Subject: Re: Some histogram magic help required - gridding/counting large dataset
Posted by David Fanning on Tue, 25 Mar 2014 15:28:04 GMT
View Forum Message <> Reply to Message

rjp23@le.ac.uk writes:

> I have 2 large datasets that I want to plot as the x and y data on a scatter plot.
>
> However I'd also like to grid this data in the x and y direction and count how many datapoints fall into each grid cell.
>
> I know I should be able to do this with hist_nd but I just can't figure out exactly how to do it

I believe your thinking about this is All Wrong. :-)

Just histogram each 1D data set into the same number of bins.

```
  h1 = cgHistogram(data1, NBins=100, Reverse_Indices=ri1)
  h2 = cgHistogram(data2, NBins=100, Reverse_Indices=ri2)
```

When you want to find out, for example how many "hits" you have in bin 25 in the first data set and bin 45 in the second data set, you do this:

```
  b25indices = cgReverseIndices(ri1, 24, COUNT=c1)
  b45indices = cgReverseIndices(ri2, 44, COUNT=c2)

  IF (c1) GT 0) && (c2 GT 0) THEN BEGIN
    indices = cgSetIntersection(b25indices, b45indices, COUNT=count)
    IF count GT 0 THEN BEGIN
      Print, 'Matches: ', count
    ENDIF ELSE Print, 'No matches.
  ELSE Print, 'No matches'
```

If you want to plot the points in this intersection of bins:

```
  cgPlot, data1[indices], data2[indices], PSYM=1
```

Cheers,

David
--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.idlcoyote.com/
Sepore ma de ni thue. ("Perhaps thou speakest truth.")