Probability & Statistics

Authors : Bill Davis, Horacio Porta and Jerry Uhl **Producer** : Bruce Carpenter
Publisher : Math Everywhere, Inc. **Distributor & Translator**: MathMonkeys, LLC

7.01 Simulations

Tutorials T1

Experience with the starred problems will be useful for understanding developments later in the c

Graphics Primitives

The variables (x, s, t, z, y) are independent of each other.

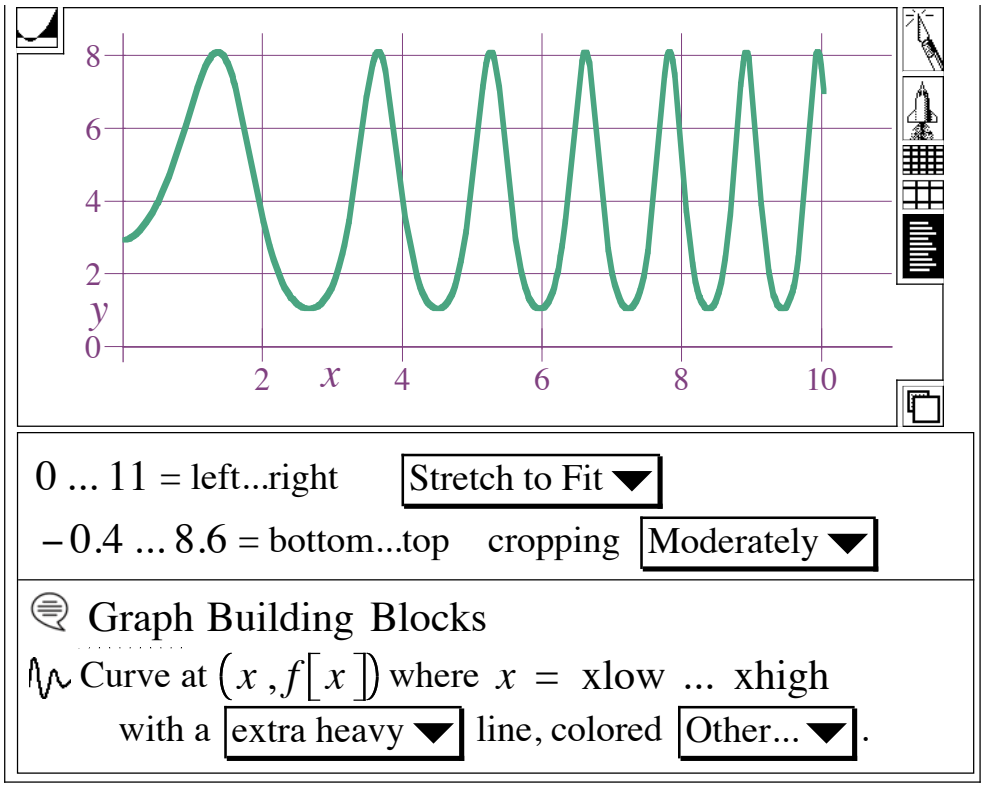
T.1) Monte Carlo estimation of integrals and other area measurements

John von Neumann and Stanislaw Ulam were great mathematicians of the early to midtwentiet
Working together on the original atomic bomb in the 1940's,
they devised the Monte Carlo method for doing approximate calculation for
problems intractable by hand.
You'll see many Monte Carlo simulations in this course.

T.1.a) Monte Carlo estimation of an integral $\int_{xlow}^{xhigh} f(x)\,dx$

Here's a function f[x] plotted from x = xlow to x = xhigh:

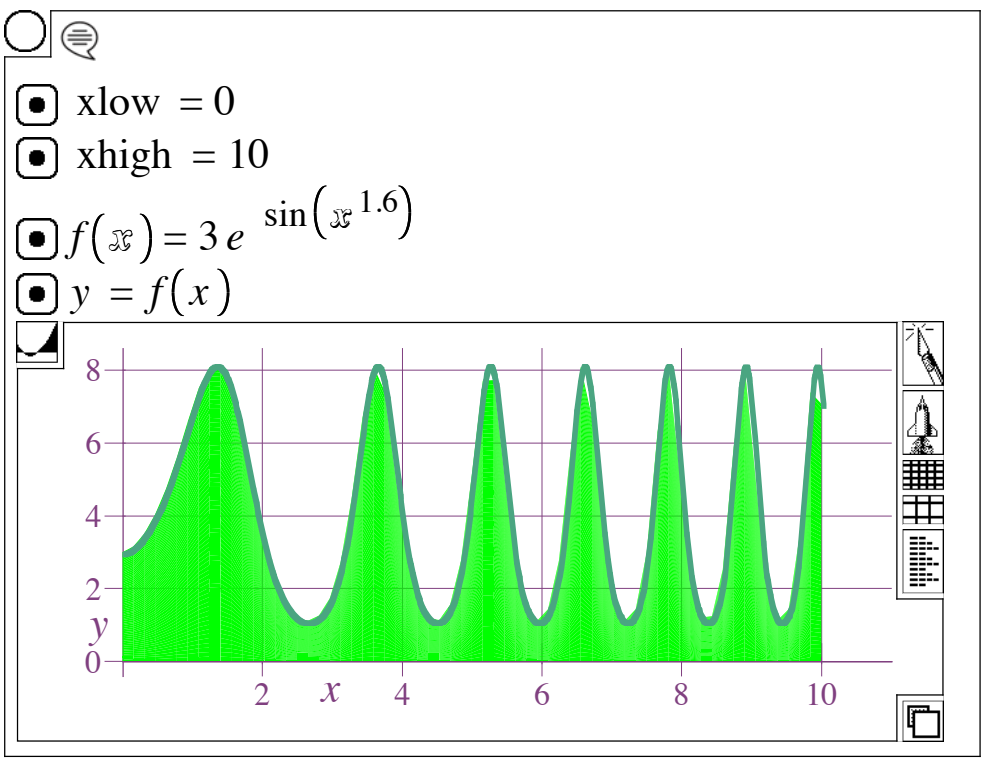
xlow = 0
xhigh = 10
$f(x) = 3\,e^{\sin(x^{1.6})}$
$y = f(x)$



