

## Re: issues with statistical test suite from http://csrc.nist.gov/rng/

**Source:** <http://www.derkeiler.com/Newsgroups/sci.crypt/2004-01/1618.html>

---

**From:** Mack ([macckone\\_at\\_a\\_nospamjunk123\\_ol.com](mailto:macckone_at_a_nospamjunk123_ol.com))

**Date:** 01/24/04

Date: Sat, 24 Jan 2004 11:16:33 GMT

On Thu, 22 Jan 2004 19:30:50 GMT, "Cristiano"  
<[cristiano.pi@NSquipo.it](mailto:cristiano.pi@NSquipo.it)> wrote:

>Mack wrote:

>> On Wed, 21 Jan 2004 21:44:53 GMT, "Cristiano"

>> <[cristiano.pi@NSquipo.it](mailto:cristiano.pi@NSquipo.it)> wrote:

>>

>>> Mack wrote:

>>>> On Tue, 20 Jan 2004 08:59:02 GMT, "Cristiano"

>>>> <[cristiano.pi@NSquipo.it](mailto:cristiano.pi@NSquipo.it)> wrote:

>>>>

>>>>> Mack wrote:

>>>>>> On Mon, 19 Jan 2004 23:18:19 GMT, "Cristiano"

>>>>>> <[cristiano.pi@NSquipo.it](mailto:cristiano.pi@NSquipo.it)> wrote:

>>>>>>

>>>>>>> Mack wrote:

>>>>>>>> On Mon, 19 Jan 2004 20:28:40 GMT, "Cristiano"

>>>>>>>> <[cristiano.pi@NSquipo.it](mailto:cristiano.pi@NSquipo.it)> wrote:

>>>>>>>>

>>>>>>>>> Mack wrote:

>>>>>>>>>> On Sat, 17 Jan 2004 17:09:26 GMT, "Cristiano"

>>>>>>>>>> <[cristiano.pi@NSquipo.it](mailto:cristiano.pi@NSquipo.it)> wrote:

>>>>>>>>>>

>>>>>>>>>>> Luke Kenneth Casson Leighton wrote:

>>>>>>>>>>>>

>>>>>>>>>>>>> the people at [csrc.nist.gov](http://csrc.nist.gov) inform me that they used

>>>>>>>>>>>>>> blum-blum-shub as the "baseline" for the lempel-ziv test (i

>>>>>>>>>>>>>> haven't asked them about the other tests) and that they

>>>>>>>>>>>>>>> then took EMPIRICALLY OBSERVED values for the mean and

>>>>>>>>>>>>>>> standard deviation of the information that generates the

>>>>>>>>>>>>>>> p-values.

>>>>>>>>>>>>>>>

>>>>>>>>>>>>>>>> They have also used sha-1 based generator to get the mean and

>>>>>>>>>>>>>>>>> the variance. They updated those values with:

>>>>>>>>>>>>>>>>> mean = 69588.20190000

>>>>>>>>>>>>>>>>> variance = 73.23726011

>>>>>>>>> *which are good enough.*  
>>>>>>>>>  
>>>>>>>>> *if they did the same on one or two other tests, it's*  
>>>>>>>>> *possible that they either didn't take a large enough*  
>>>>>>>>> *pseudo-random sample from which to derive the empirical*  
>>>>>>>>> *mean and s.d., or that there is a problem with the*  
>>>>>>>>> *pseudo-random generator that they used.*  
>>>>>>>>>  
>>>>>>>>> *either way, a skew of the p-values is, as you say,*  
>>>>>>>>> *introduced.*  
>>>>>>>>>  
>>>>>>>>> *No skewed p-values introduced.*  
>>>>>>>>>  
>>>>>>>>> *This is easily testable. The p-values are not uniformly*  
>>>>>>>>> *distributed. There is skew. I have posted examples in a*  
>>>>>>>>> *seperate message.*  
>>>>>>>>>  
>>>>>>>>> *I seen the message. We can try with an example.*  
>>>>>>>>> *Suppose you have  $n=1e6$ .*  
>>>>>>>>> *Calculate the p-values for few W's:*  
>>>>>>>>> *W p-value*  
>>>>>>>>> *69588 0,49058890723625*  
>>>>>>>>> *69589 0,537151142026133*  
>>>>>>>>> *69590 0,58320929236385*  
>>>>>>>>> *69591 0,628151659956359*  
>>>>>>>>> *69592 0,67141122813222*  
>>>>>>>>>  
>>>>>>>>> *you can clearly see that the p-value 0.51 (for example) cannot*  
>>>>>>>>> *exists because  $W = 69588.361$  cannot exists ( $W$  is an integer*  
>>>>>>>>> *number).*  
>>>>>>>>>  
>>>>>>>>> *The p-values are \*not\* skewed, they don't exist. You just need*  
>>>>>>>>> *to properly use the test.*  
>>>>>>>>>  
>>>>>>>>> *If they don't exist then there is skew in the output.*  
>>>>>>>>>  
>>>>>>>>> *No.*  
>>>>>>>>> *You can see that they are \*not\* skewed by calculating the*  
>>>>>>>>> *skewness:*  
>>>>>>>>> *it is about 0 (very good).*  
>>>>>>>>> *An other way is to see the graphical display of the sorted*  
>>>>>>>>> *p-values: you'll see that they are about evenly distributed (the*  
>>>>>>>>> *only problem*  
>>>>>>>>> *is that they "jump" over the bins).*  
>>>>>>>>>  
>>>>>>>>> *A skewed distribution is, for example, a bell shaped curve which*  
>>>>>>>>> *looks like a chi-squared one with  $df>3$ ; in other words the lack*  
>>>>>>>>> *of p-values is in a tail.*  
>>>>>>>>>  
>>>>>>>>>  
>>>>>>>>> *Perhaps a better description is bias. However you define it the*

