Astro 426/526

Fall 2019

Prof. Darcy Barron

Lecture 18: Wrap-up of analysis techniques

Miscellaneous

- Class in P&A white board classroom
- Homework from *Practical Statistics* will be posted tonight
  - Due in 2 weeks (due at start of class Wed Nov 13)
  - 5 data analysis problems, to make sure you know/are aware of techniques you should be using in your project
  - Email me to set up office hours if you need help
- Full project requirements will be posted tonight
  - No surprises, just a full write-up on expectations and grading rubric
- Today: cover the rest of the relevant analysis topics from *Practical Statistics for Astronomers*
- Next week: return to *Measuring the Universe*

Groups

- Marsquakes – Vincent, Bruno, Ryan
- Radio emission from Exoplanets – Ivey, Prescott, Susan
- Variability of Massive Stars – Pratik, Sanja

Random number generation

- Why generate random numbers?
- Often when fitting models, you want a set of numbers that are distributed how we might guess the real data will be
  - Check error propagation
  - Test our code
  - Test some technique
  - Determine expected S/N or significance
- Two kinds of random numbers
  - Uniformly distributed
  - Drawn randomly from a parent population of known frequency distribution

Pitfalls of random number generators

- Usually some function x = rand(i)
- Varying levels of "randomness"
  - RANDU
  - Lava lamp random numbers
- Any routine is generating pseudo-random numbers
- Cycle length – how long until the pseudo random cycle repeats?

Slides from Practical Statistics

Random numbers from a given distribution

How do we draw a set of random numbers following a **given** frequency distribution?

Suppose we have a way of producing random numbers that are uniformly distributed, in say the variable α; and we have a functional form for our frequency distribution $dn/dx = f(x)$. We need a transformation $x = x(\alpha)$ to distort the uniformity of α to follow $f(x)$. But we know that