The symbol:

$$\int_{xlow}^{xhigh} f(x)\,dx$$

stands for the measurement of the area between the plotted curve and the x-axis



Use uniformly distributed points and the Monte Carlo method to come up with a

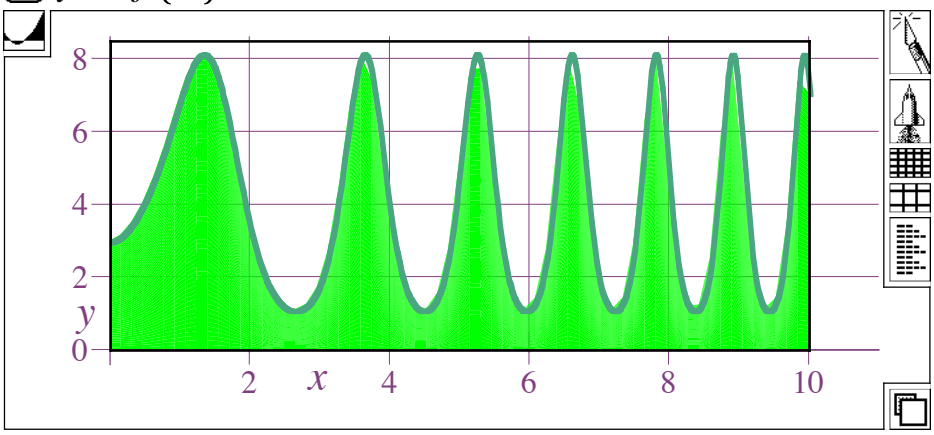
$$\int_{xlow}^{xhigh} f(x)\,dx,$$

the measurement of the area between the plotted curve and the x-axis.

Answer

Put a box around everything:

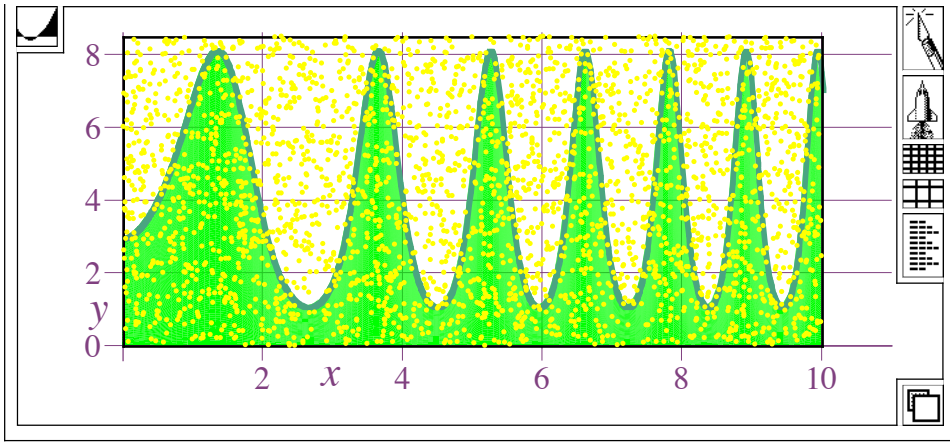
xlow = 0 xhigh = 10
ylow = 0 yhigh = 8.5
$f(x) = 3e^{\sin(x^{1.6})}$
$y = f(x)$



Your box does not have to touch everything in the area under measurement.
But, the area under measurement must be inside your box.

