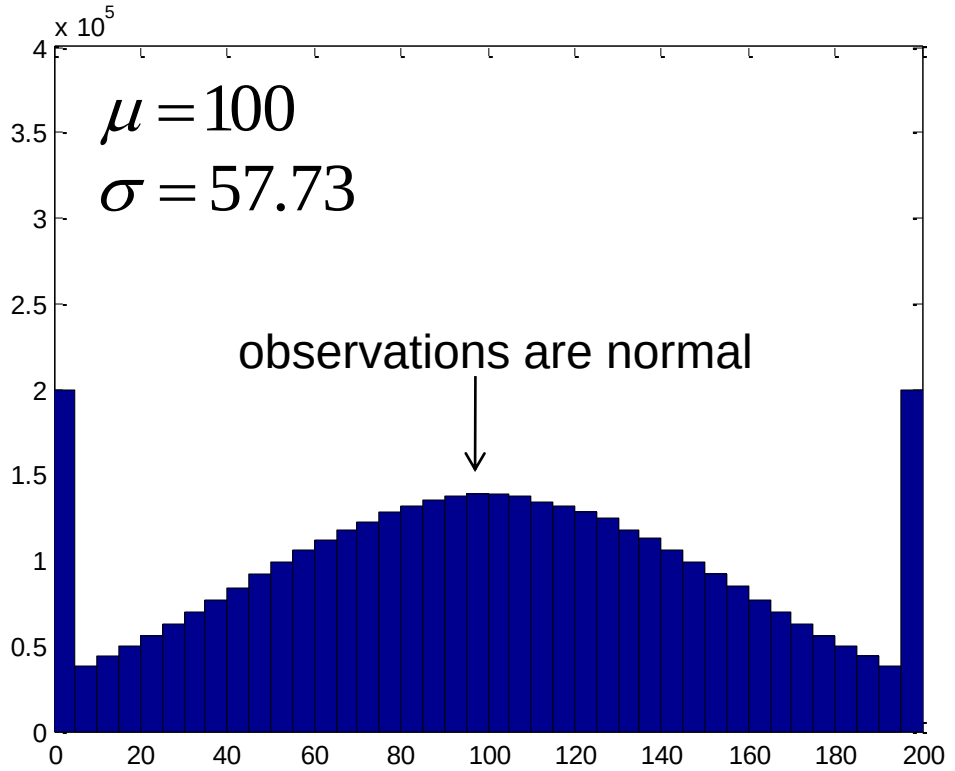




The Central Limit Theorem (CLT)

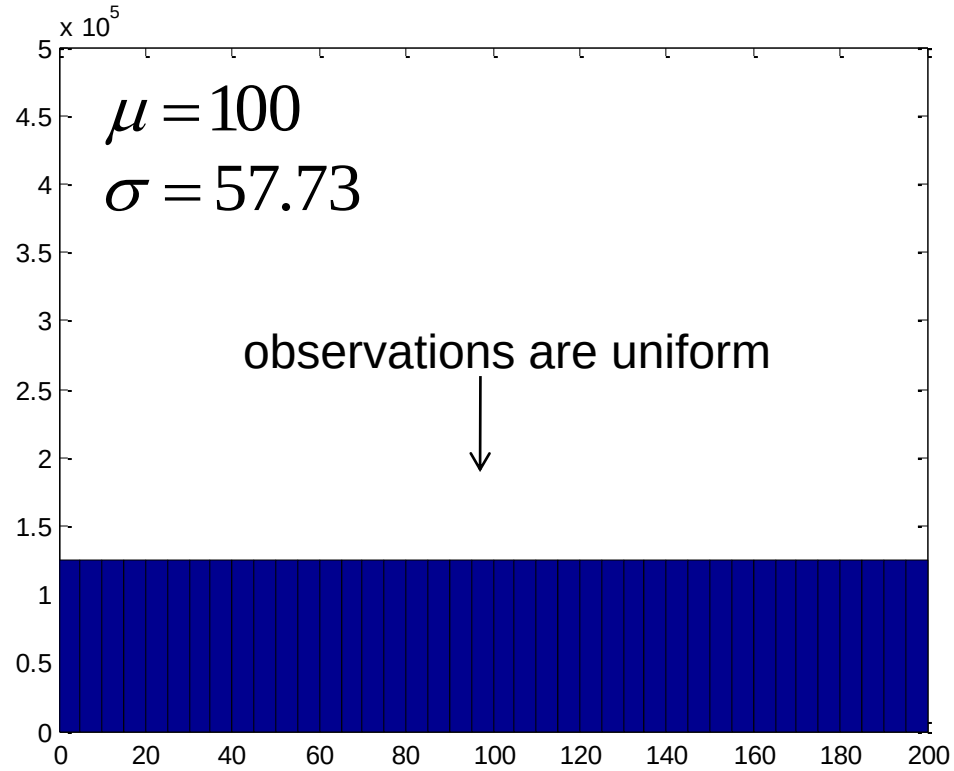
n=4 4×10^6 observations

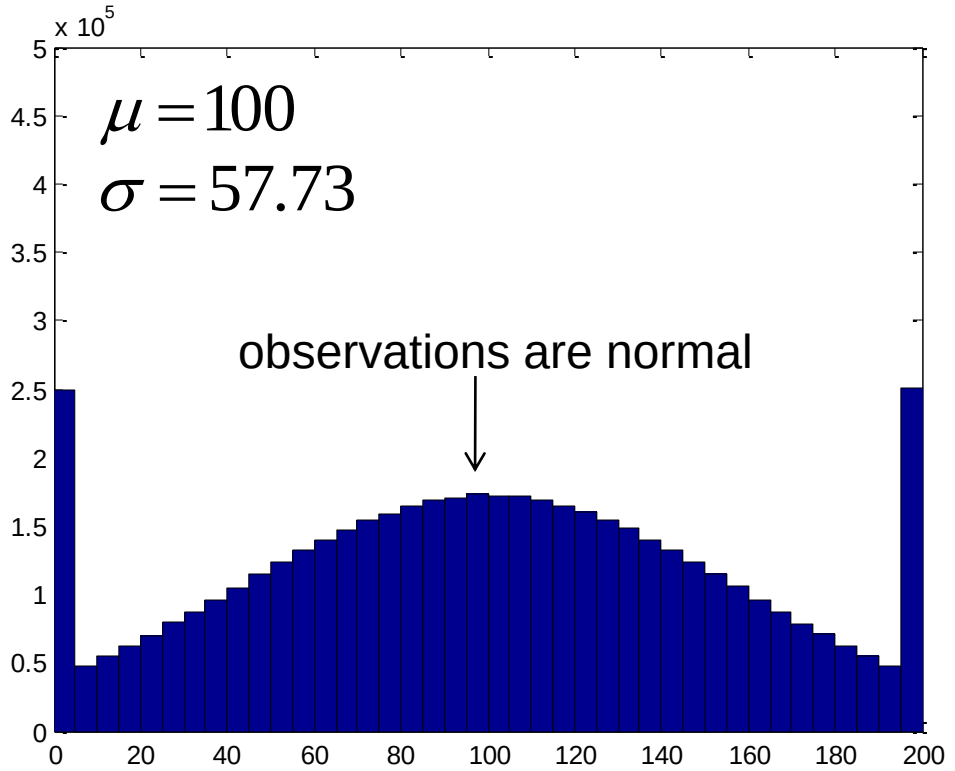




The Central Limit Theorem (CLT)

