Subject: Q: Quantil calculation in IDL?
Posted by joerg.mosthaf on Fri, 15 Oct 1999 07:00:00 GMT
View Forum Message <> Reply to Message

Hi,
I have been searching the help files and David Fannings great book, but I can't find a way to calculate 25%- and 75%-quantils. Unfortunately I don't know the english name for this so let me explain: A 75%-quantil is like the median, but with 75% instead of 50% i.e. the number in a data spread, that 75% of all data points are less or equal to. Is there a way to do this fast on an 256x256 array? I need it to cut off noise at a specific level and to get a reliable min/max value, not including data spikes. I am probably overlooking something very easy, but I just couldn't find it.

Thanks,
Joerg Mosthaf

Subject: Re: Q: Quantil calculation in IDL?
Posted by J.D. Smith on Fri, 15 Oct 1999 07:00:00 GMT
View Forum Message <> Reply to Message

Pavel Romashkin wrote:

> Am I missing something? Is 75%-quantils not just a value of the sorted array 3/4
> of the way through?
>
> ```
> IDL> a = bindgen(256,256)
> IDL> b = sort(a)
> IDL> c = n_elements(b)*3/4
> IDL> print, (a[b])[c]
>  192
> ```

A minor quibble, but a[b[c]] is faster, since it doesn't recopy the entire 256x256 array as (a[b])[c] does.

JD

```
--
 J.D. Smith                             |*|      WORK: (607) 255-5842
 Cornell University Dept. of Astronomy  |*|            (607) 255-6263
 304 Space Sciences Bldg.               |*|       FAX: (607) 255-5875
 Ithaca, NY 14853                       |*|
```

Subject: Re: Q: Quantil calculation in IDL?
Posted by J.D. Smith on Fri, 15 Oct 1999 07:00:00 GMT

View Forum Message <> Reply to Message

Joerg Mosthaf wrote:

>
> Hi,
> I have been searching the help files and David Fannings great book, but I can't
> find a way to calculate 25%- and 75%-quantils. Unfortunately I don't know the
> english name for this so let me explain: A 75%-quantil is like the median, but
> with 75% instead of 50% i.e. the number in a data spread, that 75% of all
> data points are less or equal to. Is there a way to do this fast on an
> 256x256 array? I need it to cut off noise at a specific level and to get a
> reliable min/max value, not including data spikes. I am probably overlooking
> something very easy, but I just couldn't find it.
>
> Thanks,
> Joerg Mosthaf

You can of course fully sort the data and then select element p(N-1) of the sorted array where p is your quantile percentage and N is the number of elements in your array.

However, if you're really looking for speed, you may need to consider a *selection* routine (not provided with IDL). The advantage of selection over sorting is that you don't care if everything is in order, just that everything less than the nth largest is to the left of it, everything greater is to the right, in any order. You gain as the log(N) over an optimized full sort. Numerical Recipes provides a nice selection routine, but it probably would be slower than sorting if translated to IDL. Compiling in C or Fortran and linking into IDL would provide the speed-up.

JD

--
J.D. Smith |*| WORK: (607) 255-5842
Cornell University Dept. of Astronomy |*| (607) 255-6263
304 Space Sciences Bldg. |*| FAX: (607) 255-5875
Ithaca, NY 14853 |*|

Subject: Re: Q: Quantil calculation in IDL?
Posted by Pavel Romashkin on Fri, 15 Oct 1999 07:00:00 GMT
View Forum Message <> Reply to Message

Am I missing something? Is 75%-quantils not just a value of the sorted array 3/4 of the way through?

```
IDL> a = bindgen(256,256)
IDL> b = sort(a)
IDL> c = n_elements(b)*3/4
IDL> print, (a[b])[c]
 192
```

So, 3/4 of all elements in A are less than 192. Sure enough, in this example a
median is as easy to get by replacing 3/4 with 1/2, and it is equal to 128.

