Today we start a new topic whose relevance has gained more and more importance

The topic goes under the general name *Monte Carlo techniques.* We will apply these techniques to two different cases:

1. Integration
2. Monte Carlo simulation

Monte Carlo techniques are used widely in many fields. The surprising thing about these techniques is that they rely on *random numbers* to make predictions about *deterministic systems*!

Techniques vary widely, but a sort of general pattern is:

1. Generate *random* inputs from a *probability distribution function*.
2. Do some sort of deterministic computation on the inputs
3. *Aggregate* the results.

Before we can apply these techniques, we have to understand what we mean by these

Random Number: A *random number* is a number *generated by a process*, whose outcome is unpredictable, and which cannot be subsequentially reliably reproduced.

There's a lot here. What is the process? What does it mean to be unpredictable and not reproducible?

We are going to call the process, the *probability distribution function (PDF).* This function generates a pool of numbers. A *random number* is a number randomly chosen from this pool of numbers. The *likelihood* of any particular number being chosen (or drawn) is determined by the *PDF.*

A fair die, our *PDF*

$$p(x) = \begin{cases} 1/6, & x \in \{1, \ldots, 6\} \\ 0 & \text{otherwise} \end{cases}$$

- *p(x)* is the *PDF*. Notice that there is a relationship between probability and random numbers.
- This *PDF* is *discrete*. It generates a countable set of numbers
- In this case each number is *equally likely* to appear. This is not always the case.

In the previous example, the probability of drawing a 3 is exactly 1/6. This is a feature of discrete *PDF.*

Continuous PDF form a pool of numbers that are not countable. Before we go to discuss these, do question (1) on the worksheet and S T O P

a) Countable means that the number of random numbers can be put one-to-one with integers. Basically discrete PDF arise when counting, continuous when measuring.
b) Lots of different ones
c) Most likely to occur is the 2, but any of the other numbers is as likely to be drawn as 2.
d) 0.

For continuous PDF, the likely of drawing a specific number is *zero* because there are uncountably many to choose from. For continuous *PDF*, the appropriate question to ask is what is the likelihood of drawing a number in some interval. So we have instead

$$Pr(b < x < a) = \int_a^b \underbrace{p(x),}_{PDF} \mathrm{d}x$$

PDF stuff:

o The random variable *x* must take on some value, so that integrating the *PDF* over the entire range gives:

$$\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1$$

o There are two important quantities that give information about the *central tendencies* of random numbers

(1) The central tendency is the *mean* given by

$$\mu_x = \int x \cdot p(x)\,\mathrm{d}x,$$

(2) The spread of the data about the mean, called the *variance* and given by

$$\sigma_x^2 = \int (x - \mu_x)^2\, p(x)\,\mathrm{d}x,$$

Not all *PDF* have a mean and variance. For example the *Cauchy PDF*, $p(x) = \dfrac{1}{\pi(1+x^2)}$ has infinite mean and variance

Do question(2) on the worksheet and S T O P

(2)

$$\begin{aligned}
\mu_x &= 3\int_0^1 x \cdot x^2 dx \\
&= 3\left.\frac{x^4}{4}\right|_0^1 = \frac{3}{4} \\
\sigma_x^2 &= 3\int_0^1 \left(x - \frac{3}{4}\right)^2 x^2 dx \\
&\vdots
\end{aligned}$$

In most real applications, we do not know the exact *PDF.* However, we can still get an estimate of the central tendencies by finding the *sample mean and variance* given by

$$\bar{x} = \frac{1}{N}\sum_i^N x_i \quad \text{sample mean}$$

$$\sigma_N^2 = \frac{1}{N-1}\sum_i^N (x_i - \bar{x})^2 \quad \text{sample variance}$$

Be aware that $\mu_x \neq \bar{x}$ and $\sigma_n^2 \neq \sigma_x^2$. The difference between the sample mean and true mean is given by

$$(\mu_x - \bar{x}) = \frac{\sigma_N}{\sqrt{N}}$$

where *N* is the number of samples. Thus the more data you have, the closer you get to the true mean.

Do question (3) on the worksheet and S T O P

In many cases, we have more than one random element. In such cases, we can form a *random vector*,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \mu_{\mathbf{x}} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

But the variance is more complicated because of interdependence between variables

$$\Sigma_{\mathbf{x}} = \underbrace{\begin{pmatrix} \sigma_1^2 & \sigma_1\,\rho_{12}\,\sigma_2 & \sigma_1\,\rho_{13}\,\sigma_3 & \dots \\ \sigma_2\,\rho_{21}\,\sigma_1 & \sigma_2^2 & \sigma_2\,\rho_{23}\,\sigma_3 & \dots \\ \sigma_3\,\rho_{31}\,\sigma_1 & \sigma_3\,\rho_{32}\,\sigma_2 & \sigma_3^2 & \dots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}}_{\text{covariance matrix}} \qquad -1 < \rho_{ij} < 1$$

In the course notes many different distributions are discussed. Here we only look at the *uniform distribution* because it can be used to construct many of the others.

