
Subject: Multidimensional Histograms re-visited
Posted by [jjb](#) on Sat, 05 Sep 1992 21:30:45 GMT
[View Forum Message](#) <> [Reply to Message](#)

In article <27AUG199209562770@stars.gsfc.nasa.gov> you write:
> In article <1992Aug27.075540@highwire.gsfc.nasa.gov>,
> burel@highwire.gsfc.nasa.gov (Jonathan Burelbach) writes...
>
>> Help! I am trying to create a scatterplot of 2 Landsat Image bands (ie
>> band 4 vs band 5) for clustering purposes and I need to be able to determine
>> the density of points on the plot. I have tried to create a 256x256 z array
>> by looping through the images, but this takes forever with images of any
>> size. Does anyone have any suggestions?
>
> This may do what you want. It uses HISTOGRAM to cut down on some of the
> looping.
>
> Bill Thompson
> -----
> PRO FORM_HISTO2,X,Y,HISTO,XSTEPS,YSTEPS,XDELTA,YDELTA

(code here left out)

If I understand the intent of the question and the response, I think there is a cleaner and simpler way to do this. I have imbedded this technique in several codes so I don't have a stand alone routine to offer, but what follows is an outline of the technique:

```
x = array of x values
y = array of y values
minx = minimum x value for x histogram axis
maxx = maximum x value for x histogram axis
miny = ditto for y etc.
maxy = ditto for y etc.
nx = number of histogram bins on the x axis
ny = number of histogram bins on the y axis
```

IDL code in outline form:

(Note: I usually deal with integer data for this problem so you might want to pay attention to rounding issues here more than I have....)

```
ix=long(nx*(x-minx)/(maxx-minx))
iy=long(ny*(y-miny)/(maxy-miny))
good=where((ix ge 0) and (ix lt nx) and (iy ge 0) and (iy lt ny))
```

```
(If good(0) eq -1 quit at this point....)
```

```
ixy=ix(good)+(iy(good)*nx)
```

```
hist=histogram(ixy,min=0l,max=nx*ny-1,binsize=1)
```

```
hist=reform(hist,nx,ny)
```

And we are done. Note that there is no looping at all!!!
Hope this helps.

Jeff Bloch
Space Astronomy and Astrophysics Group
Los Alamos National Laboratory

Subject: Re: Multidimensional Histograms re-visited
Posted by [thompson](#) on Sun, 06 Sep 1992 16:58:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

In article <1992Sep5.213045.14930@newshost.lanl.gov>, jjb@beta.lanl.gov (Jeffrey J Bloch) writes...

>

> In article <27AUG199209562770@stars.gsfc.nasa.gov> you write:

>> In article <1992Aug27.075540@highwire.gsfc.nasa.gov>,

>> burel@highwire.gsfc.nasa.gov (Jonathan Burelbach) writes...

>>

(message deleted)

No point in repeating everything, so I'll summarize. Jonathan Burelbach asked for an efficient technique to make 2D histograms from a set of X,Y pairs. I posted a routine that I used, and then Jeffrey J Bloch posted a technique which is much more efficient. I've tested it and it gives the same results as my routine, but much faster. I've now incorporated his technique in my own routine, and since I had already posted my less efficient routine earlier, I felt I should post the updated routine. I also corrected a small bug introduced when converting the routine from IDL version 1 to version 2.

Bill Thompson

```
-----  
PRO FORM_HISTO2,X,Y,HISTO,XSTEPS,YSTEPS,XDELTA,YDELTA  
;+  
; NAME:  
; FORM_HISTO2
```

```

; PURPOSE:
; Forms a two-dimensional histogram from a set of X,Y points.
; CALLING SEQUENCE:
; FORM_HISTO2, X, Y, HISTO, XSTEPS, YSTEPS [, XDELTA, YDELTA ]
; INPUT PARAMETERS:
; X, Y = Arrays giving the X,Y points to form the histogram from.
; OPTIONAL INPUT PARAMETERS:
; XDELTA = Spacing between histogram bins in the X direction.  If not
;   passed, then a suitable value is selected automatically.
; YDELTA = Spacing between histogram bins in the Y direction.
; OUTPUT PARAMETERS:
; HISTO = Two-dimensional array containing the histogram values.
; XSTEPS = Bin coordinate values along the X axis.
; YSTEPS = Bin coordinate values along the Y axis.
; OPTIONAL KEYWORD PARAMETERS:
; None.
; COMMON BLOCKS:
; None.
; SIDE EFFECTS:
; None.
; RESTRICTIONS:
; X and Y must have the same number of points.
; PROCEDURE:
; The number of points within bins bounded by XSTEP(I:I+1), YSTEP(I:I+1)
; are counted and stored into HISTO.
; MODIFICATION HISTORY:
; William Thompson, April 1992, incorporated into SERTS library.
; William Thompson, August 1992, greatly speeded up using suggestion by
;   Jeffrey J Bloch.
;-
;
; ON_ERROR,2
;
; Check the number of parameters.
;
IF (N_PARAMS(0) NE 5) AND (N_PARAMS(0) NE 7) THEN BEGIN
  PRINT, '*** FORM_HISTO2 must be called with 5 or 7 parameters:'
  PRINT, '      X, Y, HISTO, XSTEPS, YSTEPS [, XDELTA, YDELTA ]'
  RETURN
ENDIF
;
; Get the minimum and maximum values of X and Y.
;
BANG_C = !C
XMIN = MIN(X, MAX=XMAX)
YMIN = MIN(Y, MAX=YMAX)
!C = BANG_C
;