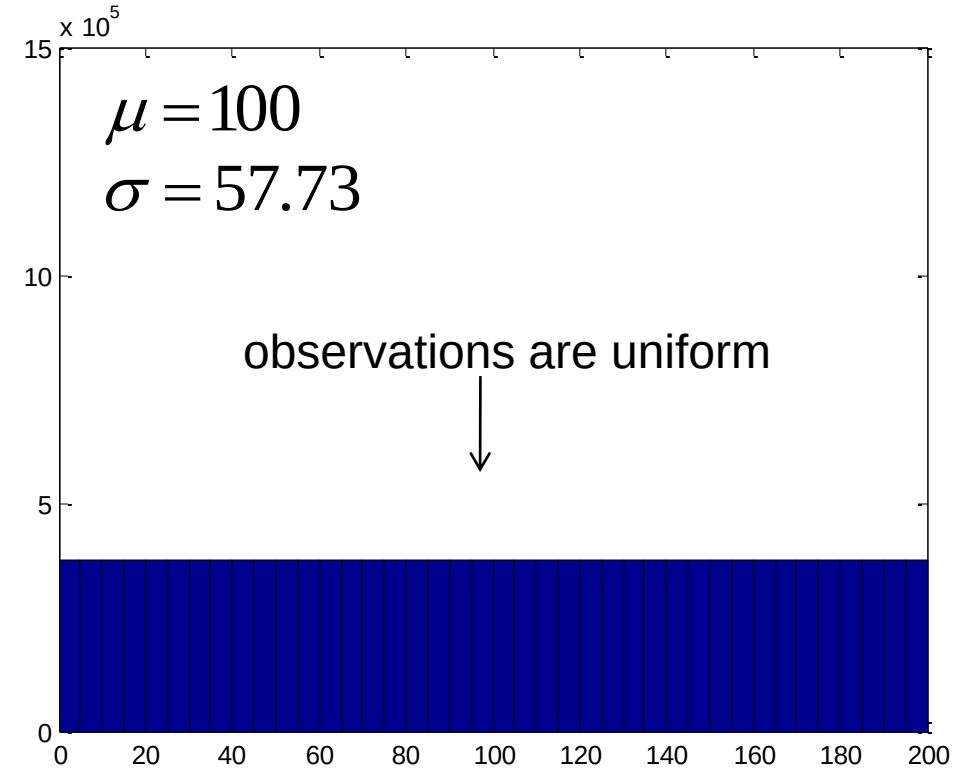
n=5 5×10^6 observations

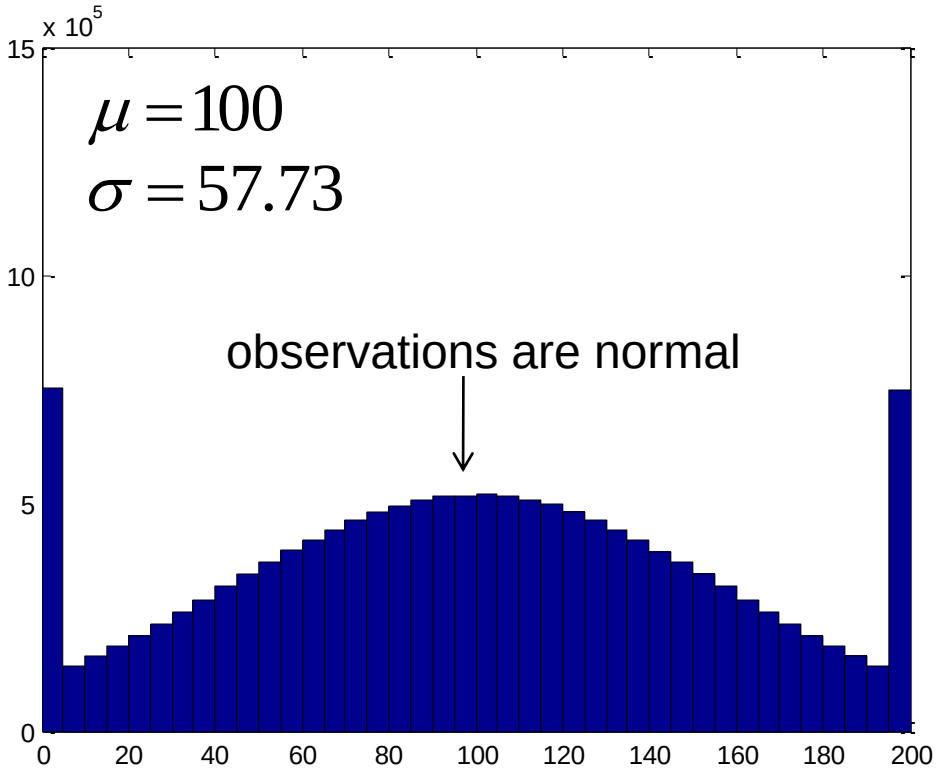




The Central Limit Theorem (CLT)

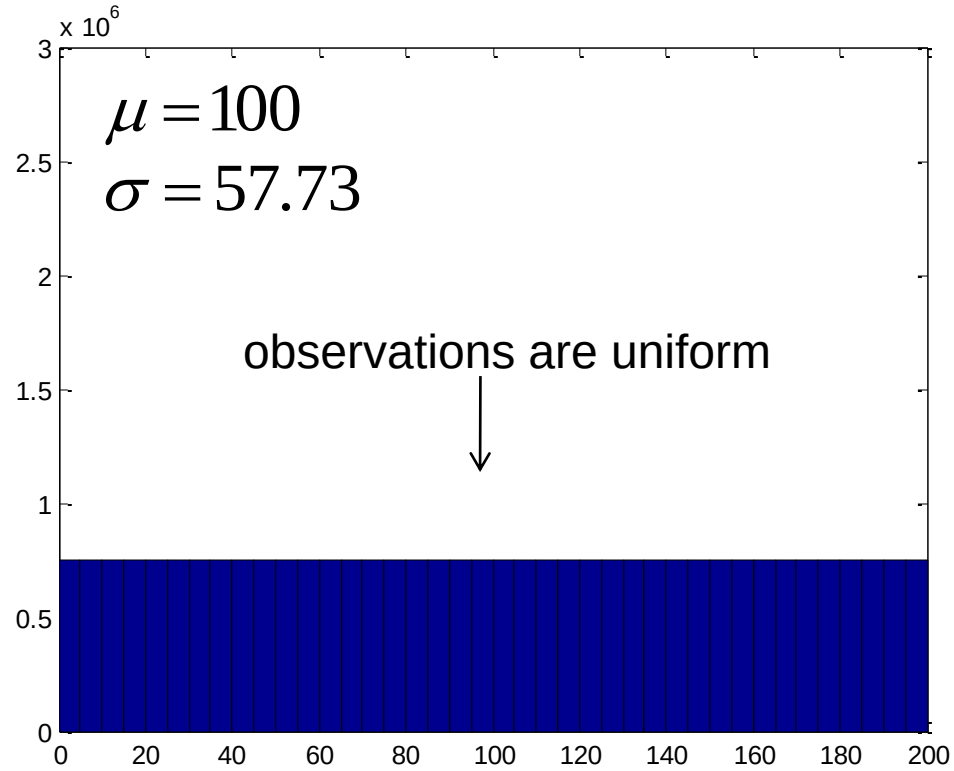
n=15 15×10^6 observations

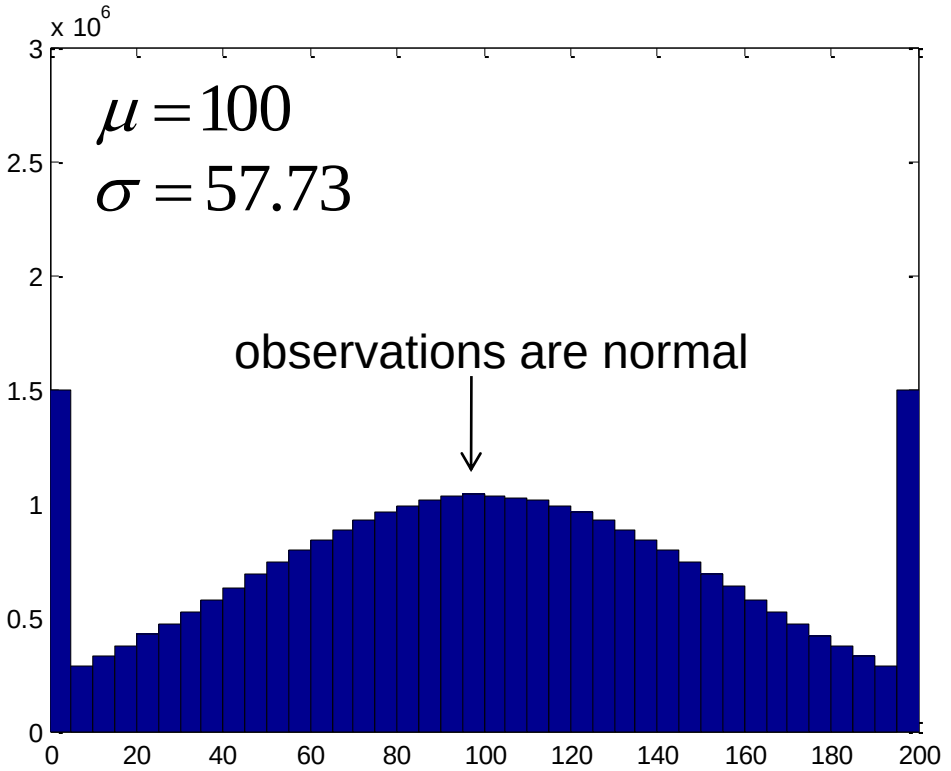




The Central Limit Theorem (CLT)