A *Uniform Distribution* is a distribution in which all the possible outcomes are equally probable . If the range is set to $a < x < b$, then the probability of drawing any number in this range is $1/(b - a)$.

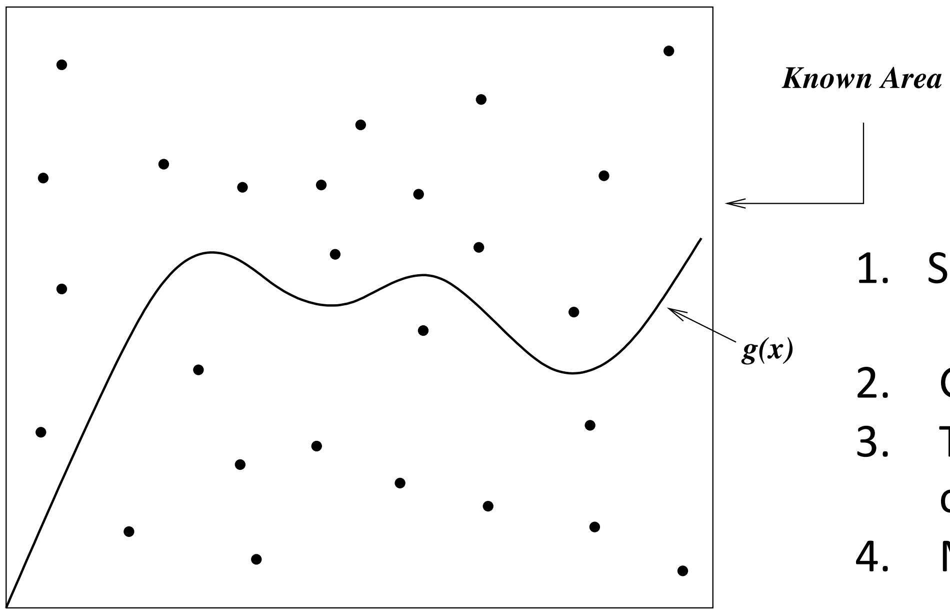
Question (4) on the worksheet will give an example of how to use this distribution to generate random numbers from an a *PDF.* Do question (4) on the worksheet and S T O P

```
function x = genx2
%Generates random variables using rejection method and pdf of x^2

j = 1;
while j<=100
   rx = 1.1*rand;
   ry = 1.1*rand;
   if ry<3*rx^2 % pdf is 3 x^2
      x(j) = rx;
      j = j + 1;
   end
end
avgx = mean(x);
varx = var(x);
fprintf('The average is %g6.4\n', avgx)
fprintf('The variance is %g6.4\n', varx)

end
```

Integration—Method I: We are now ready to start applying Monte Carlo methods to some important problems.

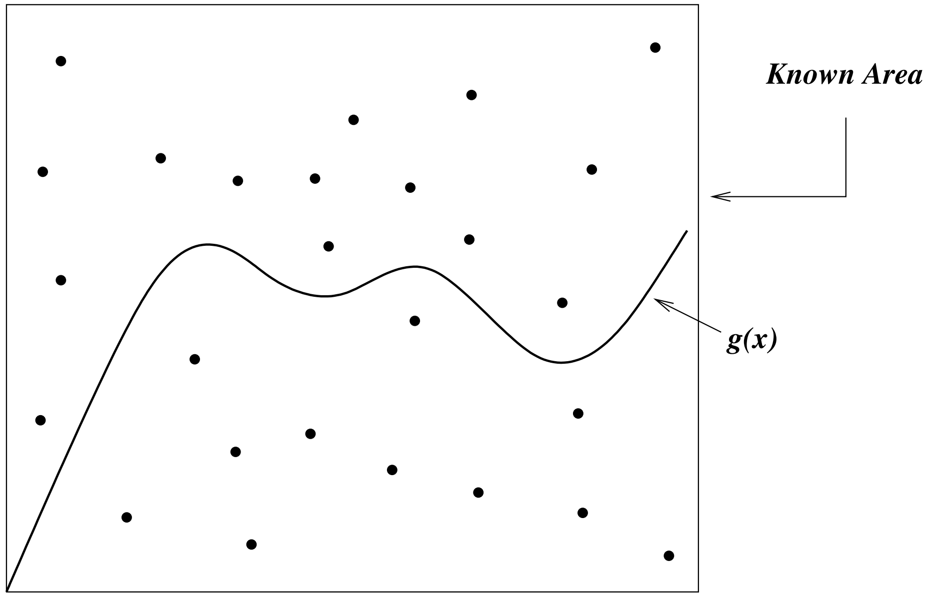
The exercise you just did on the worksheet has prepared you for one of the two methods we'll use to numerically integrate functions using Monte Carlo techniques.

