Subject: Random Sampling Without Replacement
Posted by David Fanning on Wed, 13 Oct 2010 15:46:09 GMT
View Forum Message <> Reply to Message

Folks,

Has anyone coded up an IDL algorithm to do random
sampling without replacement?

For example, suppose I want to sample values in
my 2D image. I want, say, 100 values that represent
individual pixel locations in the image. How can
I make sure I get 100 unique, but random, locations?

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: Random Sampling Without Replacement
Posted by Heinz Stege on Thu, 14 Oct 2010 01:14:00 GMT
View Forum Message <> Reply to Message

On Wed, 13 Oct 2010 09:46:09 -0600, David Fanning wrote:

> Folks,
>
> Has anyone coded up an IDL algorithm to do random
> sampling without replacement?
>
> For example, suppose I want to sample values in
> my 2D image. I want, say, 100 values that represent
> individual pixel locations in the image. How can
> I make sure I get 100 unique, but random, locations?
>
> Cheers,
>
> David

Hi all,

here is another way to do this calculation:

```
function unique_random,n,m
;
; n := total number of values
; m := number of samples
;
compile_opt defint32,strictarr,strictarrsubs
;
inds=long(randomu(seed,m)*(n-findgen(m)))
;
table=lindgen(n)
for i=0,m-1 do begin
  j=inds[i]
  inds[i]=table[j]
  table[j]=table[n-1-i]
end
;
return,inds
end
```

For a small number of samples (n=100000, m<50000) it is faster than Mike's code. And if the number of samples is not very small (n=100000, m>10000), it is even faster than JD's solution from http://tinyurl.com/26edmmq.

This is true in spite of the presence of the for-loop. I'm surprised myself. This algorithm may be a good over-all-solution for IDL.

Heinz

Subject: Re: Random Sampling Without Replacement
Posted by Heinz Stege on Thu, 14 Oct 2010 02:59:23 GMT
View Forum Message <> Reply to Message

On Thu, 14 Oct 2010 03:14:00 +0200, Heinz Stege wrote:

> function unique_random,n,m

Of course we should add the seed to the parameter list:

function unique_random,n,m,seed

Heinz

Subject: Re: Random Sampling Without Replacement
Posted by David Fanning on Thu, 14 Oct 2010 03:19:16 GMT
View Forum Message <> Reply to Message

Heinz Stege writes:

> Of course we should add the seed to the parameter list:
>
> function unique_random,n,m,seed

Actually, you should probably put the seed in a common
block, or an awful lot of your "sampling" sequences
are going to look a hell of a lot alike. :-)

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: Random Sampling Without Replacement
Posted by Heinz Stege on Thu, 14 Oct 2010 10:44:28 GMT
View Forum Message <> Reply to Message

On Wed, 13 Oct 2010 21:19:16 -0600, David Fanning wrote:

> Heinz Stege writes:
>
>> Of course we should add the seed to the parameter list:
>>
>> function unique_random,n,m,seed
>
> Actually, you should probably put the seed in a common
> block, or an awful lot of your "sampling" sequences
> are going to look a hell of a lot alike. :-)
>
> Cheers,
>
> David

Okay, I see. What I wanted to say is, that one has to take care of the
seed. And it is my preference, to put it into the parameter list.

I am afraid, that I would forget about the today's common block, and
generate a second one within another routine in half a year. :-)

Greetings, Heinz

Subject: Re: Random Sampling Without Replacement
Posted by David Fanning on Thu, 14 Oct 2010 12:27:46 GMT
View Forum Message <> Reply to Message

Heinz Stege writes:

> Okay, I see. What I wanted to say is, that one has to take care of the
> seed. And it is my preference, to put it into the parameter list.
>
> I am afraid, that I would forget about the today's common block, and
> generate a second one within another routine in half a year. :-)

Yes, half a year later you would probably be fine. However,
if you were doing this in some like of loop, maybe using
a bootstrap process or something, passing in the seed
as a parameter is often problematic. To get a truly
random sequence of numbers, the seed has remain "alive"
between calls to RandomU. Otherwise, you get the same
"random" sequence of numbers coming out of your program.

I think a lot of people don't realize this.

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")