Generate 3000 uniformly distributed random points inside the box:

xlow = 0 xhigh = 10
ylow = 0 yhigh = 8.5
$f(x) = 3e^{\sin(x^{1.6})}$
x = Random(xlow, xhigh)
Tabulate xRanNums with
y = Random(ylow, yhigh)
Tabulate yRanNums with



Break the random points {x, y} into two separate groups:

1) pointsunder = those points {x, y} under or on the curve

and

2) pointsover = those points {x, y} over the curve

xlow = 0 xhigh = 10

ylow = 0 yhigh = 8.5

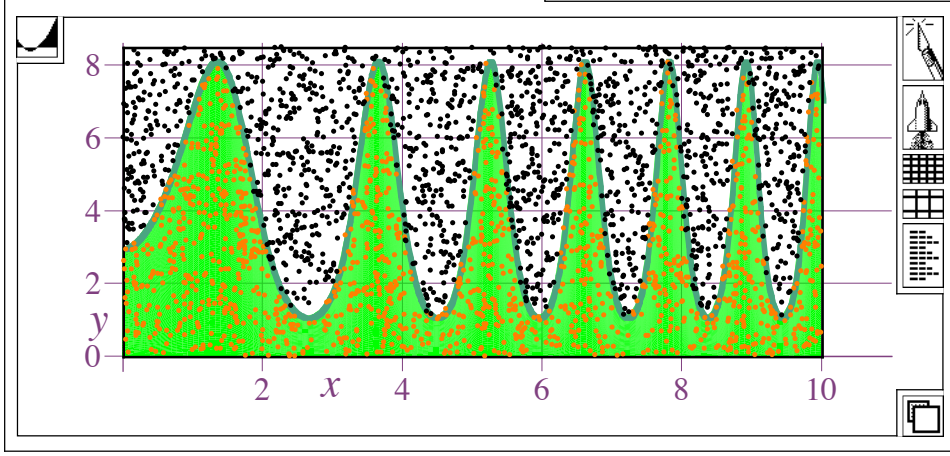
$f(x) = 3\,e^{\sin(x^{1.6})}$

x = Random (xlow , xhigh)

Tabulate xRanNums with

y = Random (ylow , yhigh)

Tabulate yRanNums with



pointsunder consists of the points {x, y} for which y ≤ f[x].

pointsover consists of the points {x, y} for which y > f[x].

The Monte carlo idea:

Because the random points are approximately uniformly distributed, you can e

$$\frac{\int_{xlow}^{xhigh} f(x)\,dx}{\text{Area enclosed by the box}} \approx \frac{\text{Number of random points under curve}}{\text{Total number of random points inside box}}.$$

so that:

$$\int_{xlow}^{xhigh} f(x)\,dx \approx \left(\frac{\text{Number of random points under curve}}{\text{Total number of random points inside box}}\right)(\text{Are}$$

Try it out:

xlow $= 0$ xhigh $= 10$

ylow $= 0$ yhigh $= 8.5$

$f(x) = 3\,e^{\sin(x^{1.6})}$

x = Random(xlow , xhigh)

Tabulate xRanNums with

y = Random(ylow , yhigh)

Tabulate yRanNums with

$$\text{UnderCount} = \sum_{k=1}^{3000} \left(\text{yRanNums}[k] \le f[\text{xRanNums}\{k\}]\right)$$

UnderCount $= 1418$ *Calculate*

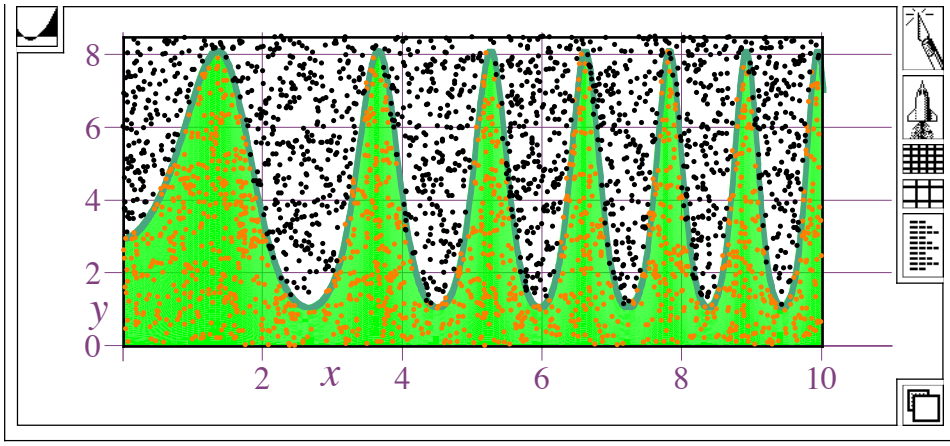
BoxArea $= (\text{xhigh} - \text{xlow})(\text{yhigh} - \text{ylow})$

BoxArea $= 85$ *Calculate*

$$\frac{\text{UnderCount}}{3000}\,\text{BoxArea}$$

$$\frac{\text{UnderCount}}{3000}\,\text{BoxArea} = \frac{12053}{300}$$ *Substitute*

$$\frac{\text{UnderCount}}{3000}\,\text{BoxArea} = 40.1766666666667$$ *Calculate*



Run 9 more Monte Carlo experiments, each involving 1000 independently ger

xlow = 0 xhigh = 10

ylow = 0 yhigh = 8.5

$f(x) = 3e^{\sin(x^{1.6})}$

x = Random(xlow , xhigh)