Suppose you need to integrate a complicated function in some interval. The rejection method we just developed can be used to this. Consider the figure shown below. Again, *the curve* is the function we want to integrate, and we enclose the function in a *rectangle* of know length and width.



Known area, the rectangle, encircles the function, g(x), whose integral we want to find.

1. Set up a *region* whose *area is easy to find*, such as a rectangle
   i. Make sure it fully encloses the function you wish to integrate
2. Generate random points
3. Take the *ratio* of those points that fall *underneath* the curve to the total number of points
4. Multiply the ratio by the known area.



Known area, the rectangle, encircles the function, g(x), whose integral we want to find.

1. Suppose area of rectangle = *10*,
2. You generate *1000* random *(x_r, y_r)* pairs of points that lie within rectangle.
3. You then calculate *$g(x_r)$* and count how many *$y_r < g(x_r)$*. Suppose you found *673* of these points.
4. What would you estimate the area of *g(x)* to be:

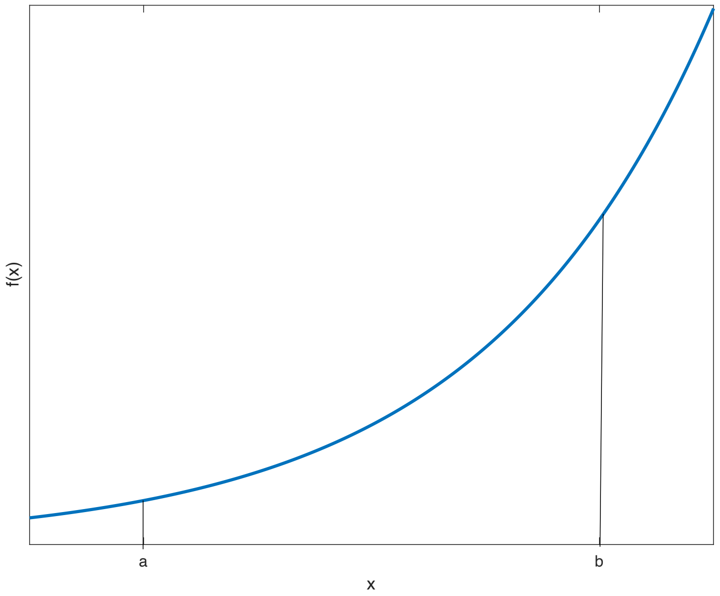
$$\text{Area } g(x) \approx \underbrace{\frac{673}{1000}}_{\text{ratio}} \times \underbrace{10}_{\text{area of rectangle}} = 6.73$$

Do questions (5) worksheet and S T O P

(5)

```
function MyMonteRegion(N,a,b)
% A simple MonteCarlo integrator

fun = @(x) (x.*exp(x));

rand('state',0);
intsum = 0;
x1 = 0:.01:1;
fofx = x1.*exp(x1);
l = exp(1);     % rectangle is 1 x exp(1)
plot(x1,fofx,'k')
hold on
for n = 1:N
    x = a+(b -a )*rand; % generating a random (x,y) in the rectangle
    y = (l)*rand;
    plot(x,y,'.r','MarkerSize',8)
    hold on
    if y<=fun(x)
        intsum = intsum + 1;
    end
end
montarea = exp(1)*intsum/N     % ratio of points x area

end
```

