Subject: Re: CHISQR_CVF question.
Posted by pgrigis on Wed, 19 Aug 2009 19:29:13 GMT
View Forum Message <> Reply to Message

Hi Bob,

I guess I am a bit confused as to what you are doing.

Is that it? (I hope I am not making too much a fool of myself :))

- start with a uniform distribution x

- fourier transform: y= FFT(x)

- study histogram of real(y), imaginary(y) and abs(y),
  compare with known distributions (gaussian, chisquare)?

If that is it, maybe you are having a histogram binsize problem?

That's what happens to me very often - I forget to account for
histogram bin widths.

Ciao,
Paolo

On Aug 19, 12:12 pm, "R.G. Stockwell" <noemai...@please.com> wrote:
> I'm just writing up a simple routine to calculate
> significance levels for an FFT of a white spectrum.
> I actually find the distribution of the spectrum, normalize
> it with respect to N number of points and variation so I
> always get the same distribution, and I am comparing it
> to the CHISQR_CVF function (with 2 degrees of freedom, for a
> power spectrum).
>
> I am off by a constant factor (which depends on degrees of freedom)
> but invariant to length of the time series, or the standard deviation
> (since those are normalized out).
>
> any ideas of what step i am missing here? I frankly can't think
> of any other parameter involved here.
>
> cheers,
> bob
>
> PS with degrees of freedom = 2, the constant factor is
> 10.914899

Subject: Re: CHISQR_CVF question.
Posted by R.G. Stockwell on Wed, 19 Aug 2009 20:42:46 GMT
View Forum Message <> Reply to Message

"Paolo" <pgrigis@gmail.com> wrote in message
news:5c82e9fd-2bb3-4f1b-9ca6-8e0586a00bc0@g19g2000vbi.google groups.com...
Hi Bob,

I guess I am a bit confused as to what you are doing.

Is that it? (I hope I am not making too much a fool of myself :))

- start with a uniform distribution x

- fourier transform: y= FFT(x)

- study histogram of real(y), imaginary(y) and abs(y),
compare with known distributions (gaussian, chisquare)?

If that is it, maybe you are having a histogram binsize problem?

That's what happens to me very often - I forget to account for
histogram bin widths.

Ciao,
Paolo
****

basically yes, abs(fft(ts))^2, and comparing it to chisquare from the
IDL functions.
I have worked on it, but I think the result is off by a factor of 2.
That is a factor of 2 too stringent.

The binsizes are fine (i think), they are correctly showing the
distribution,
and the cumulative distribution and I normalized wrt to the number of points
so that the histogram is in term of probability.

Perhaps you can check my understanding. If we have a 95% significance
level,
then if we make a spectrum with 1000 points, shouldnt 50 of them be above
that 95% line?

cheers,
bob

Subject: Re: CHISQR_CVF question.

On Aug 19, 4:42 pm, "R.G. Stockwell" <noemai...@please.com> wrote:
> "Paolo" <pgri...@gmail.com> wrote in message
>
> basically yes, abs(fft(ts))^2, and comparing it to chisquare from the
> IDL functions.
> I have worked on it, but I think the result is off by a factor of 2.
> That is a factor of 2 too stringent.
>
...
> Perhaps you can check my understanding. If we have a 95% significance
> level,
> then if we make a spectrum with 1000 points, shouldnt 50 of them be above
> that 95% line?

Let's say we have a time series, defined like this,
LC = time series values (array)
ERR = measurement uncertainty (array) of each LC point.

I define the power spectrum in the following way,
POW = ABS(FFT(LC,+1))^2 * (2 / TOTAL(ERR^2))
which is to say, it is normalized by the total variance of the time series, and a factor of 2. Assuming LC is real, then really only the first half of POW is independent.

Then POW should be distributed like a chi-square with 2 degrees of freedom. The mean value should be 2, the standard deviation should be 2. I just verified this with some random data.

I verified that CHISQR_CVF() produced reasonable numbers, compared to my own MPCHILIM() function, which also computes confidence limits. Sample code below.

Craig

```
lc = randomn(seed,2000)
err = dblarr(2000) + 1
POW = ABS(FFT(LC,+1))^2 * ( 2 / TOTAL(ERR^2) )
pow1 = pow(0:1000) ;; First half of power spectrum

print, avg(pow1)
;;  ==>     1.9769791
print, stddev(pow1)
;;  ==>     1.9997902

print, chisqr_cvf(0.05d, 2d)
```

;; ==>   5.9914659
print, mpchilim(0.05d, 2d, /slevel)
;; ==>   5.9914645
help, where(pow1 GE 5.9914645d)
;; ==> <Expression>   LONG   = Array[38]
;; (in other words, 38 out of 1000 or 3.8% of data exceeded
threshold)

Subject: Re: CHISQR_CVF question.
Posted by R.G. Stockwell on Thu, 20 Aug 2009 15:21:51 GMT
View Forum Message <> Reply to Message

> "Craig Markwardt" <craig.markwardt@gmail.com> wrote in message
> news:cab41ca6-e1a4-4f73-851f->8b25ab0c1e58@k26g2000vbp.googlegroups.com...
> On Aug 19, 4:42 pm, "R.G. Stockwell" <noemai...@please.com> wrote:
>> "Paolo" <pgri...@gmail.com> wrote in message
>>
>> basically yes, abs(fft(ts))^2, and comparing it to chisquare from the
>> IDL functions.
>> I have worked on it, but I think the result is off by a factor of 2.
>> That is a factor of 2 too stringent.
>>
...
>> Perhaps you can check my understanding. If we have a 95% significance
>> level,
>> then if we make a spectrum with 1000 points, shouldnt 50 of them be above
>> that 95% line?