Tabulate xRanNums $_1$ with

Tabulate xRanNums $_2$ with

Tabulate xRanNums $_3$ with

Tabulate xRanNums $_4$ with

Tabulate xRanNums $_5$ with

Tabulate xRanNums $_6$ with

Tabulate xRanNums $_7$ with

Tabulate xRanNums $_8$ with

Tabulate xRanNums $_9$ with

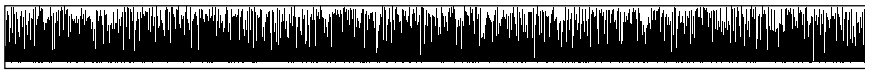
y = Random(ylow , yhigh)

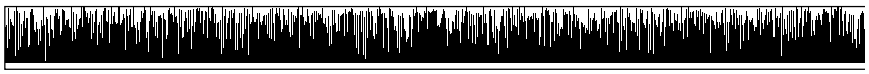
Tabulate yRanNums $_1$ with

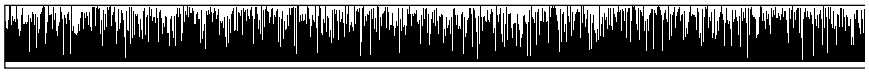
Tabulate yRanNums $_2$ with

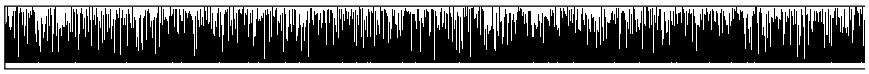
Tabulate yRanNums $_3$ with

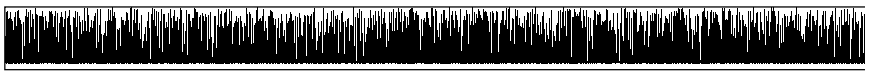
Tabulate yRanNums $_4$ with

Tabulate $yRanNums_5$ with 

Tabulate $yRanNums_6$ with 

Tabulate $yRanNums_7$ with 

Tabulate $yRanNums_8$ with 

Tabulate $yRanNums_9$ with 

$$UnderCount_1 = \sum_{k=1}^{3000} \left(yRanNums_1[k] \le f\left[xRanNums_1\{k\} \right] \right)$$

$UnderCount_1 = 1418$ *Calculate*

$$UnderCount_2 = \sum_{k=1}^{3000} \left(yRanNums_2[k] \le f\left[xRanNums_2\{k\} \right] \right)$$

$UnderCount_2 = 1376$ *Calculate*

$$UnderCount_3 = \sum_{k=1}^{3000} \left(yRanNums_3[k] \le f\left[xRanNums_3\{k\} \right] \right)$$

$UnderCount_3 = 1430$ *Calculate*

$$UnderCount_4 = \sum_{k=1}^{3000} \left(yRanNums_4[k] \le f\left[xRanNums_4\{k\} \right] \right)$$

$UnderCount_4 = 1417$ *Calculate*

$$UnderCount_5 = \sum_{k=1}^{3000} \left(yRanNums_5[k] \le f\left[xRanNums_5\{k\} \right] \right)$$

$UnderCount_5 = 1396$ *Calculate*

$$UnderCount_6 = \sum_{k=1}^{3000} \left(yRanNums_6[k] \le f\left[xRanNums_6\{k\} \right] \right)$$

$UnderCount_6 = 1454$ *Calculate*

$$UnderCount_7 = \sum_{k=1}^{3000} \left(yRanNums_7[k] \le f\left[xRanNums_7\{k\} \right] \right)$$

$UnderCount_7 = 1499$ *Calculate*

$$UnderCount_8 = \sum_{k=1}^{3000} \left(yRanNums_8[k] \le f\left[xRanNums_8\{k\} \right] \right)$$

$UnderCount_8 = 1442$ *Calculate*

$$\text{UnderCount}_9 = \sum_{k=1}^{3000} \left(\text{yRanNums}_9[k] \le f\left[\text{xRanNums}_9\{k\} \right] \right)$$

$\text{UnderCount}_9 = 1500$ *Calculate*

$$\text{BoxArea} = (\text{xhigh} - \text{xlow})(\text{yhigh} - \text{ylow})$$

$$\text{Estimate}_1 = \frac{\text{UnderCount}_1}{3000} \text{BoxArea}$$

$\text{Estimate}_1 = 40.1766666666667$ *Calculate*

$$\text{Estimate}_2 = \frac{\text{UnderCount}_2}{3000} \text{BoxArea}$$

$\text{Estimate}_2 = 38.9866666666667$ *Calculate*

$$\text{Estimate}_3 = \frac{\text{UnderCount}_3}{3000} \text{BoxArea}$$

$\text{Estimate}_3 = 40.5166666666667$ *Calculate*

$$\text{Estimate}_4 = \frac{\text{UnderCount}_4}{3000} \text{BoxArea}$$

$\text{Estimate}_4 = 40.1483333333333$ *Calculate*

$$\text{Estimate}_5 = \frac{\text{UnderCount}_5}{3000} \text{BoxArea}$$

$\text{Estimate}_5 = 39.5533333333333$ *Calculate*

$$\text{Estimate}_6 = \frac{\text{UnderCount}_6}{3000} \text{BoxArea}$$

$\text{Estimate}_6 = 41.1966666666667$ *Calculate*

$$\text{Estimate}_7 = \frac{\text{UnderCount}_7}{3000} \text{BoxArea}$$

$\text{Estimate}_7 = 42.4716666666667$ *Calculate*

$$\text{Estimate}_8 = \frac{\text{UnderCount}_8}{3000} \text{BoxArea}$$

$\text{Estimate}_8 = 40.8566666666667$ *Calculate*

$$\text{Estimate}_9 = \frac{\text{UnderCount}_9}{3000} \text{BoxArea}$$

$\text{Estimate}_9 = 42.5$ *Calculate*

Look at the average of the 10 Monte Carlo estimates:

