Subject: Pair Counts in an Annulus, for large data sets
Posted by fatcat3131 on Sat, 11 Nov 2006 05:52:55 GMT
View Forum Message <> Reply to Message

Hi There

I have a large data set (~350,000 galaxies) of x,y points. For a given radius R [R=sqrt(x^2+y^2)], I need to count the total number of pairs in the annulus R+deltaR. That is, I choose a given data point as my center, then count the number of points that lie inside that annulus. I then do this for each of my data points to get the total number of pairs. A simplified version of the code I'm using now is as follows:

```
****************
n = n_elements(x) ; number of data points
seperation = fltarr(n) ; the seperation between data points

for i=0L,n-1 do begin
  seperation = sqrt((x - x[i])^2 + (y - y[i])^2) ;distance between the
two points, centering on point "i"
  seperation[i] = 999 ;simply because I don't want it to count itself
as a pair

  if_inside = ((seperation gt (r)) and (seperation lt (r + deltar_r))
;has value "1" for points which lie inside, "0" for those outside
  counter = counter + total(if_sep) ;count up the number of pairs
endfor

num_pairs = counter / 2 ; since I don't want to count everything twice
*****************
```

I've tried my best to avoid the urge to put lots of for loops everywhere (you should have seen it before!), but I just don't know how to make it drastically more efficient. There must be a way though, because the computations for my code are just ridiculous.... Is there a way to eliminate that nasty loop I have, which would help things?

Any help you can give would be greatly appreciated. I'm very new to IDL, as you surely know. I'm and undergrad, too.

Tara.

Subject: Re: Pair Counts in an Annulus, for large data sets
Posted by David Fanning on Thu, 23 Nov 2006 15:58:25 GMT
View Forum Message <> Reply to Message

fatcat3131@gmail.com writes:

> i've been tinkering with things, and i simply don't understand the
> histogram principle (i even read the tutorial!). would you mind
> explaining in slightly more detail?
>
> it's really not clicking with me how setting the binsize to Delta-R
> would allow me to count the points which lie in the distances (R to
> Delta-R) from a point P, and do this for all N points. i am interested
> in just one Delta-R, and counting the point-point pairs for this
> Delta-R.

Humm. Let's see. You have a data set containing the
distance from some point, R. Suppose your delta-R, or
bin size, is 1. If you run a histogram on R, it will
tell you how many Rs are in the bin 0 to 1, 1 to 2, etc.
Moreover, the reverse indices of the histogram will tell
you which points in R are in each bin.

Suppose, for example, you are interested in which points
have a radius between 2 and 3 in your R vector.
This would be the third bin of the histogram:

```
h = Histogram(R, BINSIZE=1, OMIN=0, REVERSE_INDICES=ri)
Print, 'Number of points with radius between 2-3: ', h[2]
```

Now, which points in R are in this bin? That information is contained
in the reverse indices:

```
indices_of_R_in_Bin3 = ri[ri[2]:ri[3]-1]
```

Cheers,

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

Subject: Re: Pair Counts in an Annulus, for large data sets
Posted by fatcat3131 on Fri, 24 Nov 2006 02:47:27 GMT
View Forum Message <> Reply to Message

Right. And that makes perfect sense to me. The problem is, my "R" data
set is relative to some fixed origin "O". And I want to measure a
radius "Delta-R" from a data-point R(1), not from O. Then count up the
points which lie within "Delta-R" from R(1). Then repeat, measuring

Delta-R from R(2) this time. Then from R(N), etc...

So to me, a histogram on R is counting up the points which lie a certain distance from the origin, no?

If you are in fact interpretting the problem correctly already, then I'm really sorry. Just say so and I'll keep going over it until it "clicks"... I just thought you might be thinking I want to measure Delta-R from the Origin, which is not in fact the case.

Thanks for your help!

Tara (still confused!)

David Fanning wrote:
> fatcat3131@gmail.com writes:
>
>> i've been tinkering with things, and i simply don't understand the
>> histogram principle (i even read the tutorial!). would you mind
>> explaining in slightly more detail?
>>
>> it's really not clicking with me how setting the binsize to Delta-R
>> would allow me to count the points which lie in the distances (R to
>> Delta-R) from a point P, and do this for all N points. i am interested
>> in just one Delta-R, and counting the point-point pairs for this
>> Delta-R.
>
> Humm. Let's see. You have a data set containing the
> distance from some point, R. Suppose your delta-R, or
> bin size, is 1. If you run a histogram on R, it will
> tell you how many Rs are in the bin 0 to 1, 1 to 2, etc.
> Moreover, the reverse indices of the histogram will tell
> you which points in R are in each bin.
>
> Suppose, for example, you are interested in which points
> have a radius between 2 and 3 in your R vector.
> This would be the third bin of the histogram:
>
>   h = Histogram(R, BINSIZE=1, OMIN=0, REVERSE_INDICES=ri)
>   Print, 'Number of points with radius between 2-3: ', h[2]
>
> Now, which points in R are in this bin? That information is contained
> in the reverse indices:
>
>   indices_of_R_in_Bin3 = ri[ri[2]:ri[3]-1]
>
> Cheers,
>

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

Subject: Re: Pair Counts in an Annulus, for large data sets
Posted by David Fanning on Fri, 24 Nov 2006 03:20:43 GMT
View Forum Message <> Reply to Message

fatcat3131@gmail.com writes:

> Right. And that makes perfect sense to me. The problem is, my "R" data
> set is relative to some fixed origin "O". And I want to measure a
> radius "Delta-R" from a data-point R(1), not from O. Then count up the
> points which lie within "Delta-R" from R(1). Then repeat, measuring
> Delta-R from R(2) this time. Then from R(N), etc...
>
> So to me, a histogram on R is counting up the points which lie a
> certain distance from the origin, no?
>
> If you are in fact interpretting the problem correctly already, then
> I'm really sorry. Just say so and I'll keep going over it until it
> "clicks"... I just thought you might be thinking I want to measure
> Delta-R from the Origin, which is not in fact the case.

Oh, I see. I didn't read the second part of your original
post carefully. Sorry.

In this case, I think you are going to need some JD Smith
slight of hand. I would start with this article and see
if you don't get part of the way there:

  http//www.dfanning.com/code_tips/slowloops.html

This article describes a situation similar to yours, and
outlines a couple of possible ways of solving the problem.
With 350000 points, you will want the one that uses the
least amount of memory. -)

Cheers,

David
--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.