$$\frac{dn}{dx} = \frac{dn}{d\alpha}\frac{d\alpha}{dx}$$

and as $dn/d\alpha$ is uniform, thus

$$\frac{dn}{dx} = \frac{d\alpha}{dx},$$

and

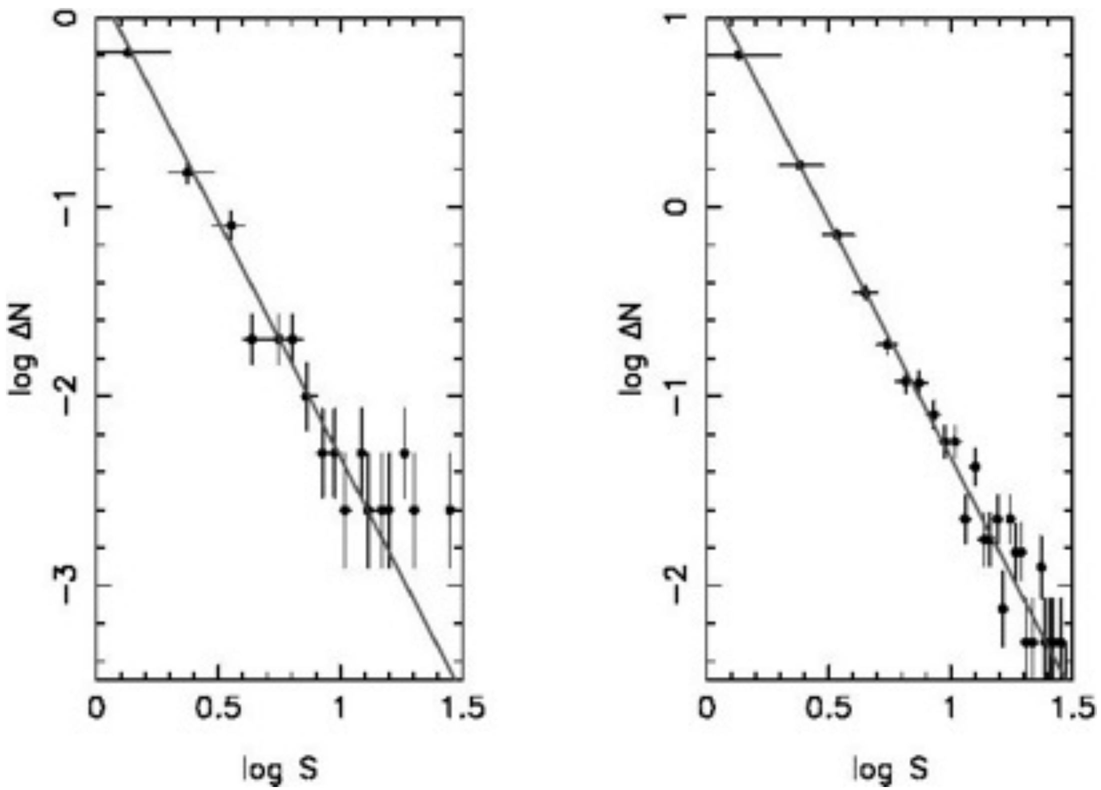
$$\alpha(x) = \int^x f(x)dx,$$

so that the required transformation is $x = x(\alpha)$.

We therefore need $x = f^{-1}(\alpha)$, the inverse function of the integral of $f(x)$.

Randoms from a given distribution 2

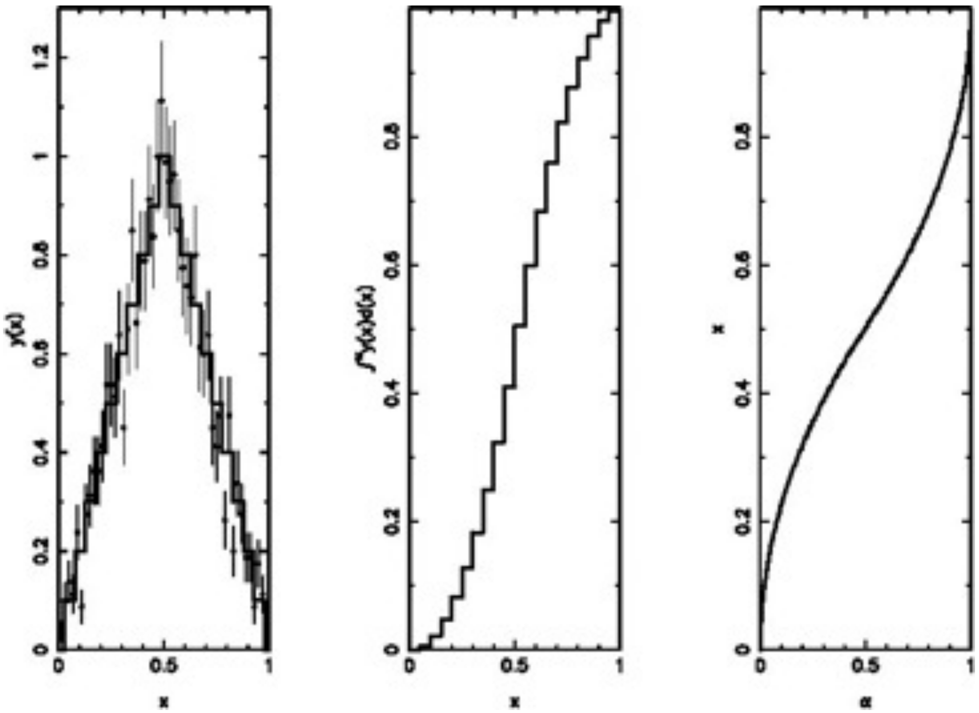
EXAMPLE: A source-count distribution is given by $\mathbf{f(x)dx = -1.5x^{-2.5}dx}$, a `Euclidean' differential source count. Here $\mathbf{d\alpha = -1.5x^{-2.5}dx}$, $\mathbf{\alpha = x^{-1.5}}$, and the transformation is $\mathbf{x = f^{-1}(\alpha) = \alpha^{-1/1.5}}$.



Differential source counts generated via Monte Carlo sampling with an initial uniform deviate, obeying the source-count law $N(>S) = kS^{-1.5}$. The straight line in each shows the anticipated count with slope -2.5. left: k = 1.0, 400 trials, right: k = 10.0, 4000 trials.

