Subject: Re: Mode function for floating point arrays
Posted by Rob Klooster on Wed, 17 Jul 2013 14:08:37 GMT
View Forum Message <> Reply to Message

Op dinsdag 9 juli 2013 15:36:46 UTC+2 schreef Matthew Argall het volgende:
> It seems like VALUE_LOCATE and HISTOGRAM solutions would have large limitations. The bin size for HISTOGRAM would have to be "2*epsilon", which would rule out data with a large dynamic range. Also, the bin should be centered on the data point so that two points falling within "epsilon" of one another do not get separated because the bins are offset.

The case of a large dynamical range is precisely the reason why I used VALUE_LOCATE instead of a plain HISTOGRAM with binsize set. Define the function like this:

```
function mode, array
  sortedarray = array[Sort(array)]
  arrayenum = sortedarray[Uniq(sortedarray)]
  mappedarray = Value_Locate(arrayenum, array)
  hist = histogram(mappedarray, min=0)
  return, arrayenum[where(hist eq max(hist))]
end
```

Example:

```
print, mode([1., 10.^8, 10.^8])
 1.00000e+008
print, mode([10.^8, 10.^8+1, 1.])
 1.00000e+008
print, mode([10.^8, 10.^8+10, 1.])
      1.00000 1.00000e+008 1.00000e+008
```

So in this case the machine precision is about 7 significant digits, as expected for floats. Note that two floats are only assumed equal when they have the exact same binary value.