

Re: FFT test with few kbits

Source: <http://www.derkeiler.com/Newsgroups/sci.crypt/2004-01/2194.html>

From: Bill Unruh (unruh_at_string.physics.ubc.ca)

Date: 01/31/04

Date: Sat, 31 Jan 2004 16:32:25 +0000 (UTC)

"Cristiano" <cristiano.pi@NSquipo.it> writes:

]Eric Backus wrote:

]> "Cristiano" <cristiano.pi@NSquipo.it> wrote in message

]> news:6xASb.169284\$VW.6887419@news3.tin.it...

]>> "Eric Backus" <eric_backus@alum.mit.edu> ha scritto nel messaggio

]>> news:1075493567.723814@cswreg.cos.agilent.com

]>>> "Cristiano" <cristiano.pi@NSquipo.it> wrote in message

]>>> news:e2xSb.168273\$VW.6834917@news3.tin.it...

]>>>> I think the easiest way to see what I'm doing is to transform this

]>>>> 20-bit sequence

]>>>> 01101110011110111110

]>>>>

]>>>> to see if you get (real and imaginary components):

]>>>> 14 0

]>>>> -0.554254269627734 1.08778525229247

]>>>> -1.19098300562505 -0.587785252292473

]>>>> -2.84785876296257 0.45105651629515

]>>>> 1.92705098312484 -2.1266270208801

]>>>> -2 -2

]>>>> -2.30901699437494 -0.951056516295152

]>>>> 0.229824774212685 -1.45105651629515

]>>>> -1.42705098312484 1.31432778029783

]>>>> 0.172288258377629 -0.0877852522924755

]>>>>

]>>>> Obviously I discard the second half (from $n/2$ to $n-1$) because it is

]>>>> identical to the first half.

]>>>

]>>> For what it's worth, I can verify that this really is the first half

]>>> of the FFT of the 20-bit sequence you gave.

]>>

]>> Thank you for the checking.

]>>

]>>> You'll notice that the very first component really is much bigger

]>>> than the rest, reflecting Ernst Lippe's observation that it should

]>>> average to $N/2$ while the rest should average to zero.

]>>

sci.crypt: Re: FFT test with few kbits

]>> Why? I agree that the first term should be about $N/2$ (because it is
]>> the numbers of 1's), but the mean of the rest (terms from 1 to $N/2$)
]>> is 0.5 or -0.5 for the real components.

]>

]> I don't see where you're getting that. For your example above, the
]> mean of the real part of the rest is -0.8888888888888888 . If you do this
]> many times with different random data, the expected value of the
]> values other than the first is zero. If you include the first value,
]> the expected value of everything is 1 (basically $N/2$ from the first
]> point, plus zero for each other point, divided by $N/2$ which is the
]> number of points you're averaging).

]>

]>

]>> The mean of the imaginary part varies from about -1.14 to 1.23 .
]>> You seem able to calculate the FFT; have you checked what you've
]>> said?

]>

]> Yes.

]I'm very surprised of that "yes".

]We have said that the first number is about $N/2$ (integer always >0).

]The sum of the real components from 1 to $N/2-1$ is always an integer number

]and it is about $1/2$ of the first component, or $N/4$ it can be <0 or >0 .

]For example, if we get a 300-bit sequence, we could get:

]component #1 = 148,

]sum from 1 to 149 of the real components = -78 ,

] $-78 / 149 = -.523$.

]I don't know how you get 0.

]>>> When you throw away the second half of the data, you can actually

]>>> keep the $n/2$ value, so you throw away only $(n/2)-1$ complex values.

]>>> The $n/2$ value will be real-only, just like the first value, but this

]>>> value is equally useful and testable and as independent as the rest

]>>> of the values.

]>>

]>> I'm not sure to understand exactly what you said.

]>

]> You start with 20 real values. The output of the DFT is 20 complex

]> values. The first of these (#0) is real-only. The following 9 (#1

]> through #9) are complex. The next value (#10) is real-only. The

]> last 9 (#11 through #19) are complex. The last 9 complex values (#11

]> through #19) are an image of (#1 through #9). So you can keep the

]> first eleven complex values out of the

]> 20. Of those eleven values, the first and last will be real-only.

]Agreed.

]>>> For the values that have non-zero imaginary parts, you should be
]>>> able to test that they have the same gaussian distribution as the
]>>> real part.
]>>
]>> Yes both components seem to be normally distributed. Here, a
]>> mathematical explanation would be very useful...
]>
]> I'm only going to hand-wave here, so it may not be too helpful. Each
]> output of the DFT is a weighted sum of N random inputs. Even though
]> the N random inputs are binary valued, something like the central
]> limit theorem says that the resulting sum will approach being
]> Gaussian distributed. For each complex output of the DFT, I believe
]> the distribution is Gaussian in Magnitude squared (sorry I'm having
]> trouble remembering the details now). I believe that the real and
]> imaginary parts are not independent of each other, but each looks
]> roughly Gaussian. For each complex output of the DFT, the phase
]> should be uniform between $+\pi$.

]I agree, that it what I seen in my experiments.
]Instead of "roughly Gaussian" I'd say "perfectly Gaussian". I seen
]incredibly beautiful bell-shaped curves; obviously I checked the shape also
]with the proper parameters and goodness-of-fit tests (KS and chi-square).

]>> As I said in my first post, the KS test over the real part gives good
]>> p-values, and the same happens for the imaginary part. These
]>> p-values should be uniformly distributed, but unfortunately it
]>> doesn't happen.
]>> The KS test over the p-values gotten from the 1st level KS test is
]>> bad for the real part, but incredibly bad for the imaginary part.
]>> For this reason I used the real part.
]>
]> If you're including the first point, which has a much larger real
]> part and exactly zero imaginary part, that might cause problems? I
]> guess I don't really know what the problem might be here.

]I don't include the first point, I use points from 1 to $N/2-1$.

a) the value of the i th number is n_i . Define n_i so the mean is zero (ie
subtract $1/2$ from your values. Then each n_i is either $+1/2$ or $-1/2$)
That also means that $\langle n_i^2 \rangle = 1/4$

b) The various n_i are supposed to be independent random variables
 $\langle n_i n_j \rangle = 0$ for $i \neq j$.

c) Define $C_k = \sum_i n_i \exp(i 2 \pi i k / N)$
where $i^2 = -1$, N is the total number.

Then

$\langle C_k C_l \rangle = 0$ if $k \neq -l$ and $\langle C_k C(-k) \rangle = \sum_i \langle n_i^2 \rangle = N/4$

Ie, the phases of various fourier components are uncorrelated.

sci.crypt: Re: FFT test with few kbits

The C_k are not gaussian distributed, although they get close for large N .