> Let's say we have a time series, defined like this,
> LC = time series values (array)
>   ERR = measurement uncertainty (array) of each LC point.

> I define the power spectrum in the following way,
 > POW = ABS(FFT(LC,+1))^2 * (2 / TOTAL(ERR^2))
> which is to say, it is normalized by the total variance of the time
> series, and a factor of 2. Assuming LC is real, then really only the
> first half of POW is independent.

Well, there you go. lol. I though I had a factor of 2 missing somewhere.
Although I need to examine that a bit more, since I do both the full + and -
spectrum, as well as just the +. It makes sense though.

thanks for the response,
cheers,
bob

Subject: Re: CHISQR_CVF question.
Posted by Craig Markwardt on Thu, 20 Aug 2009 16:14:25 GMT
View Forum Message <> Reply to Message

On Aug 20, 11:21 am, "R.G. Stockwell" <noemai...@please.com> wrote:
>> "Craig Markwardt" <craig.markwa...@gmail.com> wrote in message
>> news:cab41ca6-e1a4-4f73-851f->
8b25ab0c1__BEGIN_MASK_n#9g02mG7!__..__END_MASK_i?a63jfAD$z_
_@k26g2000vbp.googlegroups.com...
>> On Aug 19, 4:42 pm, "R.G. Stockwell" <noemai...@please.com> wrote:
>>> "Paolo" <pgri...@gmail.com> wrote in message
>
>>> basically yes, abs(fft(ts))^2, and comparing it to chisquare from the
>>> IDL functions.
>>> I have worked on it, but I think the result is off by a factor of 2.
>>> That is a factor of 2 too stringent.
>
> ...
>>> Perhaps you can check my understanding. If we have a 95% significance
>>> level,
>>> then if we make a spectrum with 1000 points, shouldnt 50 of them be above
>>> that 95% line?
>> Let's say we have a time series, defined like this,
>> LC = time series values (array)
>> ERR = measurement uncertainty (array) of each LC point.
>> I define the power spectrum in the following way,
>
> > POW = ABS(FFT(LC,+1))^2 * (2 / TOTAL(ERR^2))
>
>> which is to say, it is normalized by the total variance of the time
>> series, and a factor of 2. Assuming LC is real, then really only the
>> first half of POW is independent.
>
> Well, there you go. lol. I though I had a factor of 2 missing somewhere.
> Although I need to examine that a bit more, since I do both the full + and -
> spectrum, as well as just the +. It makes sense though.

Oh, I thought the factor was 10.9 something. OK good luck!

Craig

Subject: Re: CHISQR_CVF question.
Posted by R.G. Stockwell on Thu, 20 Aug 2009 16:53:00 GMT
View Forum Message <> Reply to Message

"Craig Markwardt" <craig.markwardt@gmail.com> wrote in message
news:cab41ca6-e1a4-4f73-851f-8b25ab0c1e58@k26g2000vbp.google groups.com...

On Aug 19, 4:42 pm, "R.G. Stockwell" <noemai...@please.com> wrote:
> "Paolo" <pgri...@gmail.com> wrote in message
>
> basically yes, abs(fft(ts))^2, and comparing it to chisquare from the
> IDL functions.
> I have worked on it, but I think the result is off by a factor of 2.
> That is a factor of 2 too stringent.
>
...
> Perhaps you can check my understanding. If we have a 95% significance
> level,
> then if we make a spectrum with 1000 points, shouldnt 50 of them be above
> that 95% line?

Let's say we have a time series, defined like this,
  LC = time series values (array)
  ERR = measurement uncertainty (array) of each LC point.

I define the power spectrum in the following way,
 POW = ABS(FFT(LC,+1))^2 * (2 / TOTAL(ERR^2))
****

Craig, Sorry but I am a bit confused here.

using the +1 direction is the "inverse" FFT here isn't it?
and hence it lacks the 1/N normalization that occurs on the "forward" FFT.
Is that right?

Also, total(err^2) happens to be equal to the length here, so i looks like
you are doing an inverse FFT ^2, and then dividing by len.

BUT, that is the same as doing the forward FFT (with 1/N), squaring it, then
multiplying
by len.

So, it almost looks like this just happens to be by coincidence the same as
pow = fft(lc, /forward)*length

And you have a factor of 2, which is coincidentally also the power of your
spectrum. and it appears that again this may have just coincidentally
cancelled out.

basically, I am starting with a normalization of the spectrum as:

 d = 120*randomn(seed,len)
spe = fft(d)
pspe = abs(spe[0:len/2-1])^2

; normalize wrt length and variance, so we always get the same result
pspe = pspe*(len)
pspe = pspe/stddev(d)^2

with this normalization, the mean of my spectrum is always the same.
(as i vary the length of the time series, and as i vary the standard
deviation,
above i have a stdev of 120).