$$\text{Average} = \frac{\sum_{k=1}^{9} \text{Estimate}_k}{9}$$

$$\text{Average} = \frac{1}{9}\text{Estimate}_9 + \frac{1}{9}\text{Estimate}_8 + \frac{1}{9}\text{Estimate}_7 + \frac{1}{9}\text{Estimate}_6 + \frac{1}{9}\text{Estima}$$

$$\text{Average} = 40.7118518518519$$ *Calculate*

Compare with the actual value of:

$$\int_{xlow}^{xhigh} f(x)\,dx$$

as calculated by LiveMath:

xlow $= 0$ xhigh $= 10$

ylow $= 0$ yhigh $= 8.5$

$f(x) = 3\,e^{\sin(x^{1.6})}$

$$\int_{xlow}^{xhigh} f(x)\,dx$$

$$\int_{xlow}^{xhigh} f(x)\,dx = 40.5342295803873$$ *Calculate*

The Law of Large numbers says:

If your experience was typical, then the average of these Monte Carlo estimates is a close enough approximation of:

$$\int_{xlow}^{xhigh} f(x)\,dx$$

for most government work.

T.1.b) Monte Carlo estimation of another area measurement

Here's the collapsed circle:

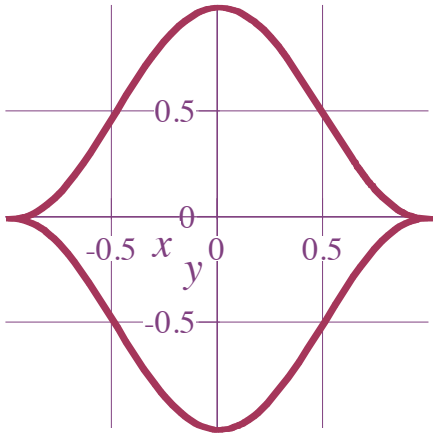
$x^2 + y^{2/5} = 1$:

xlow $= -1$

xhigh $= 1$

$xx(t) = \cos(t)$

$yy(t) = (\sin[t])^5$



Use uniformly distributed points and the Monte Carlo method to estimate the measurement of the area enclosed by the curve.

Answer

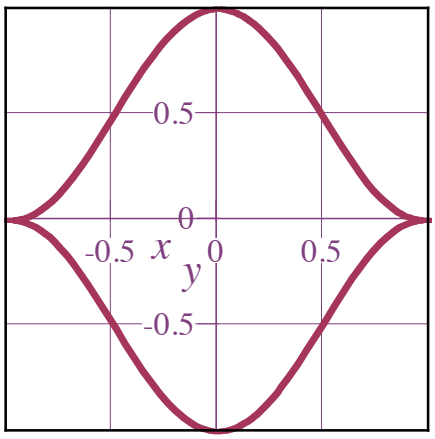
Put a box around the curve:

xlow = −1 xhigh = 1

ylow = −1 yhigh = 1

$xx(t) = \cos(t)$

$yy(t) = (\sin[t])^5$



−1 ... 1 = left...right True Proportions

−1 ... 1 = bottom...top cropping Moderately

Graph Building Blocks

Curve at (xx[t], yy[t]) where $t = 0 \ldots 2\pi$ with a extra heavy line, colored Other... .

Box

Generate 2000 uniformly distributed random points inside the box:

xlow = −1 xhigh = 1

ylow = −1 yhigh = 1

xx(t) = cos(t)

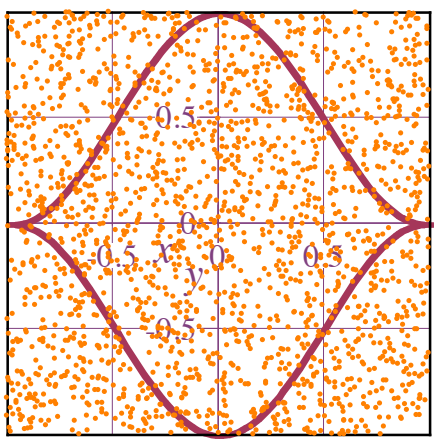
yy(t) = (sin[t])^5

x = Random(xlow , xhigh)

Tabulate xRanNums with

y = Random(ylow , yhigh)

Tabulate yRanNums with



The Monte carlo idea:

Because the points are approximately uniformly distributed, you can expect:

$$\frac{\text{Area enclosed by curve}}{\text{Area enclosed by the box}} \approx \frac{\text{Number of random points inside curve}}{\text{Total number of random points inside box}}$$

so that:

