The χ^2 Test for Goodness of Fit

John Quinn

Definition of χ^2

- The general definition of χ^2 is:

$$\chi^2 = \sum \left(\frac{\text{measured} - \text{expected}}{\text{error}} \right)^2$$

- χ^2 describes the distribution of the sum of the squares of numbers drawn from a standard normal distribution (i.e. a normal distribution with mean 0 and standard deviation 1).
- The number of independent points summed is called the number of degrees of freedom (DoF, v)
- There is a different χ^2 curve for each number of DoF (see later slide)

Definition of χ^2

- We previously met the χ^2 sum when we covered curve fitting:

$$S(pars) = \chi^2(pars) = \sum_{i=1}^{N} \left[\frac{y_i - f(x_i, pars)}{\sigma_{y_i}} \right]^2$$

- The fitting algorithm finds the best-fit parameters by minimising the χ^2 sum.
- The χ^2 sum value using the best-fit parameters is often called χ^2_{min}
- χ^2_{min} thus gives us some indication of the agreement between the data and what was expected.
- Note: "χ^2_{min}" will be used but the tests equally apply to a direct comparison of data with some expected value where no minimisation has been performed.

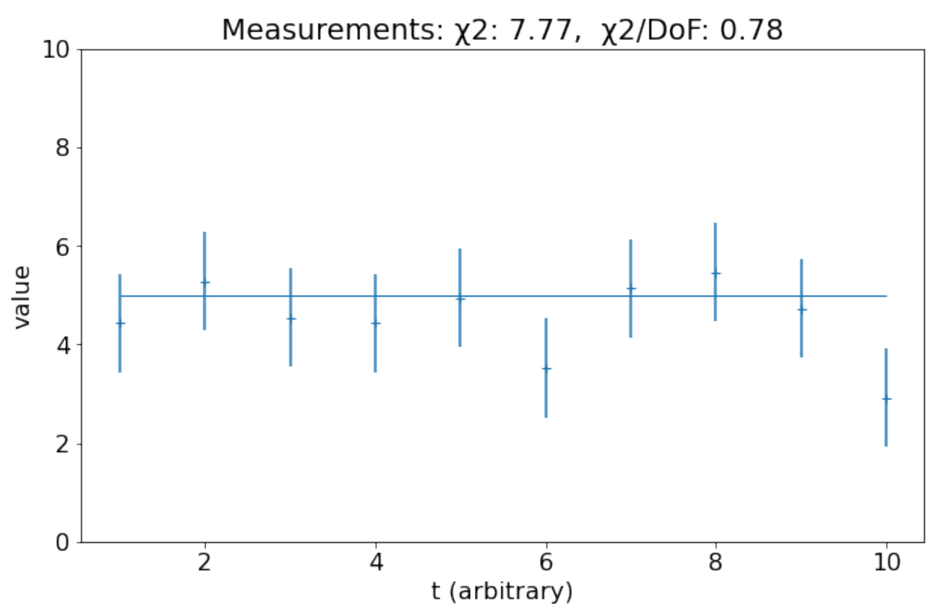
Definition of χ^2

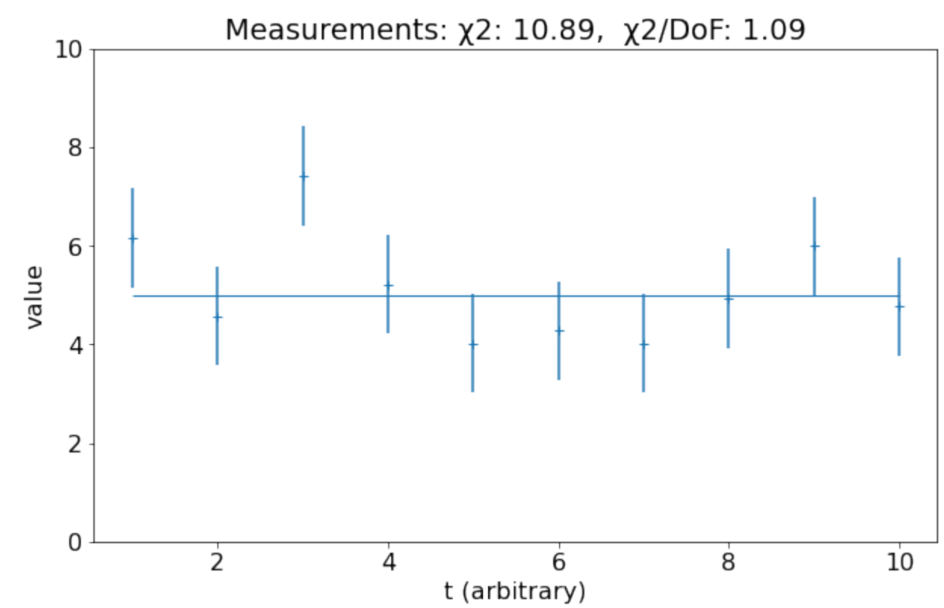
$$\chi^2 = \sum_{i=1}^{N} \left(\frac{\text{measured}_i - \text{expected}}{\text{error on measured}_i} \right)^2$$