Randoms from a given distribution 3

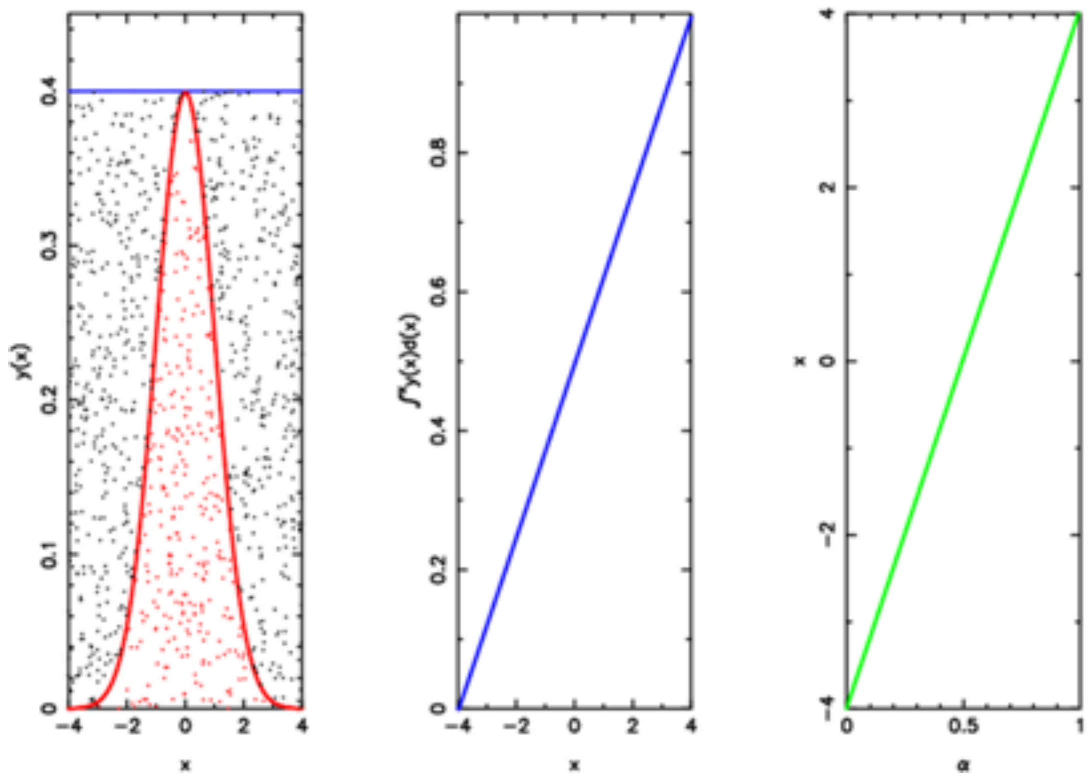
The very same procedure works if we don't have a functional form for f(x)dx.
If this is a histogram, we need simply to calculate the integral version, and perform the reverse function operation as before.
EXAMPLE:



An example of generating a Monte-Carlo distribution following a known histogram. Left: the step-ladder histogram, with points from 2000 trials, produced by a) integrating the function (middle) and b) transforming the axes to produce f^{-1} of the integrated distribution (right). The points with √N error bars in the left diagram are from drawing 2000 uniformly-distributed random numbers and transforming them according to the right diagram.