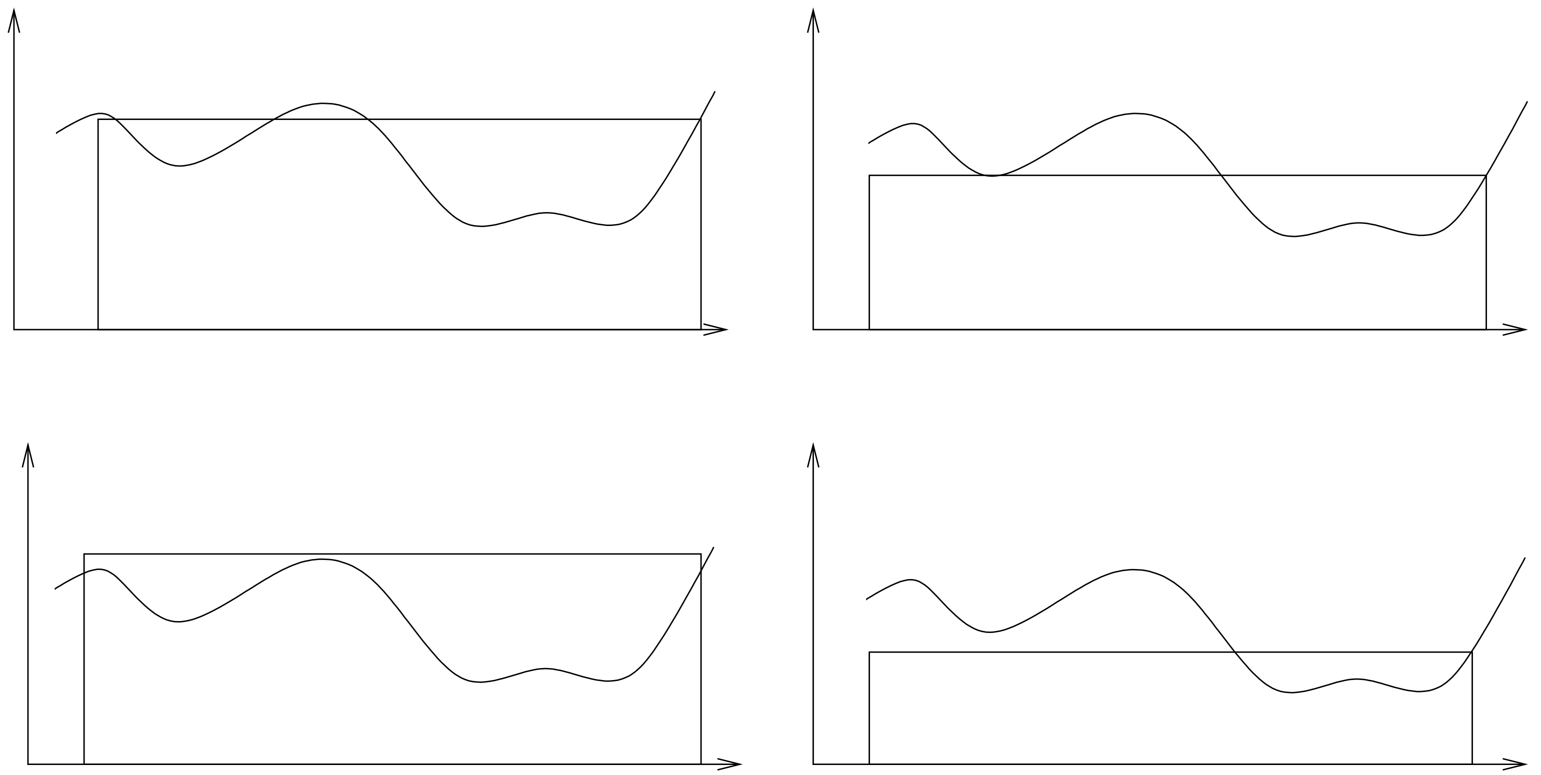
Integration—Method II: Method I is easy to implement but there is no easy way to estimate the error in the process.



$$\langle f \rangle = \frac{1}{(b-a)} \int_a^b f(x)\,\mathrm{d}\,x \;\Rightarrow\; \boxed{\int_a^b f(x)\,\mathrm{d}\,x = (b-a)\langle f \rangle}$$

$$\int_a^b f(x)\,\mathrm{d}\,x \approx \frac{(b-a)}{N} \sum_{i=1}^{N} f(x_i)$$

Easy. One picks *N* random numbers between *a, b*, calculates the sample mean, and multiplies by the length of the entire interval to find the integral



But because we are computing sample mean, we can get an estimate of the error in our calculations.

$$\int_a^b f(x)\,\mathrm{d}\,x \approx (b-a)\langle f\rangle \pm (b-a)\sqrt{\frac{\langle f^2\rangle - \langle f\rangle^2}{N}}$$

Do question (6) on the worksheet and S T O P

Modifications;

- We can use *PDF* that more closely matches the function we are trying to integrate
- We can segment the function and use our algorithm in each section. This allows for better sampling
- You can combine the previous two steps.