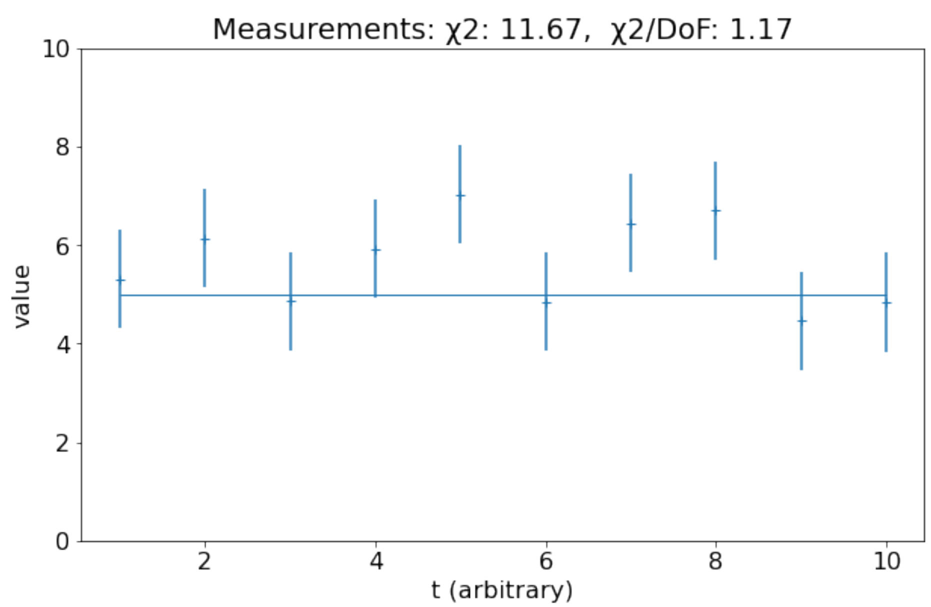
- Agreement:
  - for good agreement we would expect points to be, on average, about one standard error from the expected, and so χ^2_{min} should be ~ N.
    - this is approximately true, in fact we get $\chi^2_{min} \approx \upsilon = N - N_c$
    - where υ is called the "Number of Degrees of Freedom" and is equal to the number of data points, N, minus the number of constraints (free parameters in fit function), N_c, derived from the data.
- χ^2/υ is called the "Reduced Chi Squared" and for a good fit is ≈ 1.
- We will never get a value for $\chi 2/\upsilon$ of 1 exactly.
- So, if we get a value different from 1 what do we conclude? How much variation is expected?

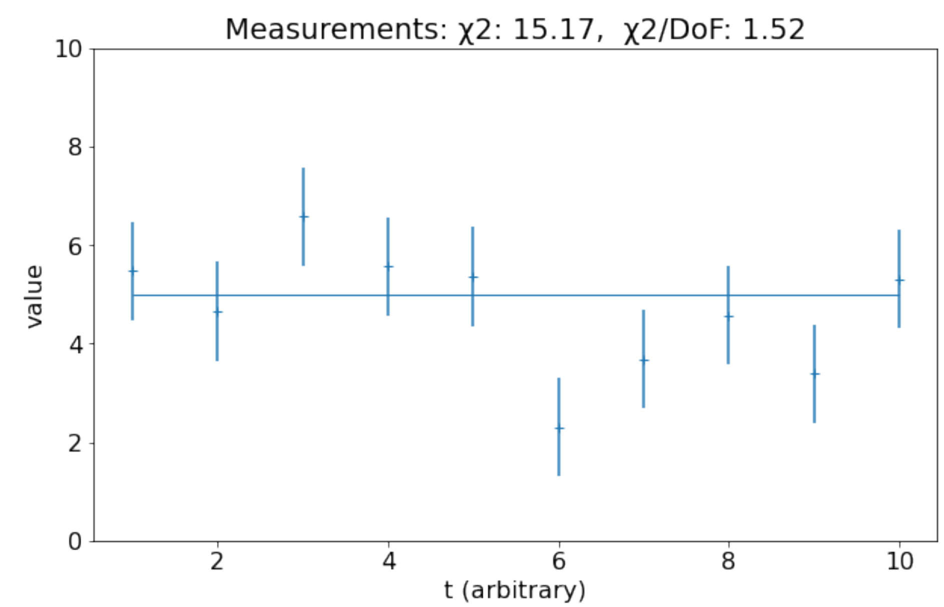
Simulate some data…

- Simulate 10 data points from a normal distribution of μ = 5 and σ = 1.
- Compare to expected value of 5 (i.e. expected value not derived from the data so N_c = 0 and ν = N = 10).