Are you saying that there should be a factor of 2 in my power spectrum,
i.e. I need a final line that states pspe = pspe*2?
Because, when I do this, I do get the expected result. By expected I mean I
calculate the number of points above the cutoff level (90%) and I find
approximately
10% above, 90% below. ditto 95%, 99%.

But, I want to justify that factor of 2.

cheers,
bob

Subject: Re: CHISQR_CVF question.
Posted by David Fanning on Thu, 20 Aug 2009 16:57:21 GMT
View Forum Message <> Reply to Message

R.G. Stockwell writes:

> Craig, Sorry but I am a bit confused here.
>
> using the +1 direction is the "inverse" FFT here isn't it?
> and hence it lacks the 1/N normalization that occurs on the "forward" FFT.
> Is that right?
>
> Also, total(err^2) happens to be equal to the length here, so i looks like
> you are doing an inverse FFT ^2, and then dividing by len.
>
> BUT, that is the same as doing the forward FFT (with 1/N), squaring it, then
> multiplying
> by len.
>
> So, it almost looks like this just happens to be by coincidence the same as
> pow = fft(lc, /forward)*length
>
> And you have a factor of 2, which is coincidentally also the power of your

> spectrum. and it appears that again this may have just coincidentally
> cancelled out.
>
>
> basically, I am starting with a normalization of the spectrum as:
>
> d = 120*randomn(seed,len)
> spe = fft(d)
> pspe = abs(spe[0:len/2-1])^2
>
> ; normalize wrt length and variance, so we always get the same result
> pspe = pspe*(len)
> pspe = pspe/stddev(d)^2
>
>
> with this normalization, the mean of my spectrum is always the same.
> (as i vary the length of the time series, and as i vary the standard
> deviation,
> above i have a stdev of 120).
>
> Are you saying that there should be a factor of 2 in my power spectrum,
> i.e. I need a final line that states pspe = pspe*2?
> Because, when I do this, I do get the expected result. By expected I mean I
> calculate the number of points above the cutoff level (90%) and I find
> approximately
> 10% above, 90% below. ditto 95%, 99%.
>
> But, I want to justify that factor of 2.

You know what? I'm just going to stick with mastering
that down-the-line backhand, thank you very much!

Cheers,

David

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

Subject: Re: CHISQR_CVF question.
Posted by Craig Markwardt on Sat, 22 Aug 2009 17:51:06 GMT
View Forum Message <> Reply to Message

On Aug 20, 12:53 pm, "R.G. Stockwell" <noemai...@please.com> wrote:
> "Craig Markwardt" <craig.markwa...@gmail.com> wrote in message

>
> news:cab41ca6-e1a4-4f73-851f-8b25ab0c1e58@k26g2000vbp.google groups.com...
> On Aug 19, 4:42 pm, "R.G. Stockwell" <noemai...@please.com> wrote:
>
>> "Paolo" <pgri...@gmail.com> wrote in message
>
>> basically yes, abs(fft(ts))^2, and comparing it to chisquare from the
>> IDL functions.
>> I have worked on it, but I think the result is off by a factor of 2.
>> That is a factor of 2 too stringent.
>
> ...
>> Perhaps you can check my understanding. If we have a 95% significance
>> level,
>> then if we make a spectrum with 1000 points, shouldnt 50 of them be above
>> that 95% line?
>
> Let's say we have a time series, defined like this,
>   LC = time series values (array)
>   ERR = measurement uncertainty (array) of each LC point.
>
> I define the power spectrum in the following way,
>   POW = ABS(FFT(LC,+1))^2 * (2 / TOTAL(ERR^2))
> **
>
> Craig, Sorry but I am a bit confused here.
>
> using the +1 direction is the "inverse" FFT here isn't it?
> and hence it lacks the 1/N normalization that occurs on the "forward" FFT.
> Is that right?
>
> Also, total(err^2) happens to be equal to the length here, so i looks like
> you are doing an inverse FFT ^2, and then dividing by len.

Bob, I was just telling you (and showing explicitly) what "works for
me." My use of the FFT(,+1) notation arises because the documentation
indicates it is faster, but also because I'll put in my own
normalization factors, thank you very much.

The "conventions" for FFT direction and normalization are so varied
across different fields, that there really is no convention!

> And you have a factor of 2, which is coincidentally also the power of your
> spectrum. and it appears that again this may have just coincidentally
> cancelled out.

I believe my power of 2 formally comes from adding + and -

frequencies, one for each. But in any case, it's a convenient scaling because as it is defined, it allows one to directly do a chi-square probability test for any given power, i.e. CHISQR_PDF() or MPCHITEST (), since each power is distributed exactly as a chi-square with 2 d.o.f. As we've seen, the scaling is nearly arbitrary, so for probability tests, I find it best to scale to a useful quantity.
[For variability measures, it's another matter.]

Craig
