Subject: RANDOMN function
Posted by fd_luni on Thu, 14 Nov 2013 13:59:11 GMT
View Forum Message <> Reply to Message

Hi

I used the RANDOMN function to add Gaussian noise to my data like this:

noise=RANDOMN(seed,N)

When I print, the mean values of RANDOMN(seed,N) I didn't get zero but something around 0.0337187.

I expect to get something very very close to zero since the RANDOMN function returns normally-distributed, floating-points with a mean of zero. Is my assumption wrong? It's correct that the median is not zero?

Many Thanks
Mar

Subject: Re: RANDOMN function
Posted by on Thu, 14 Nov 2013 14:46:01 GMT
View Forum Message <> Reply to Message

Den torsdagen den 14:e november 2013 kl. 14:59:11 UTC+1 skrev fd_...@mail.com:
> Hi
>
> I used the RANDOMN function to add Gaussian noise to my data like this:
>
> noise=RANDOMN(seed,N)
>
> When I print, the mean values of RANDOMN(seed,N) I didn't get zero but something around 0.0337187.
>
> I expect to get something very very close to zero since the RANDOMN function returns normally-distributed, floating-points with a mean of zero. Is my assumption wrong? It's correct that the median is not zero?

You will get better statistics the larger the set:

```
IDL> for e=1,9 do print,e,10d^e,mean(randomn(seed,10d^e))
      1       10.000000    0.0489258
      2       100.00000    0.0172675
      3       1000.0000   -0.0336368
      4       10000.000  -0.00799687
      5       100000.00   0.00208867
      6       1000000.0  -0.00101986
```

7 10000000. -0.000105310
8 1.0000000e+08 0.000104068
9 1.0000000e+09 2.46072e-05

Subject: Re: RANDOMN function
Posted by Russell Ryan on Thu, 14 Nov 2013 15:12:55 GMT
View Forum Message <> Reply to Message

There are lies, damned lies, and statistics.

On Thursday, November 14, 2013 9:46:01 AM UTC-5, Mats Löfdahl wrote:
> Den torsdagen den 14:e november 2013 kl. 14:59:11 UTC+1 skrev fd_...@mail.com:
>
>> Hi
>
>>
>
>> I used the RANDOMN function to add Gaussian noise to my data like this:
>
>>
>
>> noise=RANDOMN(seed,N)
>
>>
>
>> When I print, the mean values of RANDOMN(seed,N) I didn't get zero but something around 0.0337187.
>
>>
>
>> I expect to get something very very close to zero since the RANDOMN function returns normally-distributed, floating-points with a mean of zero. Is my assumption wrong? It's correct that the median is not zero?
>
>
>
> You will get better statistics the larger the set:
>
>
>
> IDL> for e=1,9 do print,e,10d^e,mean(randomn(seed,10d^e))
>
> 1 10.000000 0.0489258
>

> 2 100.00000 0.0172675
>
> 3 1000.0000 -0.0336368
>
> 4 10000.000 -0.00799687
>
> 5 100000.00 0.00208867
>
> 6 1000000.0 -0.00101986
>
> 7 10000000. -0.000105310
>
> 8 1.0000000e+08 0.000104068
>
> 9 1.0000000e+09 2.46072e-05

Subject: Re: RANDOMN function
Posted by fd_luni on Thu, 14 Nov 2013 15:15:54 GMT
View Forum Message <> Reply to Message

> You will get better statistics the larger the set:
>
>
>
> IDL> for e=1,9 do print,e,10d^e,mean(randomn(seed,10d^e))
>
> 1 10.000000 0.0489258
>
> 2 100.00000 0.0172675
>
> 3 1000.0000 -0.0336368
>
> 4 10000.000 -0.00799687
>
> 5 100000.00 0.00208867
>
> 6 1000000.0 -0.00101986
>
> 7 10000000. -0.000105310
>
> 8 1.0000000e+08 0.000104068
>
> 9 1.0000000e+09 2.46072e-05

Well, I don't see what you mean to be honest.