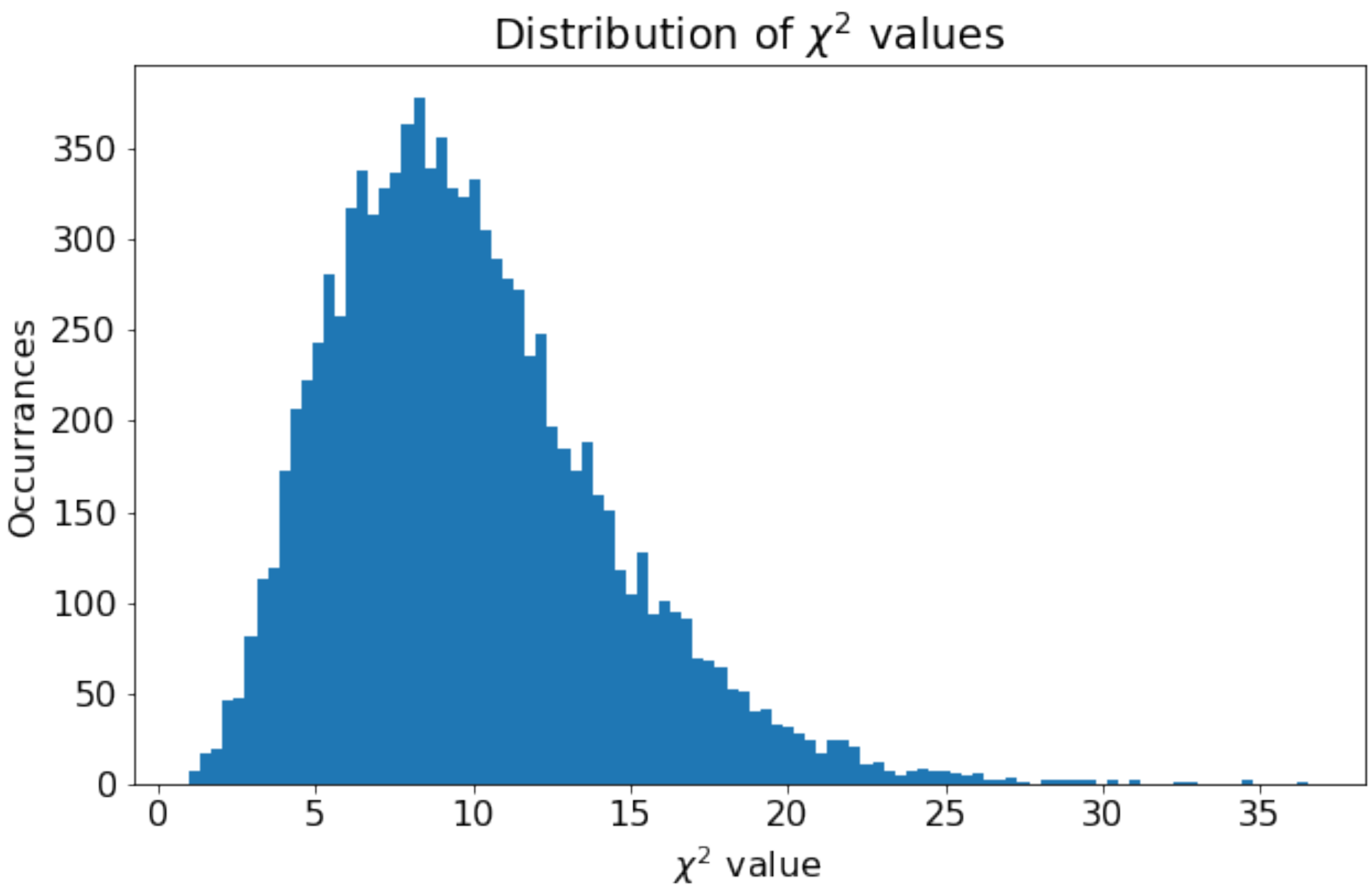






Simulate some data...

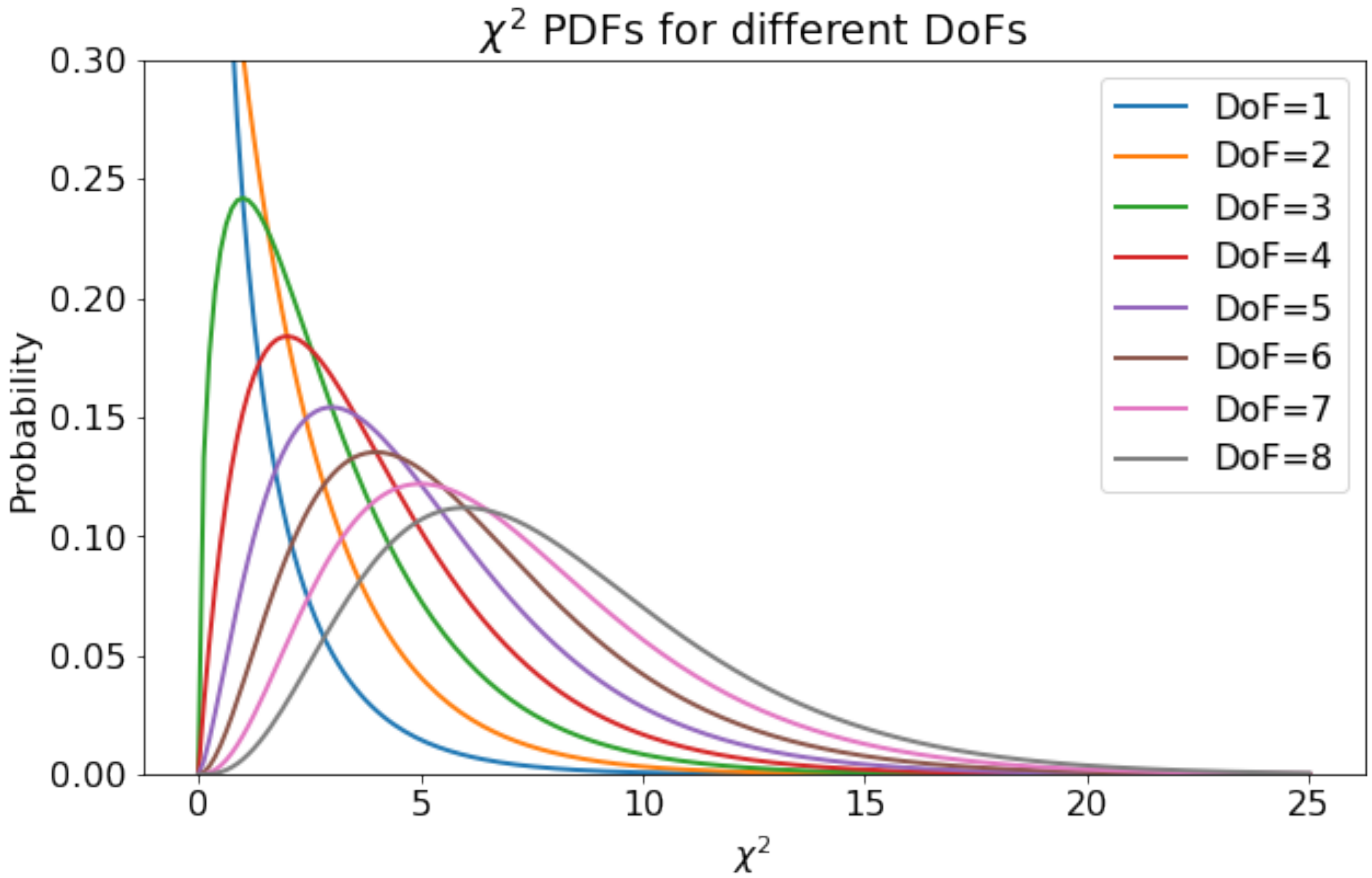
- Runs 10,000 times and look at the distribution....



- The above distribution of χ^2_{min} is expected - it is the χ^2 distribution for 10 degrees of freedom.

The χ^2 PDF

- For each number of degrees of freedom there is a χ^2 PDF which describes the expected distribution of χ^2 values for purely statistical fluctuations of the measurements.