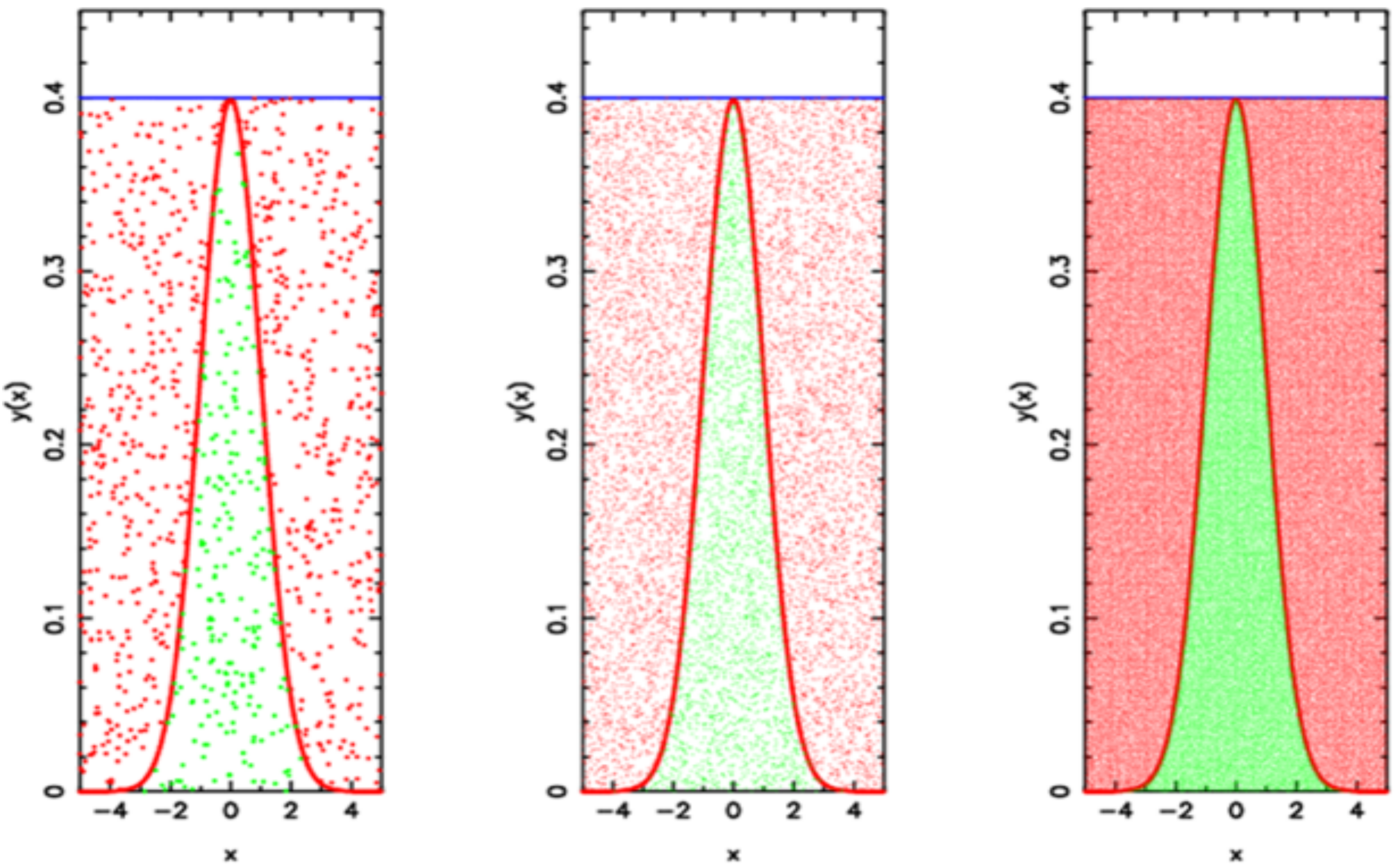
Randoms from a distribution - rejection method

1. Plot the culprit distribution – non integrable, nasty
2. Find a nicely-behaved (integrable) one which looks a bit similar and lives higher.
3. Get the random values of x for this one, via the transform method.
4. Find a random distance up the y-axis for each of these x values
5. Reject the ones which lie outside the require distribution.