>>>>> *distribution is not even.*  
>>>>>  
>>>>> *Yes, but not skewed as you and the troll insist to say.*  
>>>>>  
>>>>>  
>>>>> *The p-values of the FFT were skewed. Recheck my post with*  
>>>>> *the data. On the 1e6 x 100 test of Lempel-Ziv the data were*  
>>>>> *also skewed.*  
>>>>>  
>>>>> *In which post? There is no post with: "Skewness= ...".*  
>>>>> *I checked the LZ test also for the skewness and I found no skewed*  
>>>>> *p-values. I don't know in which other way I can say that.*  
>>>>>  
>>>>>  
>> *post: m7hn005agur5djmlkmbmcqmaelha74u38m@4ax.com*  
>  
> *It is an e-mail! O\_o*

no, newsgroup post id.

>  
>  
>> *I didn't specifically include skewness values because the*  
>> *program doesn't automatically provide them. But from the*  
>> *resultant data the skewness is obvious in several instances.*  
>>  
>> *Working with the uniform binned data.*  
>> *The expected mean is 5.5.*  
>> *The SD is*  
>> *100 = 2.8868*  
>> *1000 = 2.8737*  
>>  
>> *In the 1e6x1000 LZ case it is least obvious.*  
>> *The total number of values on the left is 522.*  
>> *The Skewness is about -.0515. We could argue about*  
>> *exactly how significant (ses=.07746) this is but since it is*  
>> *consistent across multiple tests it is relevant. The slight*  
>> *skewness is only a side effect. The real problem was*  
>> *expecting the p-values to be uniform when they are not.*  
>>  
>> *In the 1e6x100 LZ only one value exceeds the expected mean*  
>> *on the right while three do on the left.*  
>> *Skewness= -.5689 (ses=.2449) obviously skewed.*  
>  
> *Just to give an example, in the statistic process control, usually a*  
> *distribution is said good if |skewness| <= .5 (there is also the kurtosis,*  
> *but in our conversation it is irrelevant).*  
> *This is to say that .57 is not so big. If you have seen that value only one*  
> *time, it is not a problem.*

skewness should be less than  $2 \cdot \text{ses}$  which will vary by sample size.  
 $\text{ses} = \sqrt{6/n}$ .

The problem is that it isn't an isolated incident although I believe it is the largest such value. No skewness values in the opposite direction were encountered at that sample size.

>  
>  
>> for 1e5x100 LZ skewness=.2103  
>> for 1e5x1000 LZ skewness=.0823  
>  
>Very good!  
>  
>  
>> The FFT are also obviously skewed for 1e5x1000 and 1e4x1000.  
>> 1e5x100=.4701 (ses=.2449)  
>> 1e5x1000=.5163 (ses=.07746)  
>> 1e4x1000=.4598 (ses=.07746)  
>>  
>> The rank test is also skewed very slightly (insignificantly?) for  
>> 1e4x1000.  
>> 1e4x1000=.0460  
>>  
>> The FFT is definitely skewed.  
>  
>I said several times that FFT test must be used around 1e6 bits; 1e5 bits is  
>not around 1e6 bits!  
>Try to check 1e6 or 2e6 bits, you should see a smaller skewness.  
>

The 1e6 values were acceptable. You stated that my original claim that I had found skew was false. I am simply showing you the data to back up that claim. In fact the original poster stated that this was only a problem below 1e6.