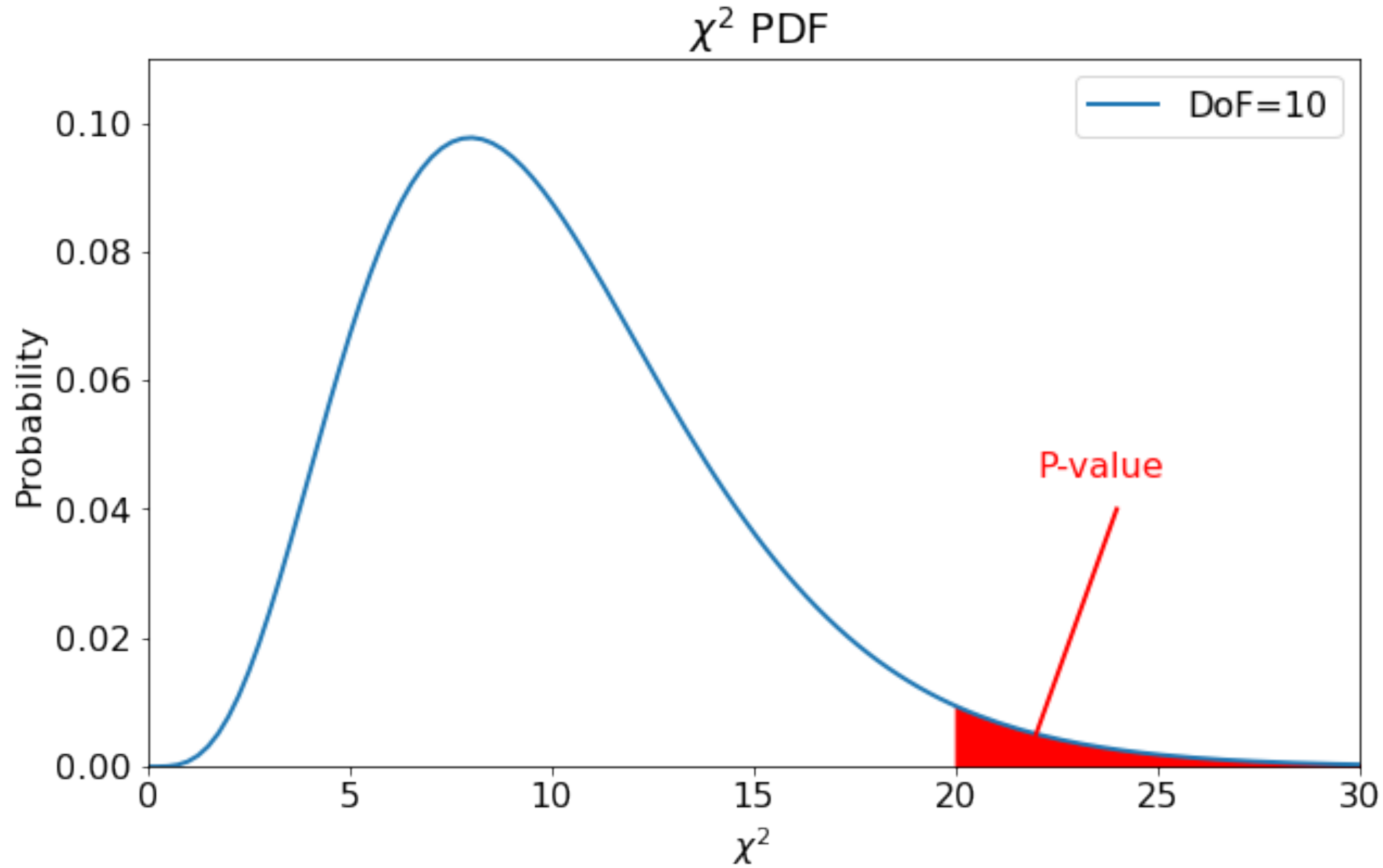
- Each χ^2 curve has:
  - an expectation value (expected mean) of ν (number of DoF)
  - standard deviation ($\sigma_{\chi^2} = \sqrt{2\nu}$) (asymmetric for small ν)

χ^2 Probability

- Very small values of χ^2_{min} indicate too good agreement between measured and expected and generally indicate that the error bars are over-estimated and not an accurate estimate of the variance within the data points, or the fit is over constrained.
- Very large values of χ^2 are rare and we can determine the probability (P-value) of obtaining a value of $\chi^2 \geq \chi^2_{min}$ by chance (for normally distributed errors if the model and data agree) is given by integrating the χ^2 distribution from the observed value to infinity:

$$P(\chi^2_{min} \leq \chi^2 \leq \infty) = \int_{\chi^2_{min}}^{\infty} P_{DF}(\chi^2; \nu)\, d\chi^2$$

- Look up in table, or
- use scipy.stats.chi2.sf()



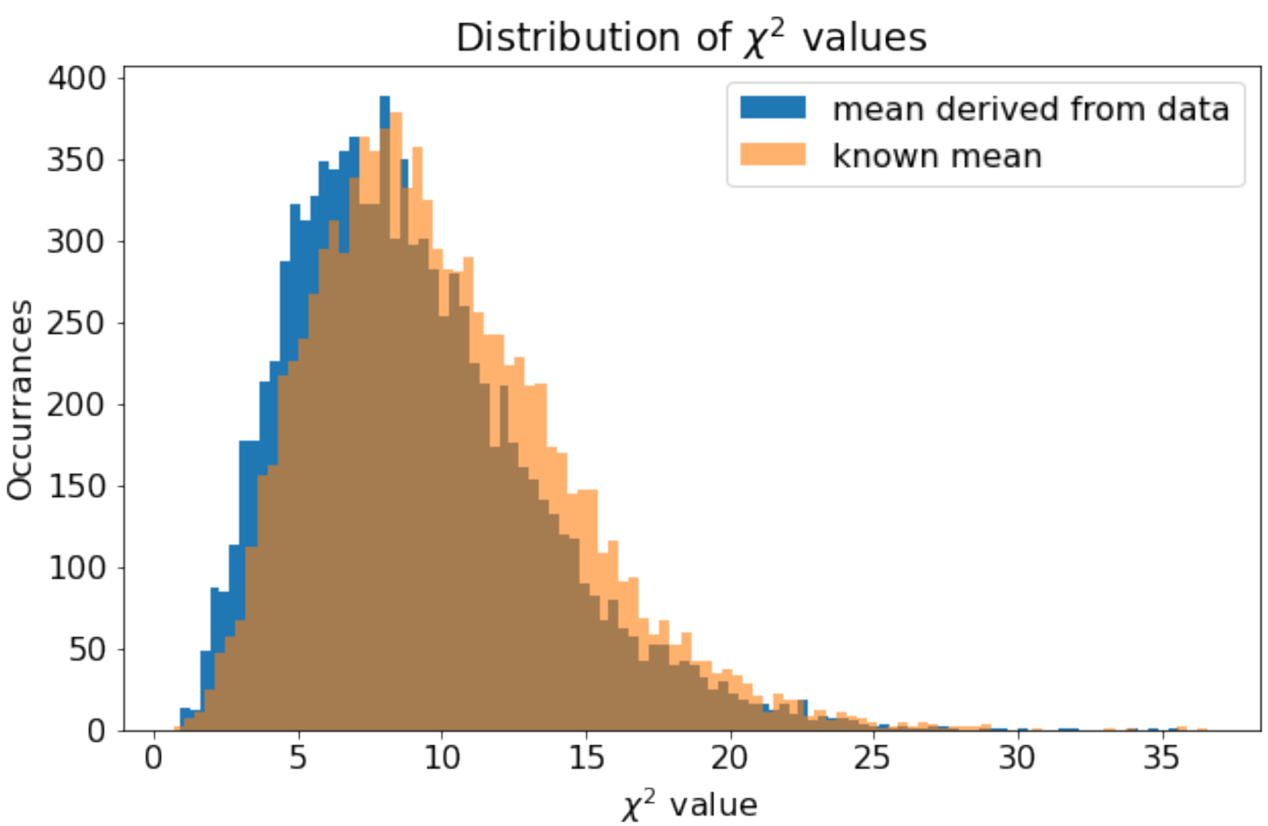
Agreement?

