ASTR4004/ASTR8004
Astronomical Computing
Lecture 09

Christoph Federrath

26 August 2016

Statistics and Monte Carlo simulations in IDL

1 Statistical functions and PDFs

1. We will use the data files from the previous lecture:
   EXTREME_proj_xy_000100
   EXTREME_proj_xy_000200
   EXTREME_proj_xy_000300
   These data are column density (2D) projections of the 3D gas density in the simulations. The 3 files are at 3 different times of the evolution. If you are interested in more details on these data and the simualtions, please have a look at Federrath et al. (2016) and references therein.

2. Compute the mean, standard deviation, skewness and kurtosis of the log-normalised column density (intensity) for these data,

   $$\mathcal{I} = \ln\left(I/\langle I\rangle\right), \tag{1}$$

   where $\langle I\rangle$ is the mean column density (intensity) and ln() is the natural logarithm (`alog` in IDL). Use the built-in functions `mean, stddev, skewness, kurtosis, moment`.

3. Write a procedure that writes the PDF locations, the histogram, the PDF (normalised histogram), and the CMF of $\mathcal{I}$ to an ASCII text file.

4. Write a procedure to read these ASCII files and data columns back into memory.

5. Write a procedure to compute the mean, standard deviation, skewness, and kurtosis of the PDFs by summation over the PDF. Compare the statistical moments of the PDF based on the data and intrinsic functions with those based on the summation over the PDF. Is there a difference? Why? How can this be fixed?

2 Averaging data, making plots with error bars, and fitting

1. Make a procedure/function that computes the PDF of a given dataset for a given bin size, min, and max, and returns the histogram, the PDF, the CDF, and the bin locations.

2. Average all the three column density PDFs (each computed with the new PDF-procedure) over time to produce a time-averaged PDF with error bars.

3. Produce a plot of the PDF including error bars.

4. Use the MPFIT library to write a procedure to fit the time-averaged PDF with a Gaussian function. Over-plot the fit.

3 Monte Carlo error propagation

1. Uncertainty analysis of measurements and derived quantities In astrophysics is extremely important, and published measurement results should always have a proper error analysis associated with them. Here we learn how to use IDL to do Monto Carlo (MC) error propagation, which goes beyond standard analytic error propagation and can handle non-Gaussian distributions for propagating uncertainties.

2. Suppose you have measured two Gaussian random variables,

$$a = 10 \pm 1 \qquad \text{and} \qquad b = 1.0 \pm 0.1, \tag{2}$$

   and you want to calculate the derived quantity

$$c = \frac{a^2}{b^2}. \tag{3}$$

3. First, calculate the uncertainty of c using standard analytic methods of error propagation.

4. Now write a program that does the MC error propagation. First make Gaussian random numbers and define a and b based on these Gaussian random distributions. Then define c based on a and b.

5. Now plot the PDFs of a, b, and c. What is special about the PDF of c? Try also a log-y axis version.

6. Compute the mean and standard deviation of c based on the PDF of c and compare to the analytic estimate.

7. Get the mode (most probable value) of c and the 16th and 84th percentile values.

8. Try changing the sample size of the Gaussians (`NRAND`) and the number of bins used for the PDFs, in order to study numerical convergence of the results.

Make use of `stop` in your scripts to turn from automatic script mode to interactive mode. Use `IDL> .cont` to continue the script after `stop`. Search function and procedure help to learn about optional keywords, which control the behaviour of functions/procedures.