Cheers,
Pavel

Joerg Mosthaf wrote:

> Hi,
> I have been searching the help files and David Fannings great book, but I can't
> find a way to calculate 25%- and 75%-quantils. Unfortunately I don't know the
> english name for this so let me explain: A 75%-quantil is like the median, but
> with 75% instead of 50% i.e. the number in a data spread, that 75% of all
> data points are less or equal to. Is there a way to do this fast on an
> 256x256 array? I need it to cut off noise at a specific level and to get a
> reliable min/max value, not including data spikes. I am probably overlooking
> something very easy, but I just couldn't find it.
>
> Thanks,
> Joerg Mosthaf

Subject: Re: Q: Quantil calculation in IDL?
Posted by davidf on Fri, 15 Oct 1999 07:00:00 GMT

View Forum Message <> Reply to Message

Joerg Mosthaf (joerg.mosthaf@urz.uni-heidelberg.de) writes:

> I have been searching the help files and David Fannings great book, but I can't
> find a way to calculate 25%- and 75%-quantils. Unfortunately I don't know the
> english name for this so let me explain: A 75%-quantil is like the median, but
> with 75% instead of 50% i.e. the number in a data spread, that 75% of all
> data points are less or equal to. Is there a way to do this fast on an
> 256x256 array? I need it to cut off noise at a specific level and to get a
> reliable min/max value, not including data spikes. I am probably overlooking
> something very easy, but I just couldn't find it.

Well, I don't know the "IDL way" to do this (Stein
Vidar or someone else can fill us in on the details),
but the "get the thing done" way to do this goes something
like this:

1. Calculate the HISTOGRAM of the pixel distribution.
2. Do a "running sum" of the number of pixels in each
   bin of the histogram until you exceed your target
   amount. The value of that histogram bin will be
   your quantil.

Sorry I can't be more explicit, but aside from not wanting
to embarrass myself, I have a plane to catch in about a half
hour. :-)

See you guys (and Rose, good to hear from *you* again!) in
a couple of weeks.

Cheers,

David
--
David Fanning, Ph.D.
Fanning Software Consulting
Phone: 970-221-0438 E-Mail: davidf@dfanning.com
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Toll-Free IDL Book Orders: 1-888-461-0155

Subject: Re: Q: Quantil calculation in IDL?
Posted by James Tappin on Mon, 18 Oct 1999 07:00:00 GMT

View Forum Message <> Reply to Message

Joerg Mosthaf wrote:
> Hi,
> I have been searching the help files and David Fannings great book, but I can't
> find a way to calculate 25%- and 75%-quantils. Unfortunately I don't know the
> english name for this so let me explain: A 75%-quantil is like the median, but
> with 75% instead of 50% i.e. the number in a data spread, that 75% of all
> data points are less or equal to. Is there a way to do this fast on an
> 256x256 array? I need it to cut off noise at a specific level and to get a
> reliable min/max value, not including data spikes. I am probably overlooking
> something very easy, but I just couldn't find it.

It's not wondrously efficient. But here is a routine that I wrote that will
find arbitrary fractiles of an array (N.B. it takes fractions rather than
percentages). It could be improved by doing a floor and a ceil and
interpolating rather than just a round.

```
CUT HERE -- fractile .pro
 function Fractile, x, frac

;+
; FRACTILE
; Return the requested fractile of the input data.
;
; Usage:
; fr = fractile(x, frac)
;
; Return:
; fr <input> The requested fractile.
;
; Arguments:
; x most input The array whose fractile(s) are to be
;    returned
; frac float input The fractile(s) to return.
;
; Restrictions:
; The input data must be a SORTable array (i.e. not complex,
; string or structure).
;
; Example:
; To find the interquartile range of a data set, try:
; q = fractile(data, [.25,.75])
; iqr = q(1)-q(0)
;
; History:
; Original: 26/9/95; SJT
;-

if (n_params() ne 2) then message, 'Incorrect number of arguments'

n = n_elements(x)
i = sort(x)

f = round(frac*n)

return, x(i(f))

end
```

Subject: Re: Q: Quantil calculation in IDL?
Posted by joerg.mosthaf on Mon, 18 Oct 1999 07:00:00 GMT

View Forum Message <> Reply to Message

J.D. Smith <jdsmith@astro.cornell.edu> wrote:
: You can of course fully sort the data and then select element p(N-1) of
: the sorted array where p is your quantile percentage and N is the number
: of elements in your array.

: However, if you're really looking for speed, you may need to consider a
: *selection* routine (not provided with IDL). The advantage of selection
: over sorting is that you don't care if everything is in order, just that
: everything less than the nth largest is to the left of it, everything
: greater is to the right, in any order. You gain as the log(N) over an
: optimized full sort. Numerical Recipes provides a nice selection
: routine, but it probably would be slower than sorting if translated to
: IDL. Compiling in C or Fortran and linking into IDL would provide the
: speed-up.