- (Hughes & Hase, Chapter 8)
- χ^2 test for goodness of fit are best evaluated using the P-value.
- Using the reduced χ^2 (χ^2_{min}/ν) as an indication strongly depends on the number of degrees of freedom and is more difficult to interpret

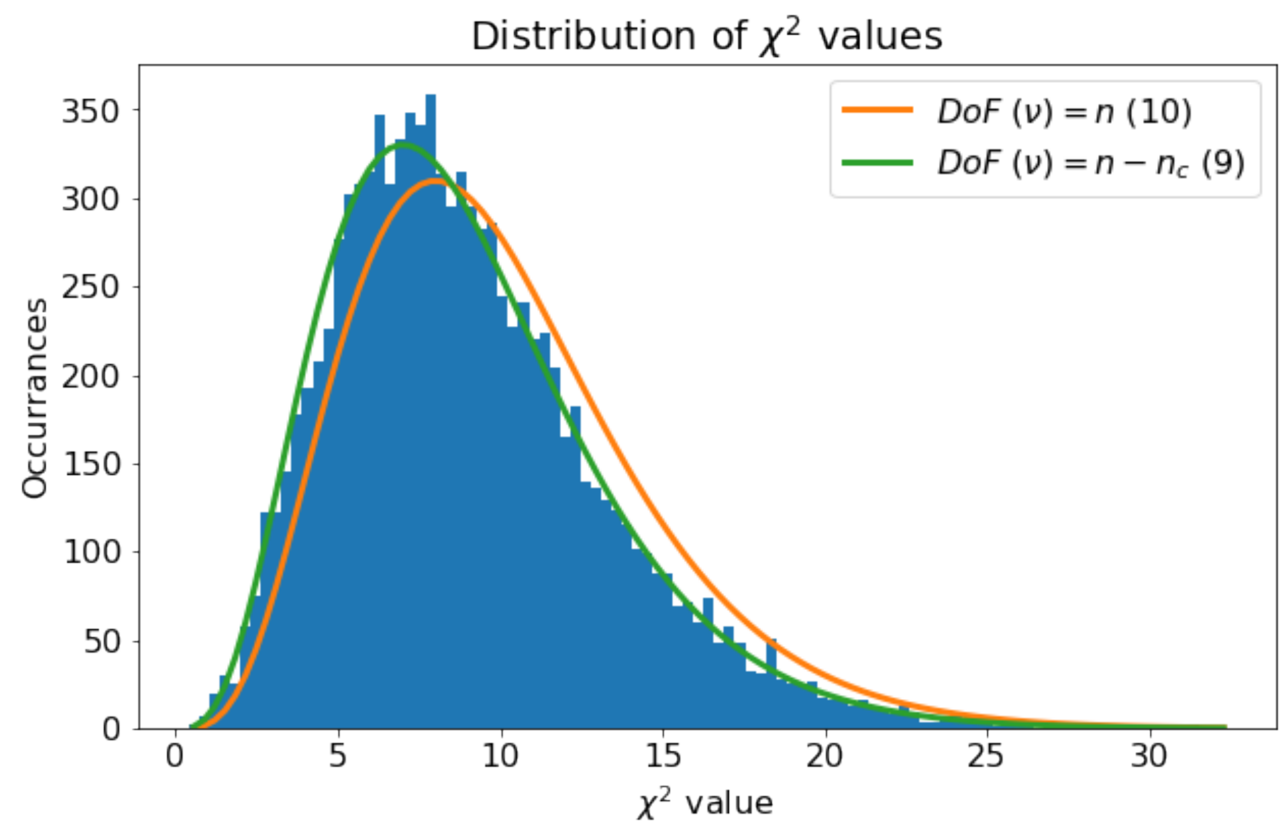
Agreement	$P(\chi^2_{min};\nu)$	χ^2_{min}/ν	Comment
Excellent	~0.5	~ 1	
Too good	➔ 1	<< 1	Errors too large / fit over constrained
Questionable	$\sim 10^{-3}$	>2 for $\nu \approx 10$ >1.5 for $50 \lesssim \nu \lesssim 100$	
Poor	$\lesssim 10^{-4}$	$\gtrsim 3$	