```

```

; If passed, then check the value of XDELTA.
;
IF N_PARAMS(0) EQ 7 THEN BEGIN
  IF N_ELEMENTS(XDELTA) NE 1 THEN BEGIN
    MESSAGE,'XDELTA must be scalar'
  END ELSE IF XDELTA LE 0 THEN BEGIN
    MESSAGE,'XDELTA must be positive'
  ENDIF
;
; If XDELTA was not passed, then determine the approximate number of histogram
; levels from the number of elements of X.
;
END ELSE BEGIN
  NX = FLOAT(N_ELEMENTS(X))
  NX = NX < 100. < (7.*ALOG10(NX) + NX/8.)
;
; Use NX to determine the spacing of the histogram levels. Break this number
; down into mantissa and exponent.
;
  XDELTA = (XMAX - XMIN) / (NX - 1)
  XPOWER = FIX(ALOG10(XDELTA))
  IF XPOWER GT ALOG10(XDELTA) THEN XPOWER = XPOWER - 1
  XDELTA = XDELTA / 10.^XPOWER
;
; Ensure that the spacing of the histogram levels is either 1,2 or 5 times
; some power of ten.
;
  XVAL = [10,5,2]
  XVALUE = 1
  FOR I = 0,2 DO IF XVAL(I) GT XDELTA THEN XVALUE = XVAL(I)
  XDELTA = XVALUE * 10.^XPOWER
;
; If X is of some integer type (byte, integer or long), then ensure that
; XDELTA is at least one.
;
  TYPE = SIZE(X)
  TYPE = TYPE(TYPE(0) + 1)
  IF ((TYPE EQ 1) OR (TYPE EQ 2) OR (TYPE EQ 3)) THEN $
    XDELTA = XDELTA > 1
ENDELSE
;
; Find the nearest multiple of XDELTA which is LE the minimum of X.
; Do the same for the maximum of X.
;
IXMIN = LONG(XMIN / XDELTA)
IXMAX = LONG(XMAX / XDELTA)
IF IXMIN*XDELTA GT XMIN THEN IXMIN = IXMIN - 1
IF IXMAX*XDELTA LE XMAX THEN IXMAX = IXMAX + 1

```

```

XMIN = IXMIN * XDELTA
XSTEPS = XMIN + XDELTA * INDGEN(IXMAX - IXMIN + 1)
;
; If passed, then check the value of YDELTA.
;
IF N_PARAMS(0) EQ 7 THEN BEGIN
    IF N_ELEMENTS(YDELTA) NE 1 THEN BEGIN
        MESSAGE,'YDELTA must be scalar'
    END ELSE IF YDELTA LE 0 THEN BEGIN
        MESSAGE,'YDELTA must be positive'
    ENDIF
;
; If YDELTA was not passed, then determine the approximate number of histogram
; levels from the number of elements of Y.
;
END ELSE BEGIN
    NY = FLOAT(N_ELEMENTS(Y))
    NY = NY < 100. < (7.*ALOG10(NY) + NY/8.)
;
; Use NY to determine the spacing of the histogram levels. Break this number
; down into mantissa and exponent.
;
    YDELTA = (YMAX - YMIN) / (NY - 1)
    YPOWER = FIX(ALOG10(YDELTA))
    IF YPOWER GT ALOG10(YDELTA) THEN YPOWER = YPOWER - 1
    YDELTA = YDELTA / 10.^YPOWER
;
; Ensure that the spacing of the histogram levels is either 1,2 or 5 times
; some power of ten.
;
    YVAL = [10,5,2]
    YVALUE = 1
    FOR I = 0,2 DO IF YVAL(I) GT YDELTA THEN YVALUE = YVAL(I)
    YDELTA = YVALUE * 10.^YPOWER
;
; If Y is of some integer type (byte, integer or long), then ensure that
; YDELTA is at least one.
;
    TYPE = SIZE(Y)
    TYPE = TYPE(TYPE(0) + 1)
    IF ((TYPE EQ 1) OR (TYPE EQ 2) OR (TYPE EQ 3)) THEN $
        YDELTA = YDELTA > 1
ENDELSE
;
; Find the nearest multiple of YDELTA which is LE the minimum of Y.
; Do the same for the maximum of Y.
;
IYMIN = LONG(YMIN / YDELTA)

```

```
IYMAX = LONG(YMAX / YDELTA)
IF IYMIN*YDELTA GT YMIN THEN IYMIN = IYMIN - 1
IF IYMAX*YDELTA LE YMAX THEN IYMAX = IYMAX + 1
YMIN = IYMIN * YDELTA
YSTEPS = YMIN + YDELTA * INDGEN(IYMAX - IYMIN + 1)
;
; Form the histogram.
;
IX = LONG((X - XMIN) / XDELTA) & NX = N_ELEMENTS(XSTEPS)
IY = LONG((Y - YMIN) / YDELTA) & NY = N_ELEMENTS(YSTEPS)
IXY = IX + IY*NX
HISTO = HISTOGRAM(IXY,MIN=0L,MAX=NX*NY-1,BINSIZE=1)
HISTO = REFORM(HISTO,NX,NY)
;
RETURN
END
```