>  
>> The LZ doesn't seem symmetrical  
>> about the mean, which it isn't because of the way the bins  
>> are positioned relative to the mean.  
>  
>What does it mean?

Think of a mean of say 5.3 ... If only integers 1–10 can occur more numbers will fall in 5 than 6 as they should. But when the skewness is calculated there will be a slight skew. This is a little different from our setting where data are discrete decimal numbers and then dumped in bins but it illustrates the point.

>  
>

>>>> *It isn't readily evident from looking at coursey  
>>>> binned data but specifically in the Lempel-Ziv and FFT case  
>>>> the p-values were not "well behaved". The rank test behaves in  
>>>> the manner expected above a certain threshold. The data falling  
>>>> in to bins that are not centered on the median naturally creates  
>>>> skew.*  
>>>  
>>> *The eyes are not good to calculate the skewness and with the bins  
>>> there is loss of information (especially with the binning which  
>>> happens with LZ and FFT).*  
>>> *If you don't know how to calculate the skewness, I can send the  
>>> code, but, please, until that, stop your inconsistent claim.*  
>>  
>> *My claim is consistent with the data provided.*  
>> *Perhaps the small amount of data is erroneously showing  
>> a skew where none exists but some instances were badly  
>> skewed.*  
>  
>*If you are looking for a perfect random distribution you should use:  
>output = counter (mod MAX+1), this way the distribution will be perfect.  
>But if you use a random generator you can't expect a perfect distribution;  
>if this happens the generator would be bad.  
>With the prng, few "strange" results must happen.*  
>

In this case it is more of a strange result with the test since the BBS generator seems to be working fine. But these results could simply be randomness in action. I am not arguing that.

>  
>>> *Also the FFT p-values are not skewed (usually I get skewness=0.1,  
>>> 0.2).*  
>>>  
>>  
>> *Are you using the sample mean or expected mean? For the 1e5 FFT I  
>> never got a skewness below .4. For 1000 samples .2 would definitely  
>> be a significant skew (2\*ses=.15492).*  
>  
>*Sure, you use that test in a bad way; n must be around 1e6 bit, do you  
>remember?*  
>*Anyway, your question seems strange. You must use the sample mean, not the  
>expected one.*

That is incorrect when you are examining a sample presumed to be from a specific distribution. That would measure skew with respect to the sample itself, not with respect to the expected distribution.

1e4 x 1000

---

C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 P-VALUE PROPORTION  
STATISTICAL TEST

0 0 0 0 0 0 0 0 1000 0.000000 \* 1.0000

Lempel-Ziv

igamc underflow error occurs for Lempel-Ziv

As an example use the LZ test with 1e4 where all samples went to bucket ten since the case statement doesn't handle this case. If you use the sample mean then it has a skew of zero. ie. the mean is ten and all samples go to bucket ten. This is obviously not what we want the test to show. We are looking for a measure of how well this conforms to the expected distribution, in this case the mean should be 5.5 and the sample is badly skewed.

>

>

>>>> *I think you have already agreed that the KS test of the p-values for these tests is not correct. Specifically it isn't the correct test.*

>>>

>>> *I totally agree with your last sentence: the KS test **\*must\*** not be used with LZ.*

>>> *But all the tests in the suite are good to check a prng (if they are properly used).*

>>>

>>

>> *No argument here.*

>>

>> *The problem is that the test suite produces a "finalAnalysisReport" that indicates failures where there are none.*

>> *This is entirely because it uses an incorrect test when producing this report.*

>

>*That problem seems common for several tests (including DH). For this reason*

>*I take each single test and then I use them in a better way.*

Diehard has only given occasional bad result ie. isolated p-values, with good data. The major problem I have with diehard is that it isn't sensitive enough with processed data from physical random number generators. Diehard doesn't give KS results except where it is appropriate. Unfortunately the output is pretty hard to read. I usually open it with a text editor and search for results of .000, .00, and .0.

I am also having to create my own test suite because nothing else meets my current needs. sts seems like a good package but it has its limitations.

>

>*Cristiano*

>

sci.crypt: Re: issues with statistical test suite from <http://csrc.nist.gov/rng/>

Leslie 'Mack' McBride

remove text between \_ marks to respond via e-mail