Area enclosed by curve ≈

$$\left(\frac{\text{Number of random points inside curve}}{\text{Total number of random points inside box}}\right).$$

Try it out remembering that:

$$x^2+y^{2/5}=1$$

is the same curve as:

$$|x|^2+|y|^{2/5}=1:$$

xlow $= -1$ xhigh $= 1$

ylow $= -1$ yhigh $= 1$

$xx(t) = \cos(t)$

$yy(t) = (\sin[t])^5$

$x = \text{Random}(\text{xlow}, \text{xhigh})$

Tabulate xRanNums with

$y = \text{Random}(\text{ylow}, \text{yhigh})$

Tabulate yRanNums with

$\text{BoxArea} = (\text{xhigh} - \text{xlow})(\text{yhigh} - \text{ylow})$

$\text{BoxArea} = 4$ *Calculate*

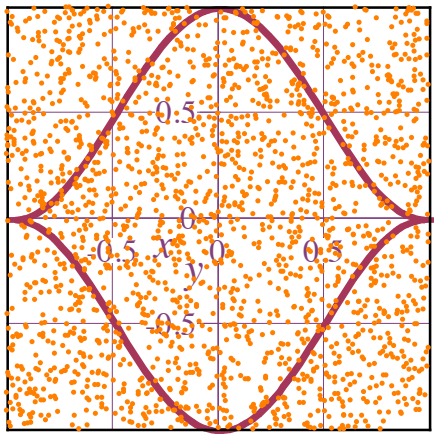
$$\text{Count} = \sum_{k=1}^{2000}\left(|\text{xRanNums}[k]|^2+|\text{yRanNums}[k]|^{\frac{2}{5}} \le 1\right)$$

$\text{Count} = 976$ *Calculate*

$$\text{AreaEstimate} = \frac{\text{Count}}{2000}\,\text{BoxArea}$$

$\text{AreaEstimate} = \frac{244}{125}$ *Substitute*

$\text{AreaEstimate} = 1.952$ *Calculate*



Now, average to try to get a better estimate:

$\text{xlow} = -1$ $\text{xhigh} = 1$

$\text{ylow} = -1$ $\text{yhigh} = 1$

$\text{xx}(t) = \cos(t)$

$\text{yy}(t) = (\sin[t])^5$

LiveMath Note: Making 20 Tabulates can get pretty tiring. So here in this example we do the same computations, but with some functional magic. Everytime these functions get Calculated for each input k, the output is a random number. Notice no k on the right-hand-side, because the random number is not dependent upon the k value.

$\text{xRandoms}(k) = \text{Random}(\text{xlow}, \text{xhigh})$

$\text{yRandoms}(k) = \text{Random}(\text{ylow}, \text{yhigh})$

$$\text{fCounts}(n) = \sum_{k=1}^{2000} \left(|\text{xRandoms}[k]|^2 + |\text{yRandoms}[k]|^{\frac{2}{5}} \le 1 \right)$$

$\text{BoxArea} = (\text{xhigh} - \text{xlow})(\text{yhigh} - \text{ylow})$

$$\text{AreaEst}(n) = \frac{\text{fCounts}(n)}{2000} \text{BoxArea}$$

$\text{AreaEst}(1)$

$\text{AreaEst}(1) = 2.022$ *Calculate*

$\text{AreaEst}(2)$

$\text{AreaEst}(2) = 1.886$ *Calculate*

$\text{AreaEst}(3)$

△ AreaEst(3) = 1.922 *Calculate*

▢ AreaEst(4)

△ AreaEst(4) = 1.954 *Calculate*

▢ AreaEst(5)

△ AreaEst(5) = 1.948 *Calculate*

▢ AreaEst(6)

△ AreaEst(6) = 1.96 *Calculate*

▢ AreaEst(7)

△ AreaEst(7) = 1.978 *Calculate*

▢ AreaEst(8)

△ AreaEst(8) = 1.996 *Calculate*

▢ AreaEst(9)

△ AreaEst(9) = 2.018 *Calculate*

▢ AreaEst(10)

△ AreaEst(10) = 1.988 *Calculate*

The average of the 10 Monte Carlo estimates is:

▢ $\text{Average} = \dfrac{\sum_{k=1}^{10} \text{AreaEst}(k)}{10}$

△ Average = 1.9718 *Calculate*

If you've had a strong calculus course such as Calculus&LiveMath: VectorCal you are in a position to check this estimate.

