Subject: Re: "bootstrap" statistics
Posted by Wayne Landsman on Mon, 27 May 2002 19:55:17 GMT
View Forum Message <> Reply to Message

wmc@bas.ac.uk wrote:

> Med Bennett <mwbennett@windra.com> wrote:
>> 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]
>
> Ah. Yes. Clever. (I have a feeling thats what one other post was trying to tell
> me but wasn't so clear).

Though the above method gives distinct values, they may not be completely random if the original RANDOMU() call returns duplicate values. The behaviour of the SORT() command when encountering duplicate values differs on different machine architectures, but on many machines the original array order is maintained (e.g. SORT([4,3,4,4,2]) returns [4,1,0,2,3]. This introduces a bias, where lower indicies are more likely to occur at the beginning of the permutation vector, and thus may bias a result that depends on completely random permuations.

However, the likelihood of RANDOMU returning duplicate values seems quite small for reasonable (<10,000?) sizes, and will be even smaller if the /DOUBLE keyword is used.

Cheers , --Wayne Landsman