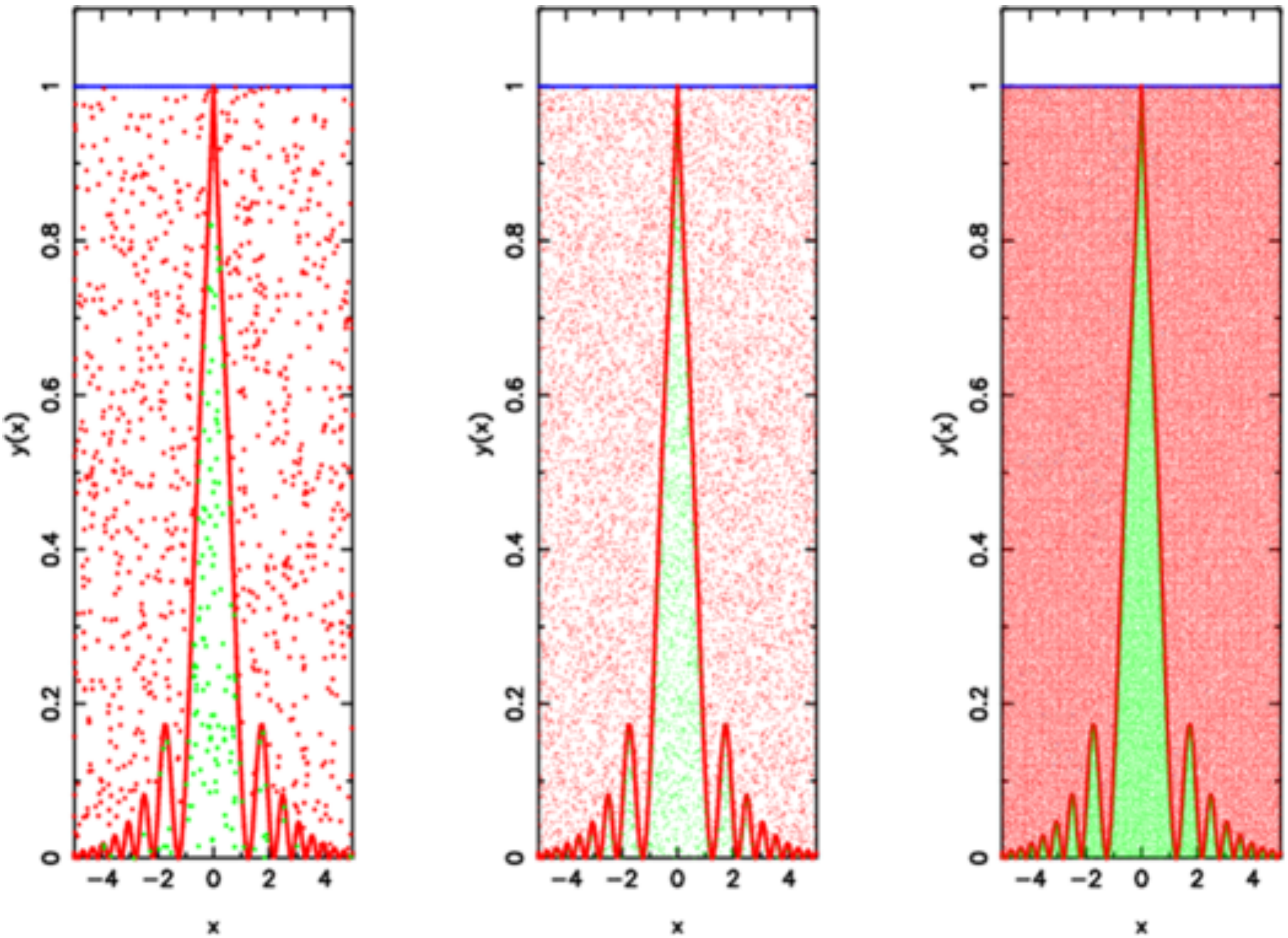
Example 1: Gaussian to +/- 5σ with elevated function (= uniform). Centre, right – transformation of the uniform function to give uniform x (and y) coverage. 1000 trials; 241 points within Gaussian; hence area estimate = (241/1000) x (0.4 x 10.0) = 0.9640.

Rejection method - Example 1 continued



Gaussians out to +/- 5 sigma. Left 1000 points, 241 in zone, area = (241/1000) x (10.*0.4) = 0.9640; Centre 10000 points, 2540 in zone, area = 1.0160; Right 10^6 points, 250830 in zone, area = 1.0015