You parameterize the plotted curve $x^2+y^{2/5} = 1$ via the parameterization:

$$\{x(t),y(t)\} = \{\cos(t),\sin(t)^5\}$$

with $0 \le t \le 2\pi$ and then calculate:

$$\int_0^{2\pi} x(t)\, y'(t)\,dt:$$

○

▣ xlow = −1 ▣ xhigh = 1

▣ ylow = −1 ▣ yhigh = 1

▣ xx(t) = cos(t)

$yy(t) = (\sin[t])^5$

$xx'(t) = \left[\frac{d}{dt} xx(t)\right]_{t=t}$

$xx'(t) = -\sin(t)$ *Substitute*

$yy'(t) = \left[\frac{d}{dt} yy(t)\right]_{t=t}$

$yy'(t) = 5\cos(t)(\sin[t])^4$ *Substitute*

$\int_0^{2\pi} xx(t)\, yy'(t)\, dt$

$\int_0^{2\pi} xx(t)\, yy'(t)\, dt = 1.96349540648435$ *Calculate*

Because of the Law of Large Numbers, the smart money bets that the average

T.1.c) Monte Carlo estimation of another area measurement

Here's the collapsed circle:

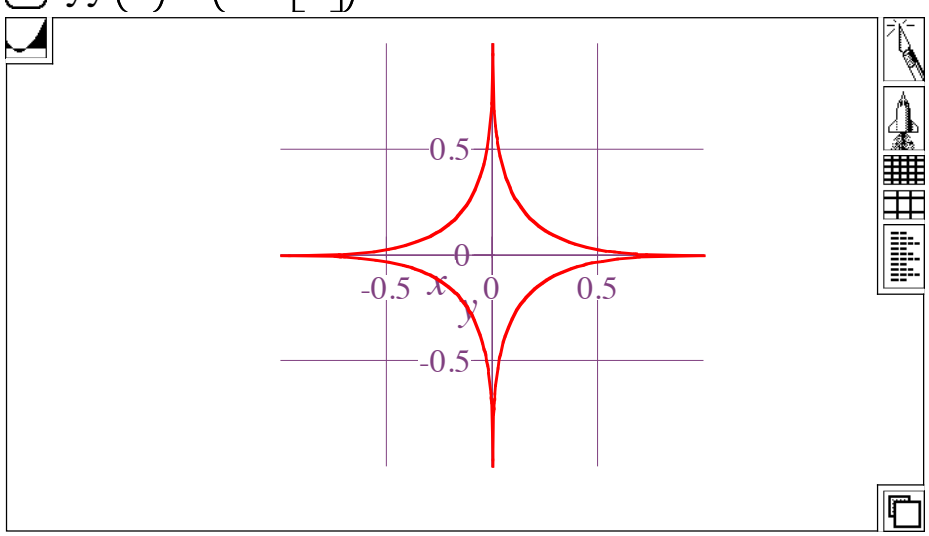
$x^{2/5} + y^{2/5} = 1$:

$xlow = -1$

$xhigh = 1$

$xx(t) = (\cos[t])^5$

$yy(t) = (\sin[t])^5$



Use uniformly distributed points to estimate the measurement of the area enclose

Answer

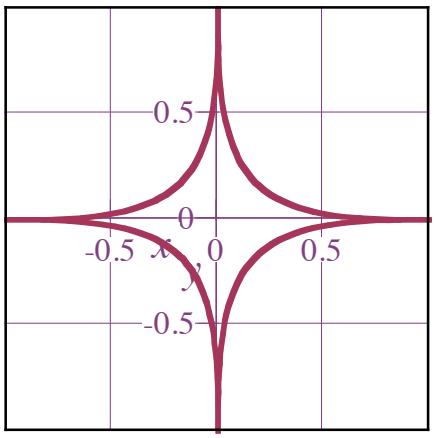
Put a box around the circle:

$xlow = -1$ $xhigh = 1$

$ylow = -1$ $yhigh = 1$

$xx(t) = (\cos[t])^5$

$yy(t) = (\sin[t])^5$



Generate 2500 uniformly distributed random points inside the box:

$xlow = -1$ $xhigh = 1$

$ylow = -1$ $yhigh = 1$

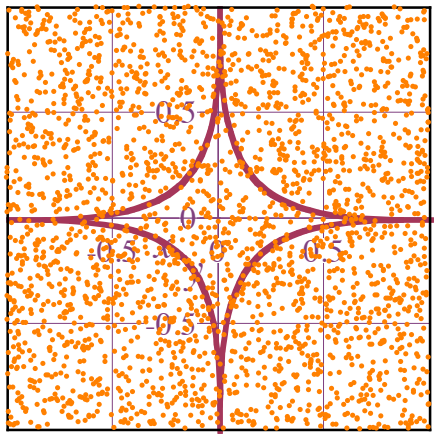
$xx(t) = (\cos[t])^5$

$yy(t) = (\sin[t])^5$

LiveMath Note: Using the functional approach to generating random numbers as demonstrated in the previous example

$xRandoms(k) = Random(xlow, xhigh)$

$yRandoms(k) = Random(ylow, yhigh)$



– 1 ... 1 = left...right True Proportions

– 1 ... 1 = bottom...top cropping Moderately

Graph Building Blocks

Curve at (xx[t], yy[t]) where $t = 0 \ldots 2\pi$ with a extra heavy line, colored Other... .

Box

Scatter plot of (xRandoms[k], yRandoms[k]) where $k = 1 \ldots 2500$ using 2 point spots colored Orange .

