Subject: Some histogram magic help required - gridding/counting large dataset
Posted by rjp23 on Tue, 25 Mar 2014 10:11:03 GMT
View Forum Message <> Reply to Message

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.

Cheers

Rob

Subject: Re: Some histogram magic help required - gridding/counting large dataset
Posted by Andy Sayer on Tue, 25 Mar 2014 12:56:48 GMT
View Forum Message <> Reply to Message

If it is a 2D (x/y) dataset you could use the built-in function hist_2d:
http://www.exelisvis.com/docs/HIST_2D.html

What exactly is the issue you are having? You can specify bin sizes and min/max values with keywords min1, max1, bin1, min2, max2, bin2.

Andy

On Tuesday, March 25, 2014 6:11:03 AM UTC-4, rj...@le.ac.uk wrote:
> 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.
>
>
>
> Cheers
>
>
>
> Rob

Subject: Re: Some histogram magic help required - gridding/counting large dataset
Posted by rjp23 on Tue, 25 Mar 2014 14:59:41 GMT
View Forum Message <> Reply to Message

It's two 1D datasets with e.g. 20,000 values.

How I've used hist_nd before was to put these into a 2D dataset [20,000, 2] and then run hist_nd on that data but that doesn't seem to work and is only returning 1 bin, even though I'm specifying e.g. [100,2] bins.

On Tuesday, March 25, 2014 12:56:48 PM UTC, AMS wrote:
> If it is a 2D (x/y) dataset you could use the built-in function hist_2d:
http://www.exelisvis.com/docs/HIST_2D.html
>
>
>
> What exactly is the issue you are having? You can specify bin sizes and min/max values with keywords min1, max1, bin1, min2, max2, bin2.
>
>
>
> Andy
>
>

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.")

Subject: Re: Some histogram magic help required - gridding/counting large dataset
Posted by Phillip Bitzer on Tue, 25 Mar 2014 16:10:52 GMT
View Forum Message <> Reply to Message

On Tuesday, March 25, 2014 5:11:03 AM UTC-5, rj...@le.ac.uk wrote:
> 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.
>
>
>

Here's a snippet for how to do what (I think) you're asking:

https://www.dropbox.com/s/j9ffh1ns3wcguq4/histo_notes.pdf

In the notes, h2d will contain the counts per bin.