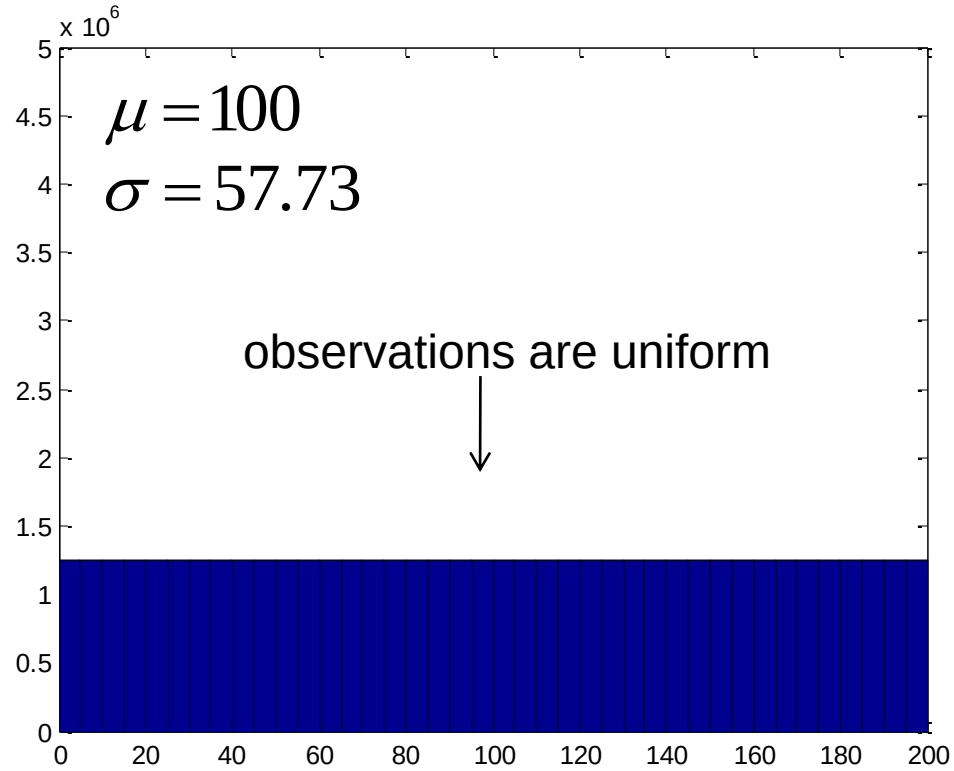
n=30 30×10^6 observations

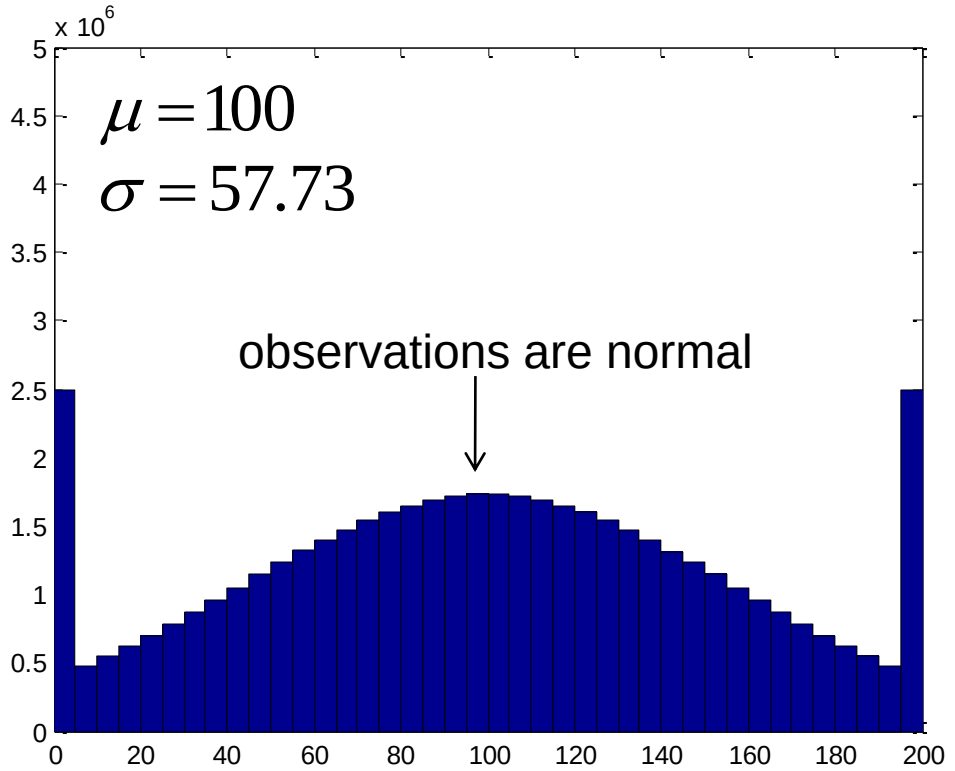




The Central Limit Theorem (CLT)

n=50 50×10^6 observations





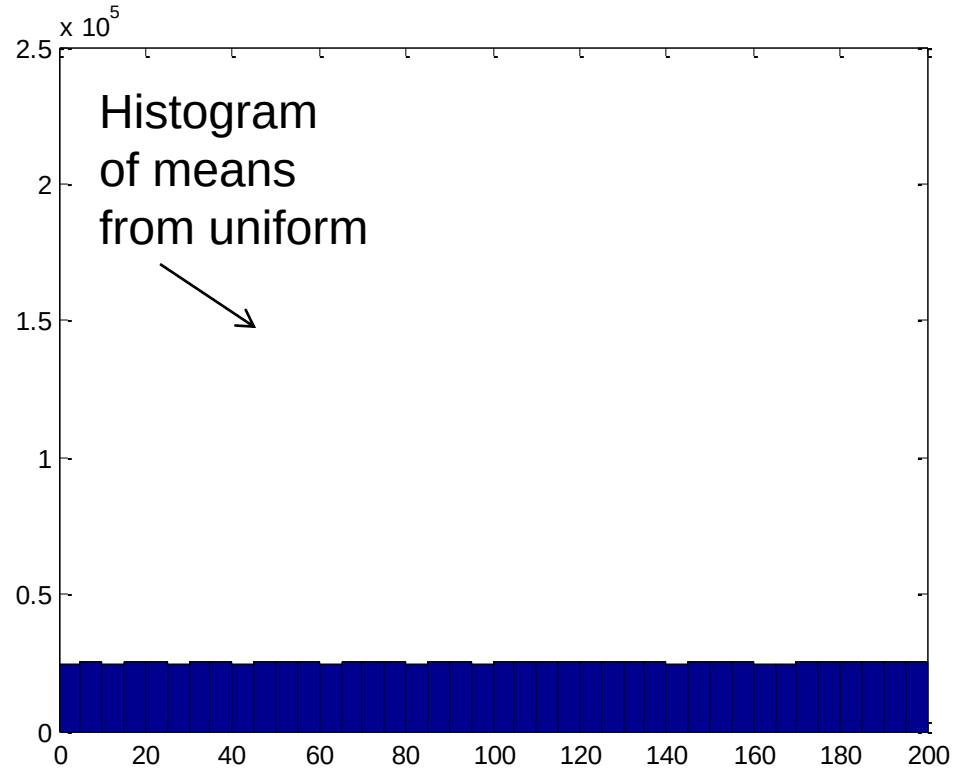
The Central Limit Theorem (CLT)

