Subject: Re: setintersection assumes sets have no repetitions?
Posted by penteado on Tue, 05 Mar 2013 18:44:04 GMT
View Forum Message <> Reply to Message

I just noticed that the documentation changed, to say that indices_a and indices_b require the intersection elements to be unique, while positions does not.

I understand the reasons to assume the sets have unique elements (both from the mathematical concept and from practical applications like identifiers from databases). But since the positions keyword was there to handle repetitions, I got confused when the code crashed on arrays with repetitions.

The use I had with repeated elements happened when I was working with matches of my list of sources in the sky (each with its unique identifier) with other such lists (with their own unique identifiers), by finding which objects in one list were within a certain distance of each object in the other list. Occasionally, one object in my list would have more than one match, and, as such, its identifier would appear more than once in the list of results.

I eventually resorted to using histogram, through my wrapper that puts the reverse indices into hashes or lists:

```
a=[1,2,3,4,5]
b=[1,2,2,2]
h=histogram_pp(a,binsize=1,reverse_hash=rh)
print,rh
;5:           4
;1:           0
;3:           2
;2:           1
;4:           3
ib=lonarr(n_elements(b))
foreach el,b,i do ib[i]=rh[el]
print,ib
;           0           1           1           1
print,b[ib]
;       1       2       2       2
```

histogram_pp is at
http://www.ppenteado.net/idl/pp_lib/doc/histogram_pp.html