Well, I will first try the suggestion from Mr. Romashkin, and if it is lacking
in speed, then I'll look into a selection routine in C.

Thank you for the answers,

Joerg Mosthaf

Subject: Re: Q: Quantil calculation in IDL?
Posted by m218003 on Mon, 25 Oct 1999 07:00:00 GMT

View Forum Message <> Reply to Message

In article <7uensq$8vu$1@usenet.bham.ac.uk>,
 James Tappin <sjt@star.sr.bham.ac.uk> writes:
> Joerg Mosthaf wrote:
>> Hi,
>> I have been searching the help files and David Fannings great book,
>> but I can't
>> find a way to calculate 25%- and 75%-quantils.
>> ...

Please find attached my PERCENTILES function which uses the SORT method
and allows for calculating arbitrary percentiles.

Cheers,
Martin.

[[[[[[[[[[
[[Dr. Martin Schultz   Max-Planck-Institut fuer Meteorologie   [[
[[Bundesstr. 55, 20146 Hamburg   [[
[[phone: +49 40 41173-308 or 416   [[
[[fax: +49 40 41173-298   [[
[[martin.schultz@dkrz.de   [[
[[[[[[[[[[

; $Id: percentiles.pro,v 1.11 1999/05/20 16:20:42 mgs Exp $

```
 ;-------------------------------------------------------- --
;+
; NAME:
;       PERCENTILES
;
; PURPOSE:
;       compute percentiles of a data array
;
; CATEGORY:
;       statistical function
;
; CALLING SEQUENCE:
;       Y = PERCENTILES(DATA [,VALUE=value-array])
;
; INPUTS:
;       DATA --> the vector containing the data
;
; KEYWORD PARAMETERS:
;       VALUE --> compute specified percentiles
;       default is a standard set of min, 25%, median (=50%), 75%, and max
;       which can be used for box- and whisker plots.
;       The values in the VALUE array must lie between 0. and 1. !
;
; OUTPUTS:
;       The function returns an array with the percentile values or
;       -1 if no data was passed or value contains invalid numbers.
;
; SUBROUTINES:
;
; REQUIREMENTS:
;
; NOTES:
;
; EXAMPLE:
;      x = (findgen(31)-15.)*0.2     ; create sample data
;      y = exp(-x^2)/3.14159         ; compute some Gauss distribution
;      p = percentiles(y,value=[0.05,0.1,0.9,0.95])
;      print,p
;
;      IDL prints :  3.92826e-05  0.000125309     0.305829     0.318310

;
; MODIFICATION HISTORY:
;       mgs, 03 Aug 1997: VERSION 1.00
;       mgs, 20 Feb 1998: - improved speed and memory usage
;                (after tip from Stein Vidar on newsgroup)
;
;-
```

```
; Copyright (C) 1997, Martin Schultz, Harvard University
; This software is provided as is without any warranty
; whatsoever. It may be freely used, copied or distributed
; for non-commercial purposes. This copyright notice must be
; kept with any copy of this software. If this software shall
; be used commercially or sold as part of a larger package,
; please contact the author to arrange payment.
; Bugs and comments should be directed to mgs@io.harvard.edu
; with subject "IDL routine percentiles"
 ;------------------------------------------------------------- --


function percentiles,data,value=value

result = -1
n = n_elements(data)
if (n le 0) then return,result   ; error : data not defined

; check if speficic percentiles requested - if not: set standard
if(not keyword_set(value)) then value = [ 0., 0.25, 0.5, 0.75, 1.0 ]

; create a temporary copy of the data and sort
; tmp = data
; tmp = tmp(sort(tmp))
; NO: simply save the sorted index array
  ix = sort(data)

; loop through percentile values, get indices and add to result
; This is all we need since computing percentiles is nothing more
; than counting in a sorted array.
for i=0L,n_elements(value)-1 do begin

   if(value(i) lt 0. OR value(i) gt 1.) then return,-1

   if(value(i) le 0.5) then ind = long(value(i)*n)    $
   else ind = long(value(i)*(n+1))
   if (ind ge n) then ind = n-1L    ; small fix for small n
                                  ; (or value eq 1.)

;  if(i eq 0) then result = tmp(ind)  $
;  else result = [result, tmp(ind) ]
; ## change number 2
   if(i eq 0) then result = data(ix(ind))  $
   else result = [result, data(ix(ind)) ]
endfor

return,result
end
```

File Attachments

1) percentiles.pro, downloaded 151 times