The Monte carlo idea:

Because the points are approximately uniformly distributed, you can expect:

$$\frac{\text{Area enclosed by curve}}{\text{Area enclosed by the box}} \approx \frac{\text{Number of random points inside curve}}{\text{Total number of random points inside box}}$$

so that:

$$\text{Area enclosed by curve} \approx \left(\frac{\text{Number of random points inside curve}}{\text{Total number of random points inside box}}\right).$$

Try it out remembering that:

$$x^{2/5} + y^{2/5} = 1$$

is the same curve as:

$|x|^{2/5} + |y|^{2/5} = 1$:

$xlow = -1$ $xhigh = 1$

$ylow = -1$ $yhigh = 1$

$xx(t) = (\cos[t])^5$

$yy(t) = (\sin[t])^5$

LiveMath Note: Using the functional approach to generating random numbers as demonstrated in the previous example

$xRandoms(k) = Random(xlow, xhigh)$

$yRandoms(k) = Random(ylow, yhigh)$

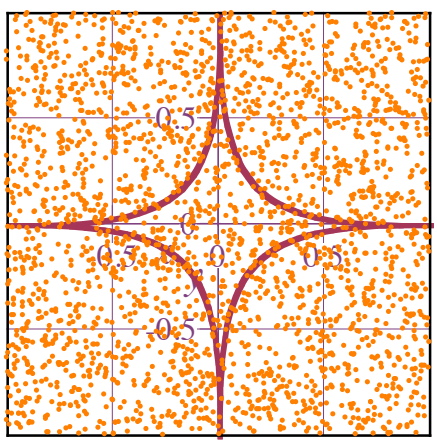
$$fCounts(n) = \sum_{k=1}^{2500} \left(|xRandoms[k]|^{\frac{2}{5}} + |yRandoms[k]|^{\frac{2}{5}} \le 1 \right)$$

$BoxArea = (xhigh - xlow)(yhigh - ylow)$

$$AreaEst(n) = \frac{fCounts(n)}{2500} BoxArea$$

$AreaEst(1)$

$AreaEst(1) = 0.392$ *Calculate*



Now, average to try to get a better estimate:

$xlow = -1$ $xhigh = 1$

$ylow = -1$ $yhigh = 1$

$xx(t) = (\cos[t])^5$

$yy(t) = (\sin[t])^5$

LiveMath Note: Using the functional approach to generating random numbers as demonstrated in the previous example

$xRandoms(k) = Random(xlow, xhigh)$

$yRandoms(k) = Random(ylow, yhigh)$

$fCounts(n) = \sum_{k=1}^{2500} \left(|xRandoms[k]|^{\frac{2}{5}} + |yRandoms[k]|^{\frac{2}{5}} \le 1 \right)$

$BoxArea = (xhigh - xlow)(yhigh - ylow)$

$AreaEst(n) = \frac{fCounts(n)}{2500} BoxArea$

$B_1 = (AreaEst[1], AreaEst[2], AreaEst[3], AreaEst[4], AreaEst[5])$

$B_1 = (0.3936, 0.3952, 0.368, 0.3232, 0.3584)$ *Calculate*

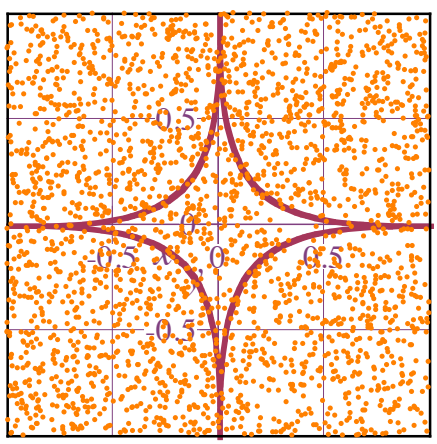
$B_2 = (AreaEst[6], AreaEst[7], AreaEst[8], AreaEst[9], AreaEst[10])$

$B_2 = (0.3568, 0.3264, 0.3824, 0.3472, 0.352)$ *Calculate*

The average of the 10 Monte Carlo estimates is:

$$Average = \frac{\sum_{k=1}^{10} AreaEst(k)}{10}$$

$Average = 0.37088$ *Calculate*



If you've had a strong calculus course such as Calculus&LiveMath: VectorCal you are in a position to check this estimate.

You parameterize the plotted curve $x^{2/5}+y^{2/5} = 1$ via the parameterization:

$$\{x(t), y(t)\} = \{\cos(t)^5, \sin(t)^5\}$$

with $0 \le t \le 2\pi$ and then calculate:

$$\int_0^{2\pi} x(t)\, y'(t)\, dt:$$

xlow $= -1$ xhigh $= 1$

ylow $= -1$ yhigh $= 1$

$xx(t) = (\cos[t])^5$

$yy(t) = (\sin[t])^5$

$xx'(t) = \left.\frac{d}{dt} xx(t)\right|_{t=t}$

$xx'(t) = -5(\cos[t])^4 \sin(t)$ *Substitute*

$yy'(t) = \left.\frac{d}{dt} yy(t)\right|_{t=t}$

$yy'(t) = 5\cos(t)(\sin[t])^4$ *Substitute*

$\int_0^{2\pi} xx(t)\, yy'(t)\, dt$

$\int_0^{2\pi} xx(t)\, yy'(t)\, dt = 0.368155387771826$ *Calculate*

Because of the Law of Large Numbers, the smart money bets that the average