Randoms via rejection - Example 2

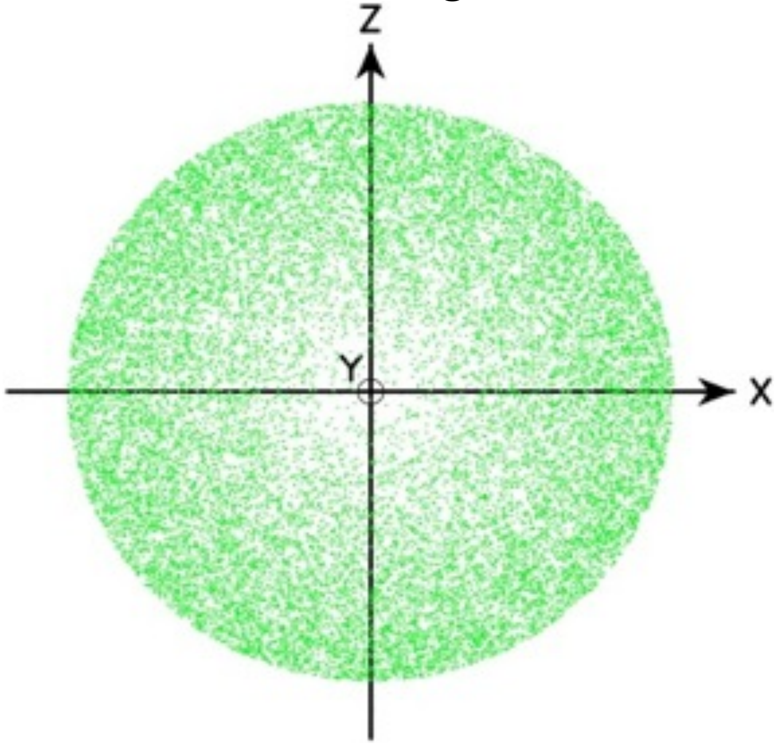


$f(x)=\exp^{-|x|}\cos(x^2)$. Left, 1000 points, 152 in zone, area = (152/1000) x (10.*1.0) =1.5200; Centre 10000 points, 1444 in zone, area = 1.4440; Right 10^6 points, 140229 in zone, area = 1.4023.

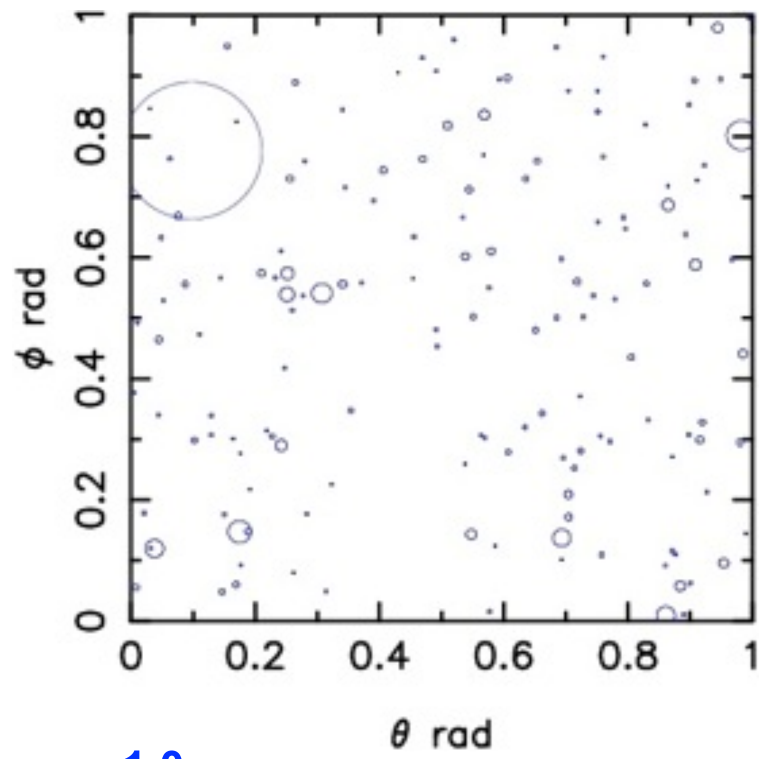
Randoms from a distribution - Example 3

Introducing the toy universe

Thin slice though centre



A view of the "sky"



Assume a flat Euclidean universe with $r_{max} = 1.0$.

Populate this with 10^6 objects randomly but uniformly distributed, i.e. 10^6 values of $(r_i, \theta_i, \varphi_i)$

Assign each a luminosity L = 1.0, so that each produces a flux (at the centre) of $1/r_i^2$

Monte Carlo (random-number) Integration

Numerical integration is a very important use of Monte Carlo.

Highly technical! Outline here only.