Number of Degrees of Freedom

- Recall: the number of degrees of freedom is $\nu = N - N_c$
  - N is the number of data points
  - N_c is the number of constraints (parameters) derived from the data
- Examine distribution of χ^2 for the 10 simulated data points again, but this time we will not compare it the expected mean value (5) known in advance (parameter which went into simulated data) but compare to the mean value derived from the data…



simulated data for both cases



simulated data (mean derived from data) + theoretical χ^2 PDFs

Counts - the Poisson case

- Recall:
  - the Poisson distribution describes the expected number of counts in an interval for a process with a known/expected mean.
  - for a sufficient number of mean counts (≳5) the Poisson distribution is represented by Gaussian of mean μ and standard deviation $\sigma = \sqrt{\mu}$.
- Thus,

$$\chi^2 = \sum_{i=i}^{N} \left(\frac{\text{observed counts}_i - \text{expected counts}}{\text{error on counts}_i} \right)^2$$

$$= \sum_{i=i}^{N} \left(\frac{O_i - E_i}{\sigma_{E_i}} \right)^2$$

$$= \sum_{i=i}^{N} \left(\frac{O_i - E_i}{\sqrt{E_i}} \right)^2$$

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}$$

Counts - the Poisson case

Measurements with Gaussian Errors

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2$$

measured errors

Measurements with Poisson Errors

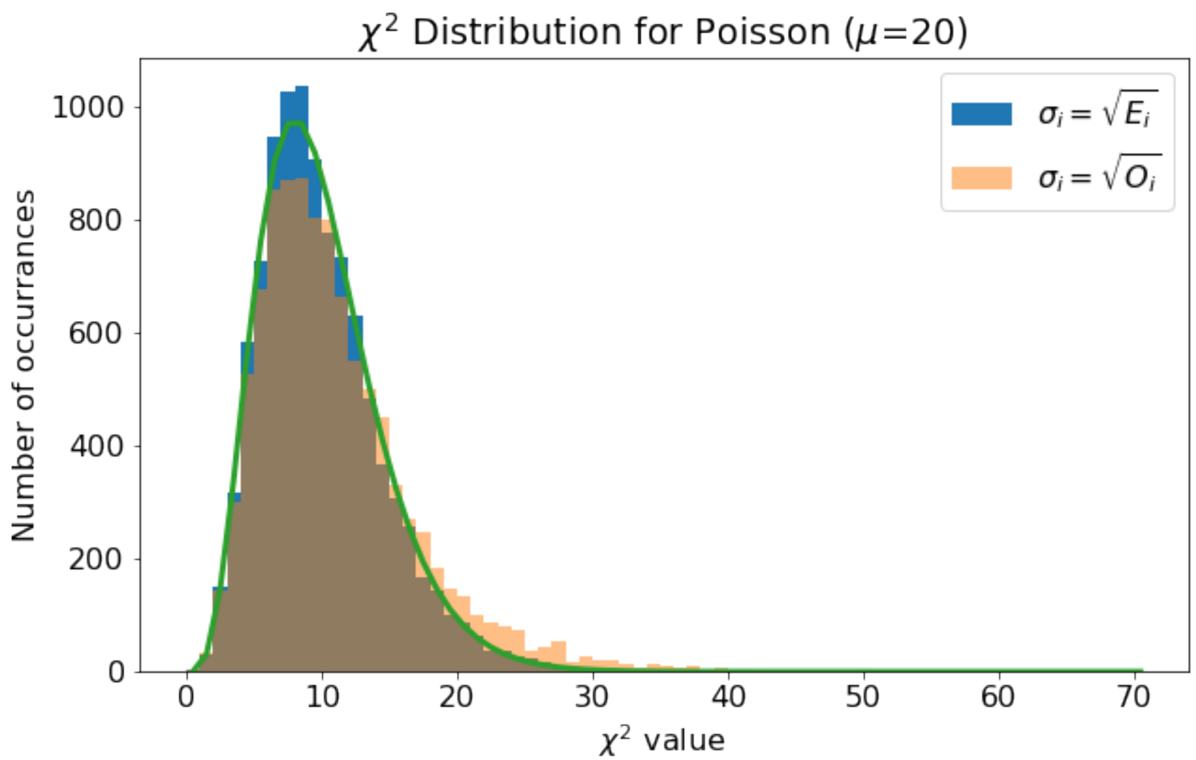
$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}$$

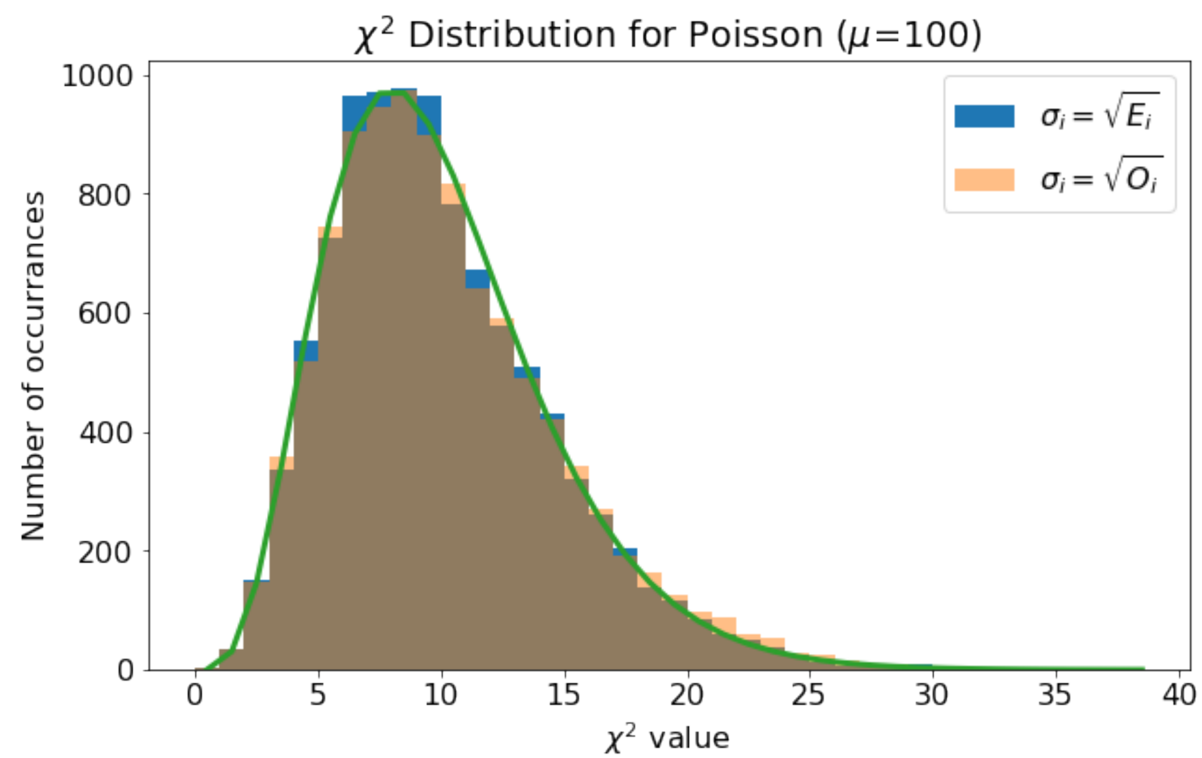
error on expected counts^2

- For counting experiments we use the error from the number of counts expected $\sqrt{E_i}$, rather than the error on each individual counts observed ($\sqrt{O_i}$).
- Why?
  - we are comparing to a known distribution from theory with expected fluctuations about E with standard deviation $\sqrt{E}$
  - with fluctuations possibly giving small numbers of counts (O_i) the estimated error (from $\sqrt{O_i}$) may not be Gaussian and can lead to anomalously large χ^2 values
    - consider extreme case of $O_i = 0$ counts which gives an infinite χ^2!