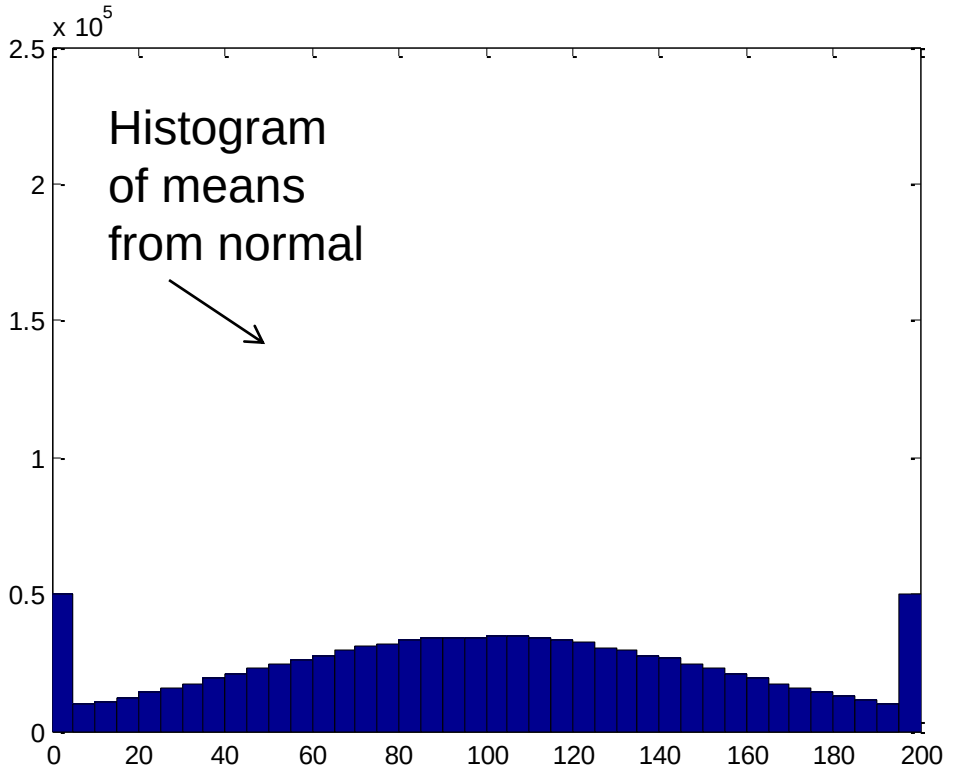
$\mu = 100$

$\sigma = 57.73$

n=1

1×10^6 means





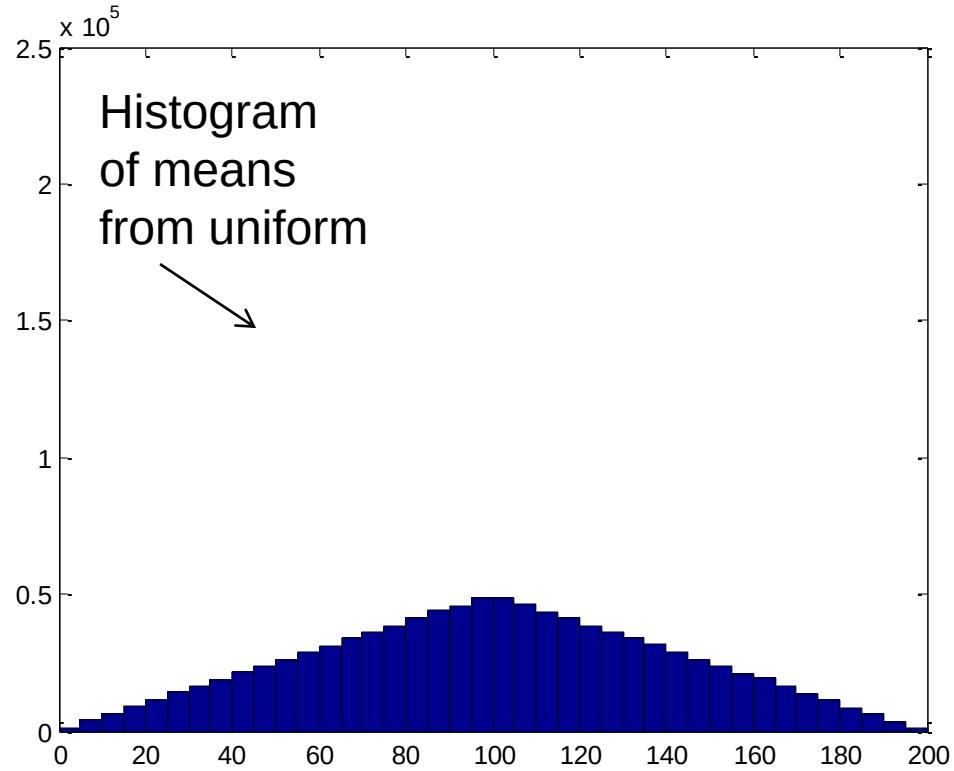
The Central Limit Theorem (CLT)