Suppose we have a probability distribution **f(x)** defined for **a** $<$ **x** $\leq$ **b**.
Draw **N** random numbers **X**, uniformly distributed between **a** and **b**, and calculate the function at these points.
Add these values of the function up, normalize - and this is our answer.

$$\int_a^b f(x)\,dx \simeq \frac{(b-a)}{N}\sum_i f(X_i).$$

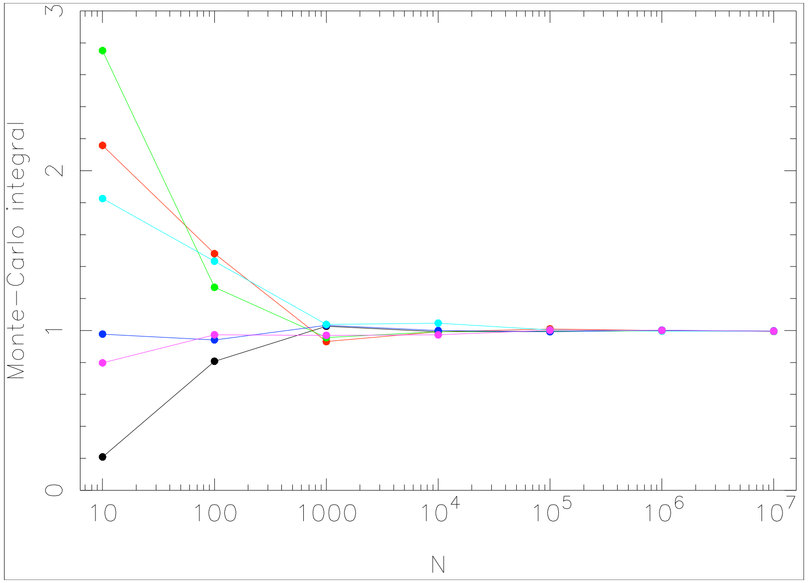
This is **Monte Carlo integration**.

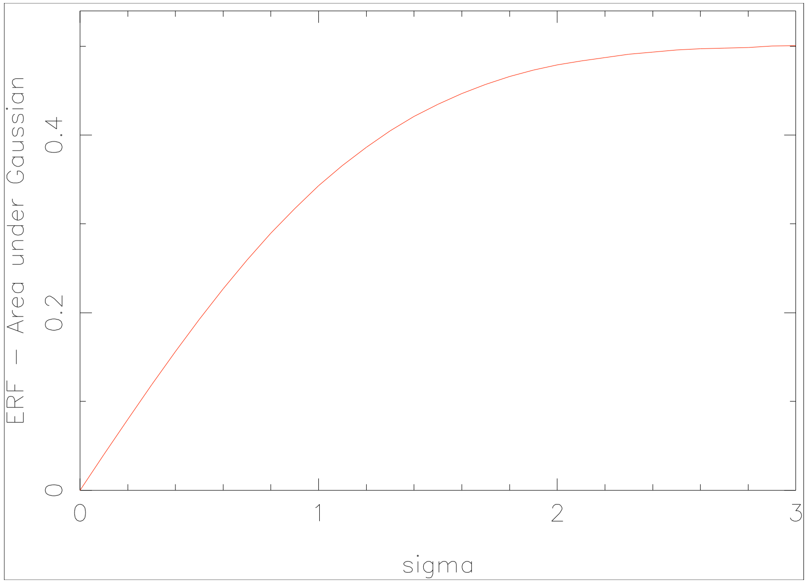
If the $\mathbf{X}_i$ are drawn from the distribution f itself, then they will sample the regions where **f** is large and the integration (for the same number of points) will be far more accurate. This technique is called importance sampling.

M C Integration - Example (Gaussian)

Use a uniform random-number generator such as the function routine *ran1* of *Numerical Recipes*; make N calls to it, scaling the $(0 \rightarrow 1)$ random numbers to the range of σs required, say $k\sigma$. For each resulting value x_i, compute $f(x_i) = \frac{1}{\sigma\sqrt{2\pi}}\exp[-\frac{x_i^2}{2\sigma^2}]$. The integral from 0 to $k\sigma$ is simply

$$\frac{k}{N}\sum_{i=1}^{N} f_i(x).$$





Right – the result, using N=10^6. Left – using +/- 10σ, and varying N. The different curves are the results of different starting indices for the random-number generator. This mindless sum shows how stable MC integration is for well-behaved functions; we have uniformly sampled +/- 10σ, and the function is really a spike between +/- 2σ.