Counts - the Poisson case

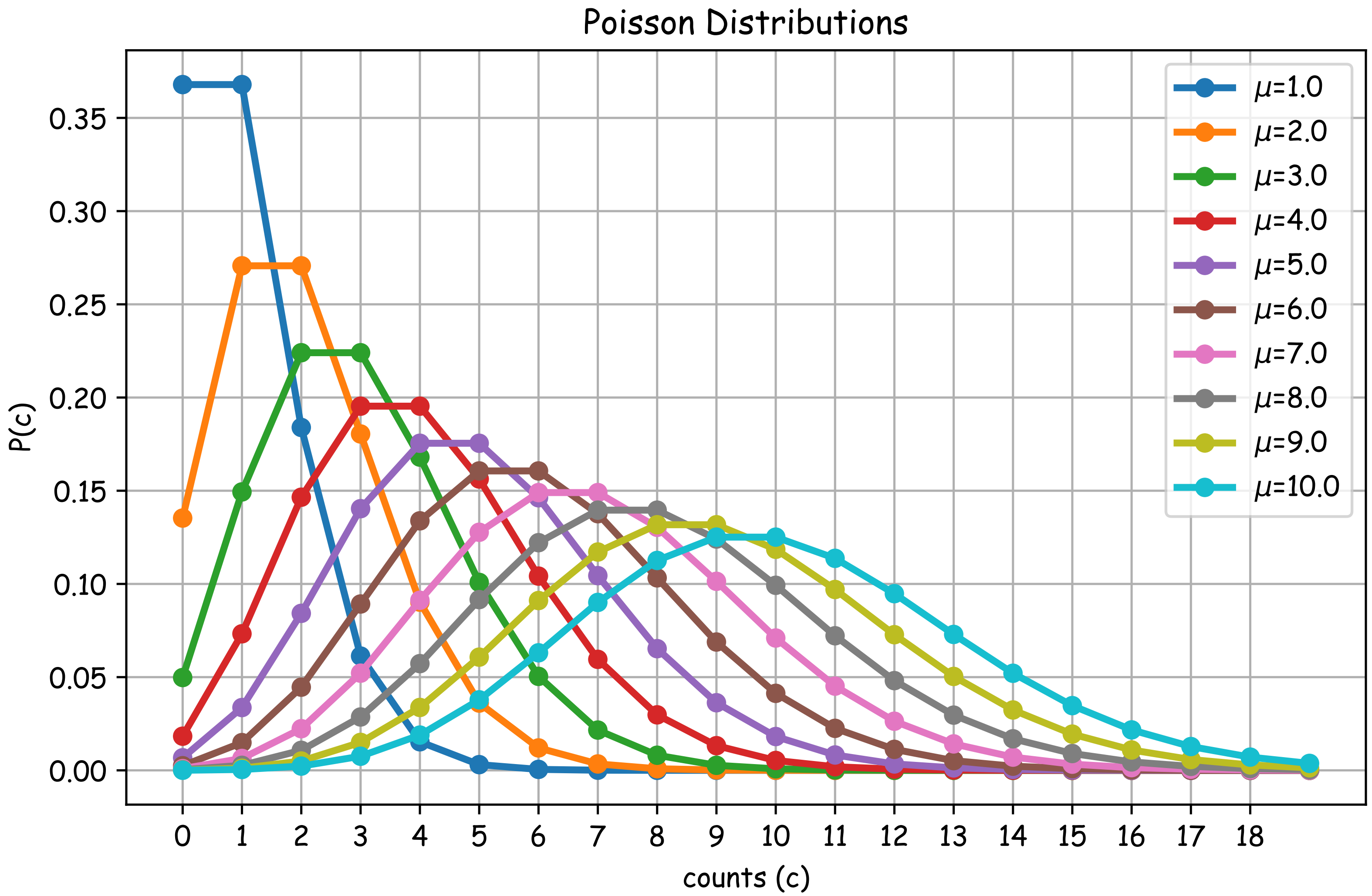
- As the mean number of counts expected becomes large, the two give the same result, but even for moderately large means ($\mu \approx 20$) the distribution is skewed





(Above is 10,000 simulations of 10 data points each drawn from a Poisson distribution of the appropriate mean and calculating the χ^2 sum for those 10 points being consistent with the expected value.)

