Subject: Re: "bootstrap" statistics
Posted by Med Bennett on Sat, 25 May 2002 14:43:19 GMT
View Forum Message <> Reply to Message

wmc@bas.ac.uk wrote:

> Hello group. I want to do what I think of as "bootstrap" statistics, viz
> given a timeseries I take a random subsample (with, say, half the number
> of elements), compute some statistic (say, then mean); then take another
> random subsample; then again lots of times (say 1000 or 10000) and end
> up with a distribution of the statistic concerned.
>
> So: to do this I need a means to generate n/2 random (non-repeating)
> indices from 0...n-1. At the moment I do this by:
> once I have m indices I generate one more at
> random; see if its in the list of m; if not, good; if it is, generate another
> one. This is hideously inefficient and slow: there *must* be a better way.
>
> I have found:
>
> http://www.astro.washington.edu/deutsch-bin/getpro/library14 .html?PERMUTE
>
> which is a somewhat better way, but still slow. Is there a no-loops version?
>
> -W.
>
> --
> William M Connolley | wmc@bas.ac.uk | http://www.nerc-bas.ac.uk/icd/wmc/
> Climate Modeller, British Antarctic Survey | Disclaimer: I speak for myself
> I'm a .signature virus! copy me into your .signature file & help me spread!

The way to do this without using the same value more than once is to use the sort command. First, generate m uniform random numbers, then find the sort indices of the random numbers. For example, if you want to randomly pick 50 out of 100 values:

```
IDL> junk = randomu(seed,100)
IDL> s=sort(junk)
IDL> print,s[0:49]
       1      96      16      19      69      97
85      91      59
      54      77       2      95      52      22
88       0      39
      44      70       8      63      50      82
41      43      42
      57      68      98      20      46      26
60      94      12
      35      72      51      14      71      64
```

78 29 89
83 10 92 58 75

--

Remove all w's from my email address to reply!

ICBM Coordinates: 40.027666N, 105.289188W, 1670 m