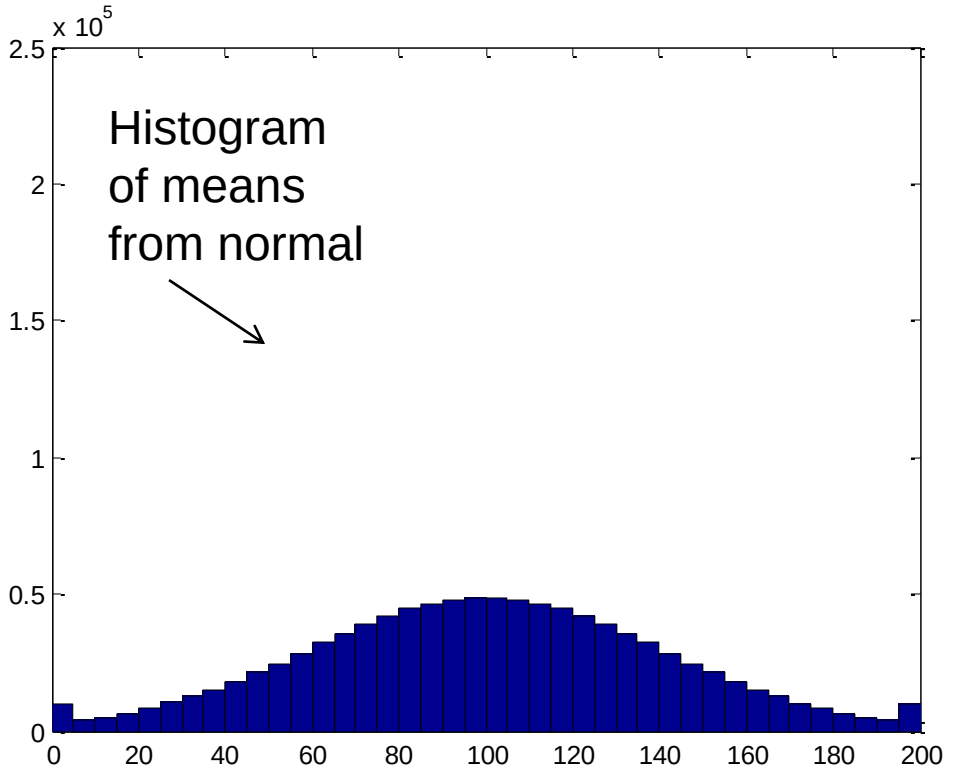
$\mu = 100$

$\sigma = 57.73$

n=2

1×10^6 means





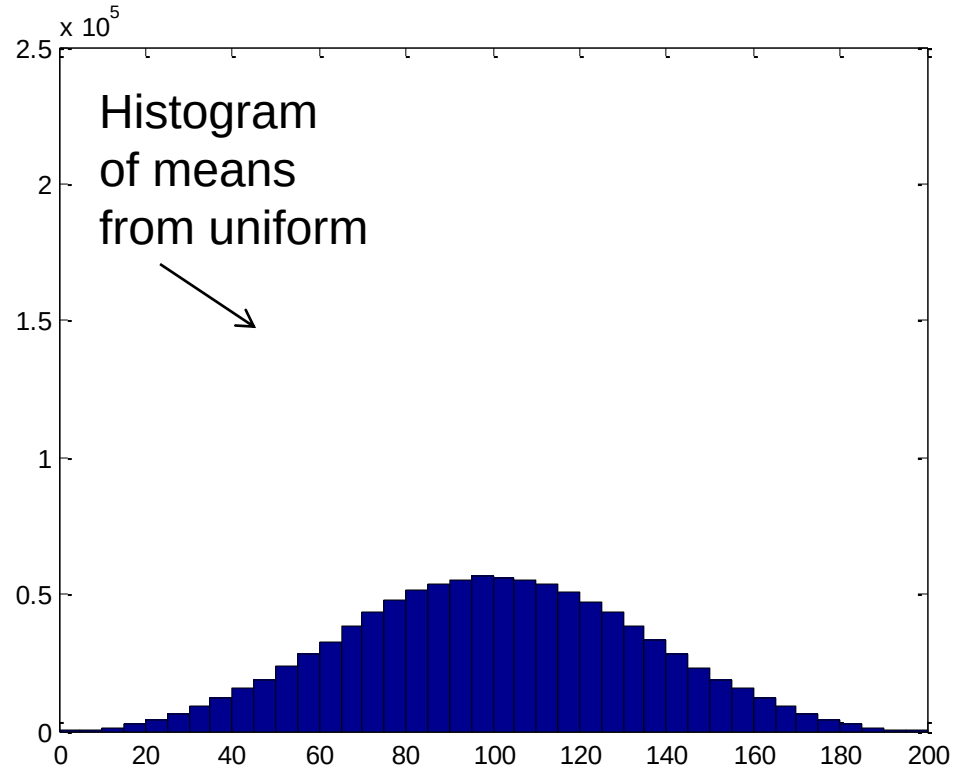
The Central Limit Theorem (CLT)

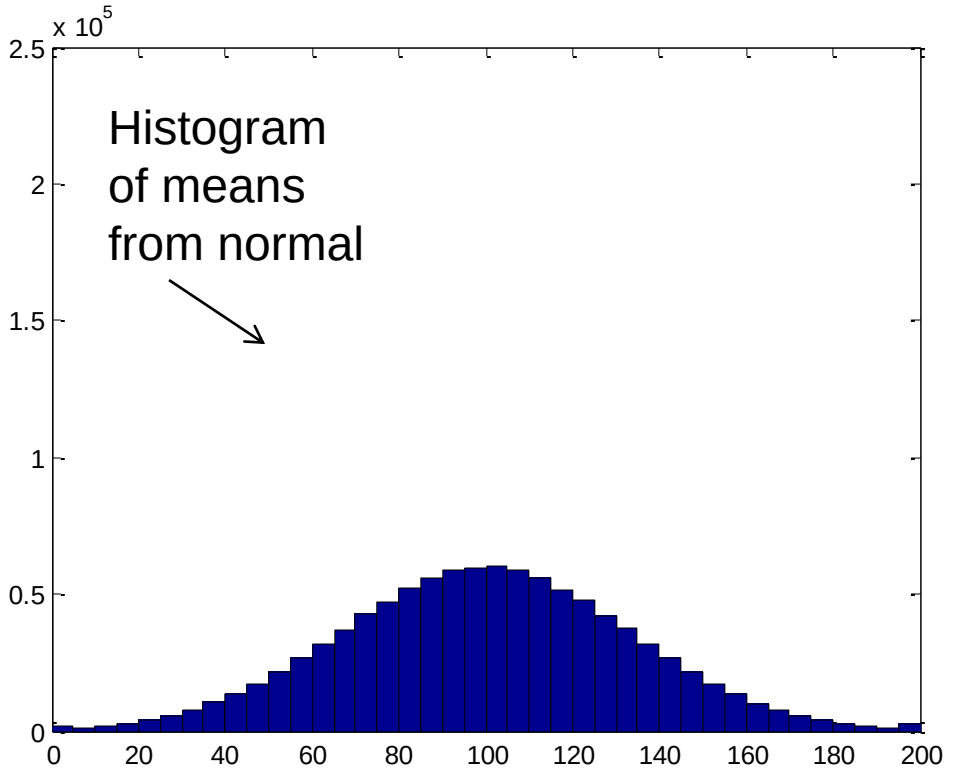
$\mu = 100$

$\sigma = 57.73$

n=3

1×10^6 means





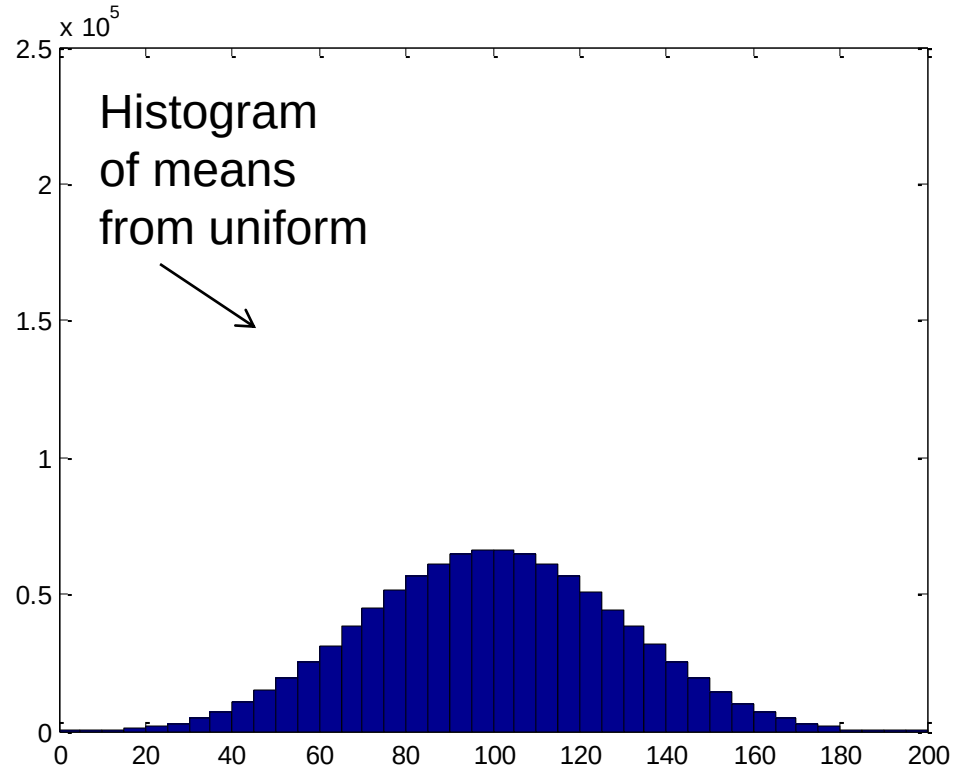
The Central Limit Theorem (CLT)