Subject: Re: RANDOMN function
Posted by fd_luni on Thu, 14 Nov 2013 15:25:08 GMT
View Forum Message <> Reply to Message

On Thursday, 14 November 2013 15:12:55 UTC, rr...@stsci.edu wrote:
> There are lies, damned lies, and statistics.

What do you mean by saying that there are lies, damned lies and statistics?
The fact that the mean of the values generated by RANDOMN(seed,N) is not zero it's not wrong?

Subject: Re: RANDOMN function
Posted by on Thu, 14 Nov 2013 15:38:47 GMT
View Forum Message <> Reply to Message

Den torsdagen den 14:e november 2013 kl. 16:15:54 UTC+1 skrev fd_...@mail.com:
>> You will get better statistics the larger the set:
>
>> IDL> for e=1,9 do print,e,10d^e,mean(randomn(seed,10d^e))
>
>> 1 10.000000 0.0489258
>> 2 100.00000 0.0172675
>> 3 1000.0000 -0.0336368
>> 4 10000.000 -0.00799687
>> 5 100000.00 0.00208867
>> 6 1000000.0 -0.00101986
>> 7 10000000. -0.000105310
>> 8 1.0000000e+08 0.000104068
>> 9 1.0000000e+09 2.46072e-05
>
> Well, I don't see what you mean to be honest.

Just that a larger set is more likely to reproduce the statistics of the distribution that was used to generate it. How many times do you have to roll a dice to get 1/6 of each number to some specified accuracy?

Subject: Re: RANDOMN function
Posted by Craig Markwardt on Thu, 14 Nov 2013 15:47:29 GMT
View Forum Message <> Reply to Message

On Thursday, November 14, 2013 10:15:54 AM UTC-5, fd_...@mail.com wrote:
>> You will get better statistics the larger the set:
...
>
> Well, I don't see what you mean to be honest.

According to probability theory, the *expectation* mean value is zero.

However, you are dealing with statistical theory as well, not just probability, because you are sampling a random variable. Do you expect flipping 1000 coins to always produce *exactly* 500 heads and 500 tails? No. There will be some variation.

The average of N random variables is also a random variable. You will get a different answer if you take an average of new random samples.

According to statistics theory, the *sample* mean value will tend to zero, but only as the number of samples becomes very large. The expected standard deviation of the mean value of N gaussians with unit variance is 1.0/SQRT(N).

Based on the value you reported, I'm going to guess you were averaging about 1000 gaussians, right?

Craig

Subject: Re: RANDOMN function
Posted by fd_luni on Thu, 14 Nov 2013 16:50:58 GMT
View Forum Message <> Reply to Message

Oh I see now. Yes right, I averaged 1000 gaussians.

Subject: Re: RANDOMN function
Posted by Russell Ryan on Thu, 14 Nov 2013 18:19:27 GMT
View Forum Message <> Reply to Message

On Thursday, November 14, 2013 10:25:08 AM UTC-5, fd_...@mail.com wrote:
> On Thursday, 14 November 2013 15:12:55 UTC, rr...@stsci.edu wrote:
>
>> There are lies, damned lies, and statistics.
>
>
>
> What do you mean by saying that there are lies, damned lies and statistics?
>
> The fact that the mean of the values generated by RANDOMN(seed,N) is not zero it's not wrong?

It's a saying from Benjamin Disraeli (at least credited to him by wikipedia). I think his original meaning was that statistics can be used (or misused) to present any argument you want, and is now often just a joke about statistics. Your problem was really rooted in the central limit theorem:

http://en.wikipedia.org/wiki/Central_limit_theorem

Ultimately, the accuracy of any statistical quantity is increased by a larger sample. You were simply quoting an average of a bunch of random numbers and were worried that it wasn't zero, as you correctly expect. But, even if you do a billion numbers the average won't be exactly zero. It will only be zero within some range, which you would probably concede is fine if you also compute the variance on the average. In which case, you'll see (I'm willing to bet) that your 1000 random numbers were actually mean zero, within the variance.
