Subject: Re: Automatic Binsize Calculations
Posted by Foldy Lajos on Sun, 29 May 2011 18:24:35 GMT
View Forum Message <> Reply to Message

On Sun, 29 May 2011, David Fanning wrote:

> Gianguido Cianci writes:
>
>> Here's what I came up with, using sshist_2d.pro
>> (http://tinyurl.com/3on7bzx) that automagically finds bin size:
>
> I don't have a television, so while I listened to Djokovic
> defeat Gasquet on the French Open Radio I was fooling
> around using the 1D version of sshist to calculate
> a default bin size for cgHistoplot. What I discovered
> is that I get completely different results depending
> on the data type of the input data!
>
> I modified sshist a bit to get the bin size out of it
> as a keyword:
>
> ; Author: Shigenobu Hirose at JAMSTEC
> ; based on original paper
> ; Shimazaki and Shinomoto, Neural Computation 19, 1503-1527, 2007
> ; http://toyoizumilab.brain.riken.jp/hideaki/res/histogram.htm l
> ;
> function sshist, data, x=x, cost=cost, nbin=nbin, binsize=binsize
>
> COMPILE_OPT idl2
>
> nbin_min = 2
> nbin_max = 200
>
> ntrial = nbin_max - nbin_min + 1
>
> nbin = INDGEN(ntrial) + nbin_min
>
> delta = FLTARR(ntrial)
> cost = FLTARR(ntrial)
>
> for n = 0, ntrial-1 do begin
> delta[n] = (MAX(data) - MIN(data)) / (nbin[n] - 1)
>
> k = HISTOGRAM(data, nbins=nbin[n])
>
> kmean = MEAN(k)
> kvari = MEAN((k - kmean)^2)
> cost[n] = (2. * kmean - kvari) / delta[n]^2

>  endfor
>
>  n = (WHERE(cost eq MIN(cost)))[0]
>  k = HISTOGRAM(data, nbins=nbin[n], locations=x, reverse_indices=ri)
>
>  if arg_present(binsize) then binsize = delta[n]
>  return, k
>
> end
>
> But, look at this:
>
> IDL> void = sshist(cgdemodata(21), binsize=bs) & print, bs
>      9.00000
> IDL> void = sshist(fix(cgdemodata(21)), binsize=bs) & print, bs
>      1.00000
> IDL> void = sshist(long(cgdemodata(21)), binsize=bs) & print, bs
>      1.00000
> IDL> void = sshist(float(cgdemodata(21)), binsize=bs) & print, bs
>      1.33684
>
> I have NO idea why this is occurring. :-(
>
> Cheers,
>
> David

My result is worse:

```
IDL> print, !version
{ x86_64 linux unix linux 8.1 Mar  9 2011      64      64}
IDL> void = sshist(cgdemodata(21), binsize=bs) & print, bs
% Compiled module: CGDEMODATA.
% Loaded DLM: JPEG.
% Compiled module: SSHIST.
% Compiled module: MEAN.
% Compiled module: MOMENT.
% Array dimensions must be greater than 0.
% Execution halted at: SSHIST             26    .../sshist.pro
%                    $MAIN$
IDL>
```

Removing 'reverse_indices=ri' from histogram:

```
IDL> void = sshist(cgdemodata(21), binsize=bs) & print, bs
       21.0000
```

```
IDL> void = sshist(fix(cgdemodata(21)), binsize=bs) & print, bs
      1.00000
IDL> void = sshist(long(cgdemodata(21)), binsize=bs) & print, bs
      1.00000
IDL> void = sshist(float(cgdemodata(21)), binsize=bs) & print, bs
      1.33684
```

regards,
Lajos

---