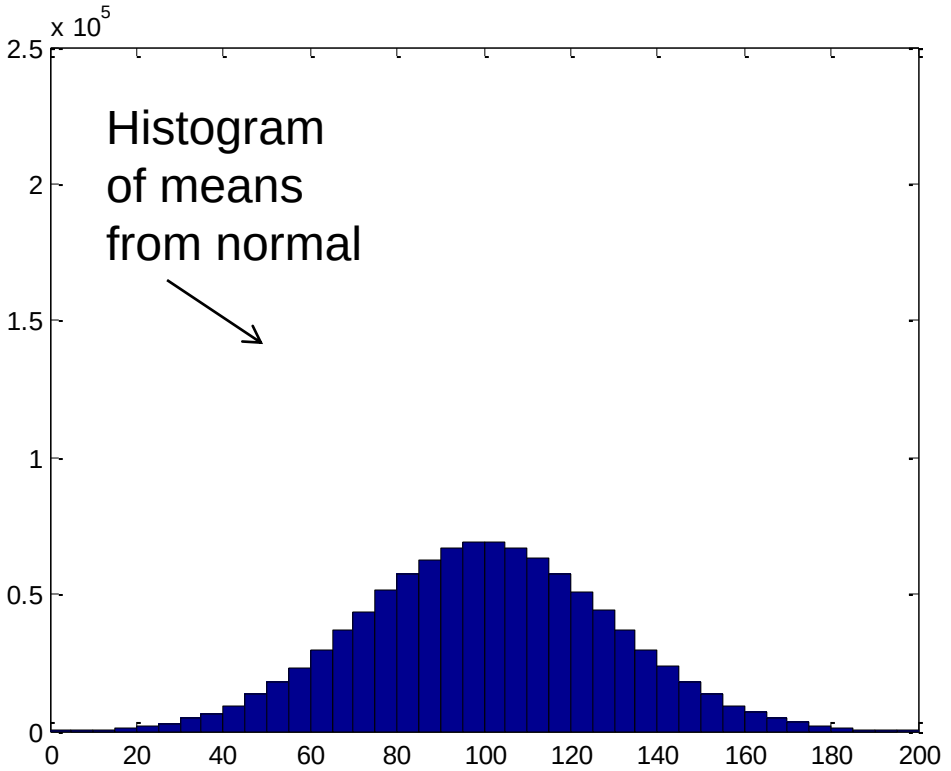
$\mu = 100$

$\sigma = 57.73$

n=4

1×10^6 means





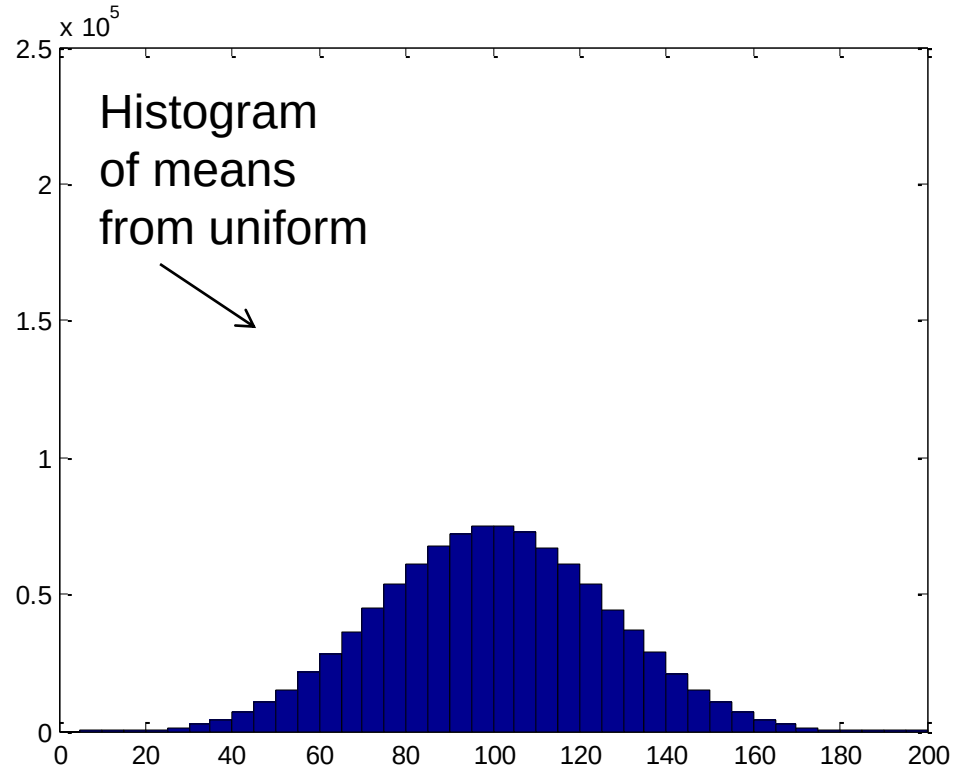
The Central Limit Theorem (CLT)

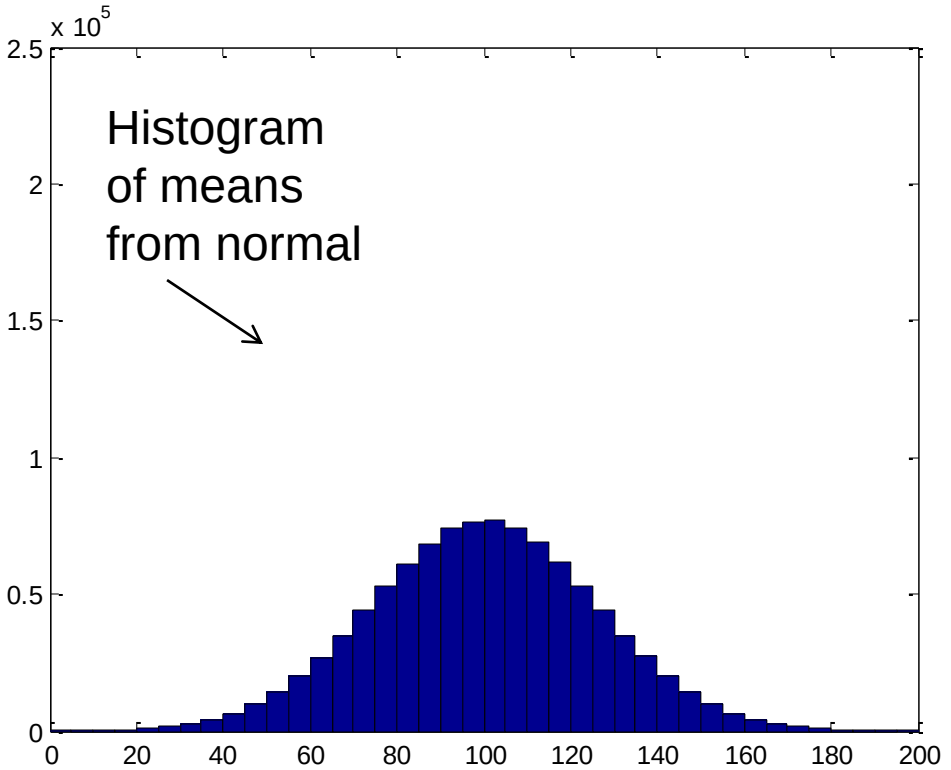
$\mu = 100$

$\sigma = 57.73$

n=5

1×10^6 means





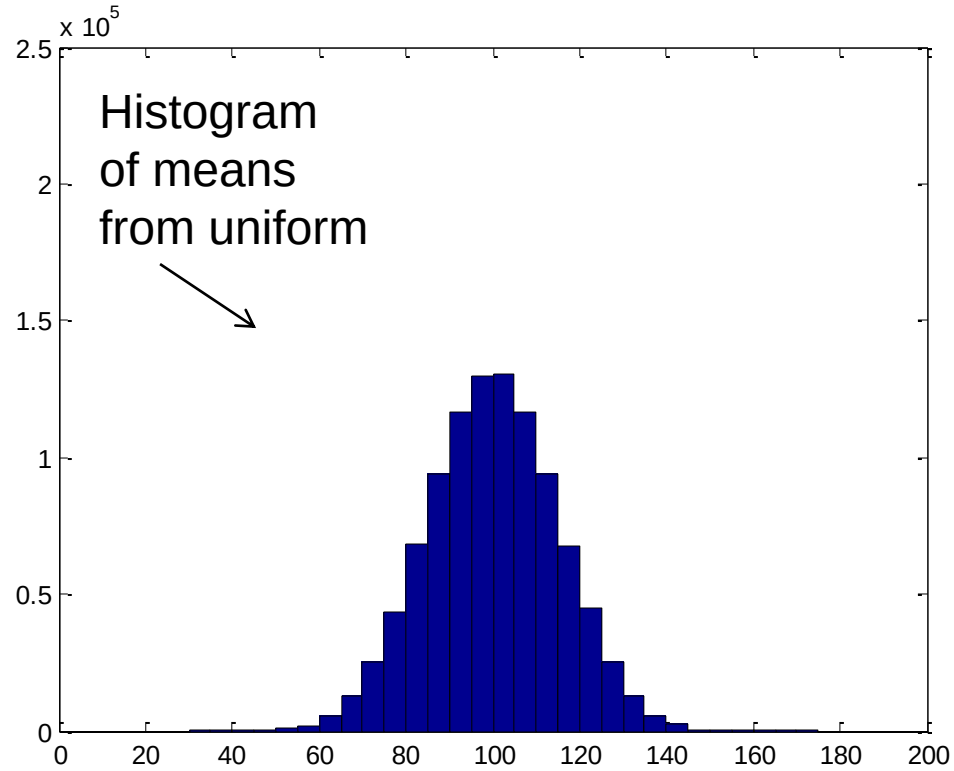
The Central Limit Theorem (CLT)