Counts - the Poisson case

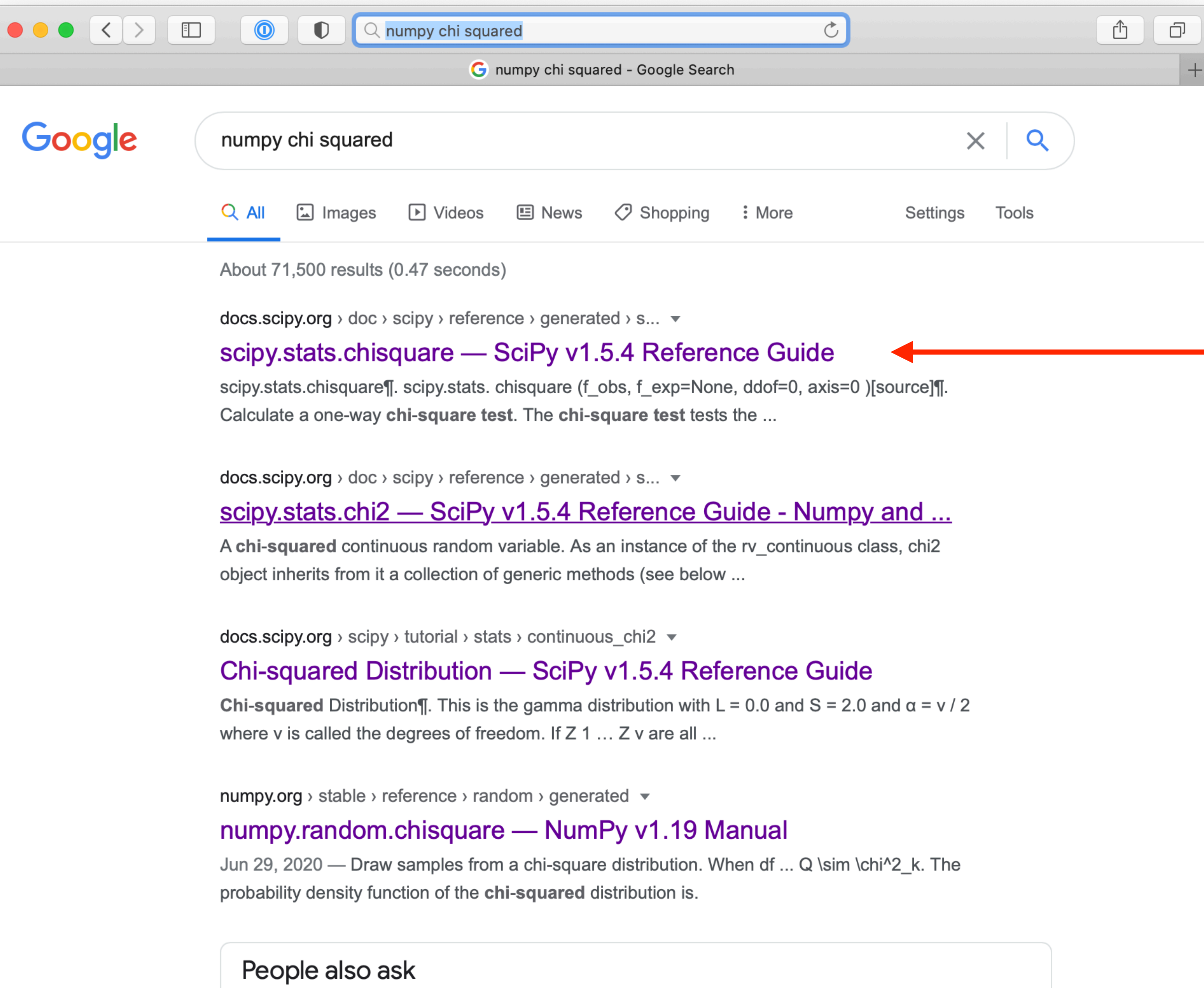


- For χ^2 tests involving counts, including comparing histograms always use $\sqrt{E_i}$ and $E_i > 5$ is the accepted minimum expected value per bin (re-bin the data if necessary if a histogram)

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(O_i - E_i\right)^2}{E_i}$$

Use the correct χ^2 test!

- Google: numpy chi squared:



Top hit:
scipy.stats.chisquare

Use the appropriate χ^2 test!

docs.scipy.org/doc/scipy/reference/generated/scip

scipy.stats.chisquare — SciPy v1.5.4 Reference Guide

SciPy.org

SciPy.org | Docs | SciPy v1.5.4 Reference Guide | Statistical functions (scipy.stats) | index | modules | next | previous

scipy.stats.chisquare

scipy.stats.**chisquare**(*f_obs*, *f_exp=None*, *ddof=0*, *axis=0*) [source]

Calculate a one-way chi-square test.

The chi-square test tests the null hypothesis that the categorical data has the given frequencies.

Parameters: **f_obs** : ***array_like***

Observed frequencies in each category.

f_exp : ***array_like, optional***

Expected frequencies in each category. By default the categories are assumed to be equally likely.

ddof : ***int, optional***

"Delta degrees of freedom": adjustment to the degrees of freedom for the p-value. The p-value is computed using a chi-squared distribution with `k - 1 - ddof` degrees of freedom, where *k* is the number of observed frequencies. The default value of *ddof* is 0.

axis : ***int or None, optional***

The axis of the broadcast result of *f_obs* and *f_exp* along which to apply the test. If axis is None, all values in *f_obs* are treated as a single data set. Default is 0.

Returns: **chisq** : ***float or ndarray***

The chi-squared test statistic. The value is a float if *axis* is None or *f_obs* and *f_exp* are 1-D.

p : ***float or ndarray***

The p-value of the test. The value is a float if *ddof* and the return value *chisq* are scalars.

See also:

scipy.stats.power_divergence

Notes

This test is invalid when the observed or expected frequencies in each category are too small. A typical rule is that all of the observed and expected frequencies should be at least 5.

The default degrees of freedom, k-1, are for the case when no parameters of the distribution are estimated. If p parameters are estimated by efficient maximum likelihood then the correct degrees of freedom are k-1-p. If the parameters are estimated in a different way, then the dof can be between k-1-p and k-1. However, it is also possible that the asymptotic distribution is not chi-square, in which case this test is not appropriate.

Previous topic

scipy.stats.ttest_rel

Next topic

scipy.stats.power_divergence

Quick search

search

Note: "frequencies" = "counts"

This is for Poisson counting experiments where error is taken as the $\sqrt{E_i}$.

It makes no sense to apply this to data and errors that are not counts (e.g. taking the square root of a length or voltage or intensity as the error).

Conclusions

- χ^2_{min} can be used to compare the agreement between measurements and theory/best-fit function.
- While the reduced χ^2 is a useful indication for good agreement (for good agreement we expect $\chi^2/\nu \approx 1$) it is better to calculate the P-value for the given observed.
- The correct number of degrees of freedom must be used for calculating the reduced chi-square or P-value.
  - in Python use: scipy.stats.chi2.sf()
- P-values ~10^{-3} are questionable and $\lesssim 10^{-4}$ indicate bad agreement.
- Make sure to use the correct form of the chi-square test and errors whether you are looking at measurement errors or counting errors:

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i)}{\sigma_{y_i}} \right)^2 \qquad \chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}$$

- Don't blindly apply formula (or scipy/numpy) functions without understanding exactly what they are for!