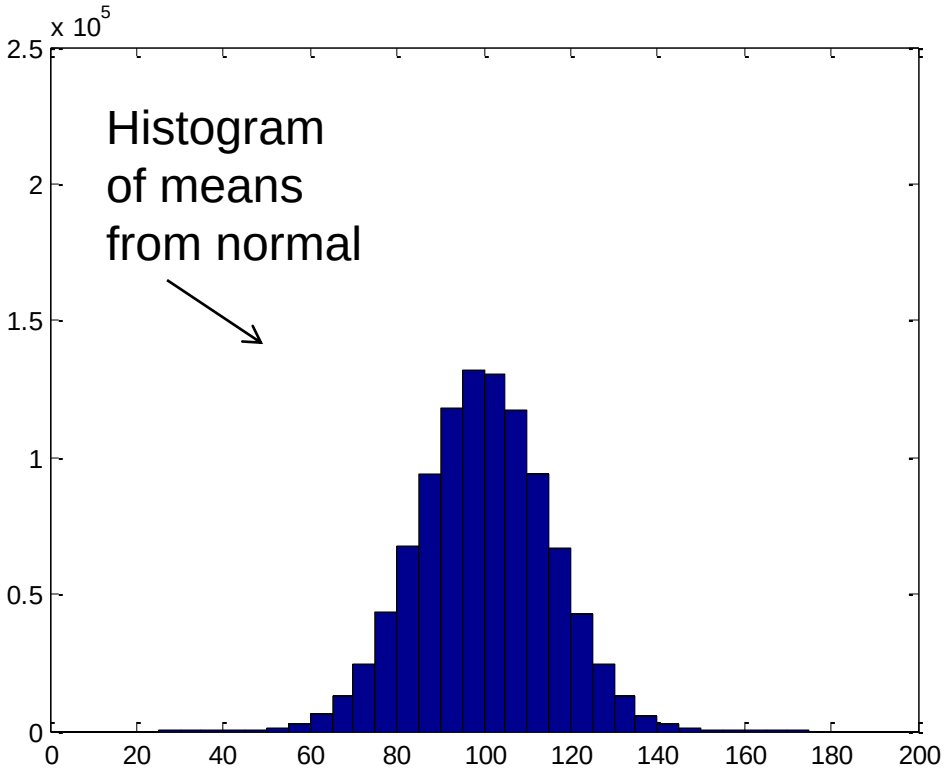
$\mu = 100$

$\sigma = 57.73$

n=15

1×10^6 means





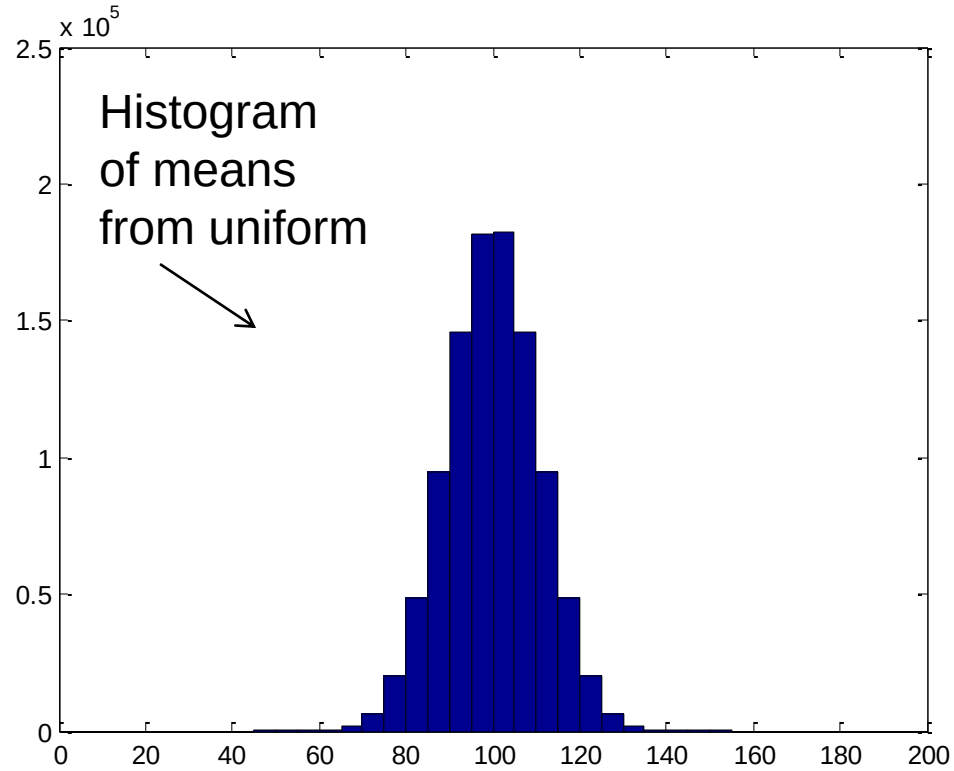
The Central Limit Theorem (CLT)

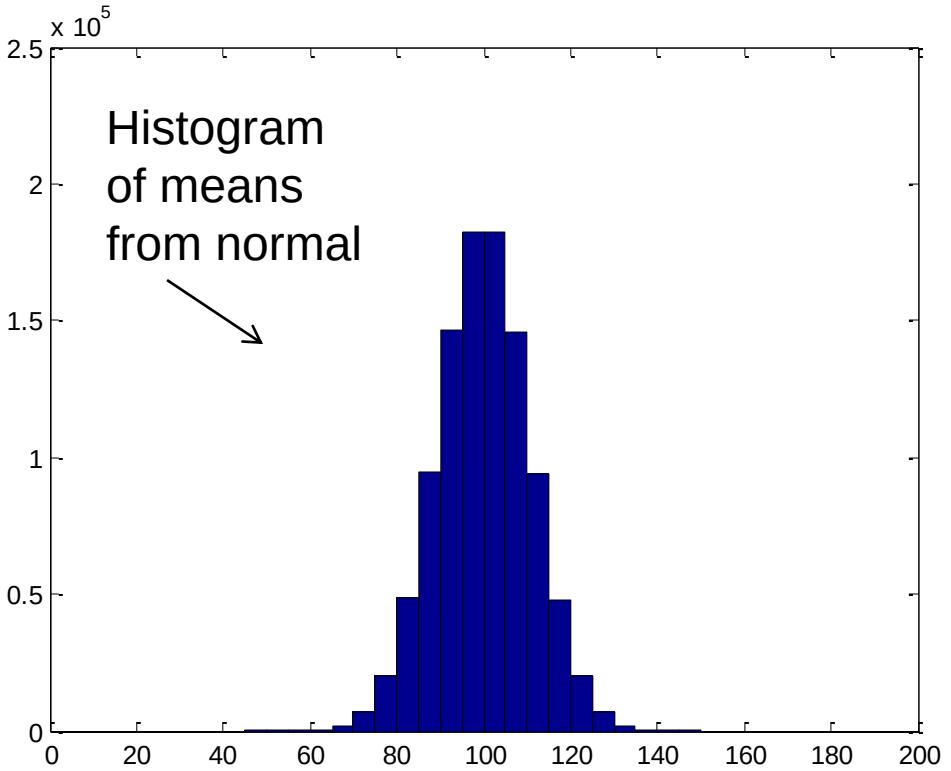
$\mu = 100$

$\sigma = 57.73$

n=30

1×10^6 means



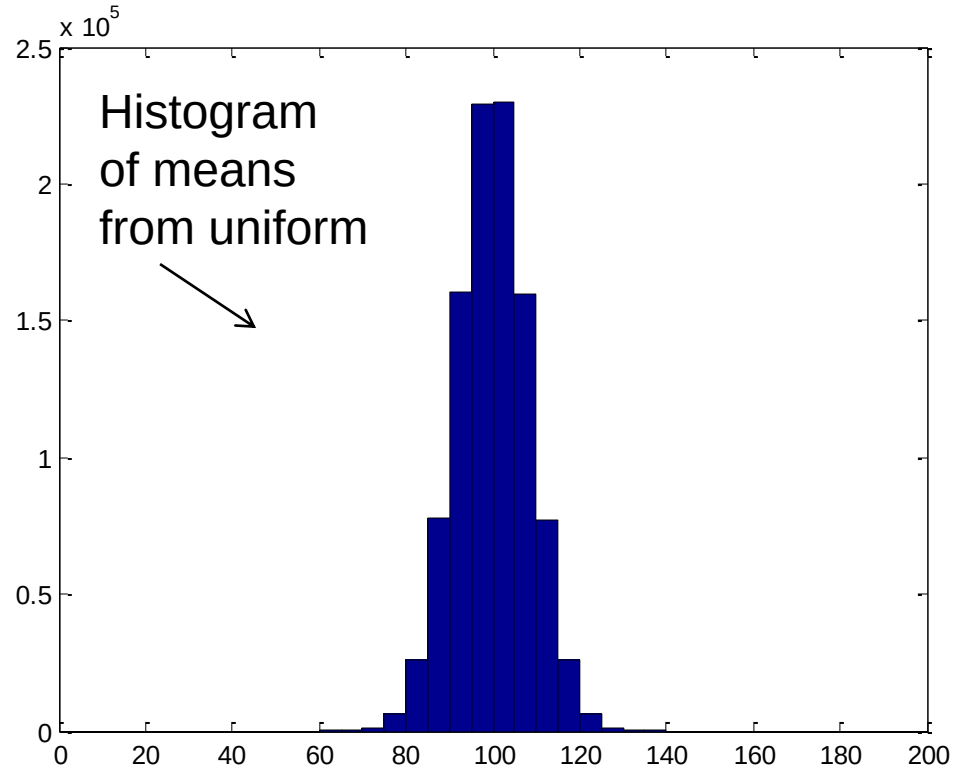


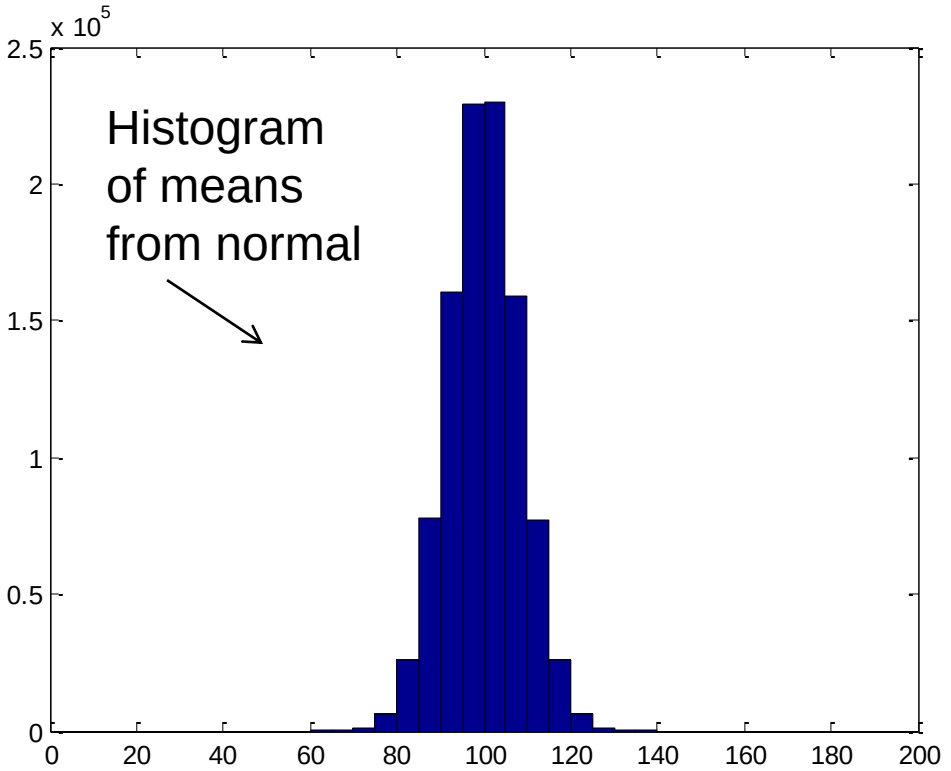
The Central Limit Theorem (CLT)

$\mu = 100$

$\sigma = 57.73$

n=50 1×10^6 means





The Central Limit Theorem (CLT)

With a population mean μ and standard deviation σ.

Random samples of size n, for large n, the distribution of the sample means quickly becomes normally distributed with

$$\mu_{\bar{x}} = \mu, \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Generally $n \geq 30$ is sufficiently large but much less often fine!

Homework 9:

1) Let u_1~uniform(0,1), u_2~uniform(0,1), u_1 and u_2 independent.
   a) Derive the distribution of $y_1 = u_1 + u_2$. Make a plot.
   b) Derive the distribution of $x = (u_1 + u_2)/2$. Make a plot.
   (This is the triangular distribution i.e. x~triangular(0,1).)

2) Let y_1~trangular(0,2), u_3~uniform(0,1), y_1, and u_3 independent.
   a) Derive the distribution of $y_3 = y_1 + u_3$. Make a plot.
   b) Derive the distribution of $y = (y_1 + u_3)/3$. Make a plot.
   c) Compare the distribution of u_1 to x to y.

Homework 9:

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

3) Let x_1, x_2, x_3 be independent standard Cauchy RVs.

a) Derive the distribution of $y_1 = x_1 + x_2$.

b) Derive the distribution of $y = (x_1 + x_2)/2$.

c) Derive the distribution of $y_3 = y_1 + x_3$.

d) Derive the distribution of $y = (y_1 + x_3)/3$.

e) Comment on what the distribution of the average of n independent Cauchy RVs is. Comment on the central limit theorem.

Homework 9:

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

4) Generate 10^6 standard Cauchy random variates.
   a) Make a histogram of the 10^6 Cauchy random variates.
   b) Compute the sample mean and variance.

5) Generate 2×10^6 standard Cauchy random variates.
   a) Make a histogram of the 2×10^6 Cauchy random variates.
   b) Compute the sample mean and variance of 2×10^6 variates.
   c) Compute the mean of every 2 to obtain 10^6 means.
   c) Make a histogram of the 10^6 means then of 10^6 variances.
   d) Compute the mean and variance of the 10^6 means.

Homework 9:

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

6) Repeat question 5) but change 2 to 3.

7) Comment on 4)-6). What did you learn?

Homework 9:

8) In the old days (before my time), I've heard that it was computationally expensive to generate normal(0,1) random numbers. So people used to generate 12 uniforms,

sum them, then subtract 6.

$$z = (u_1 + u_2 + u_3 + u_4 + u_5 + u_6 + u_7 + u_8 + u_9 + u_{10} + u_{11} + u_{12}) - 6$$

Then z was supposed to be normal (0,1).

Generate a 10^6 by 12 array, sum each row

and subtract 6 to make up a 13^{th} column.

Is this a good way to do it? Present a yes or no case.