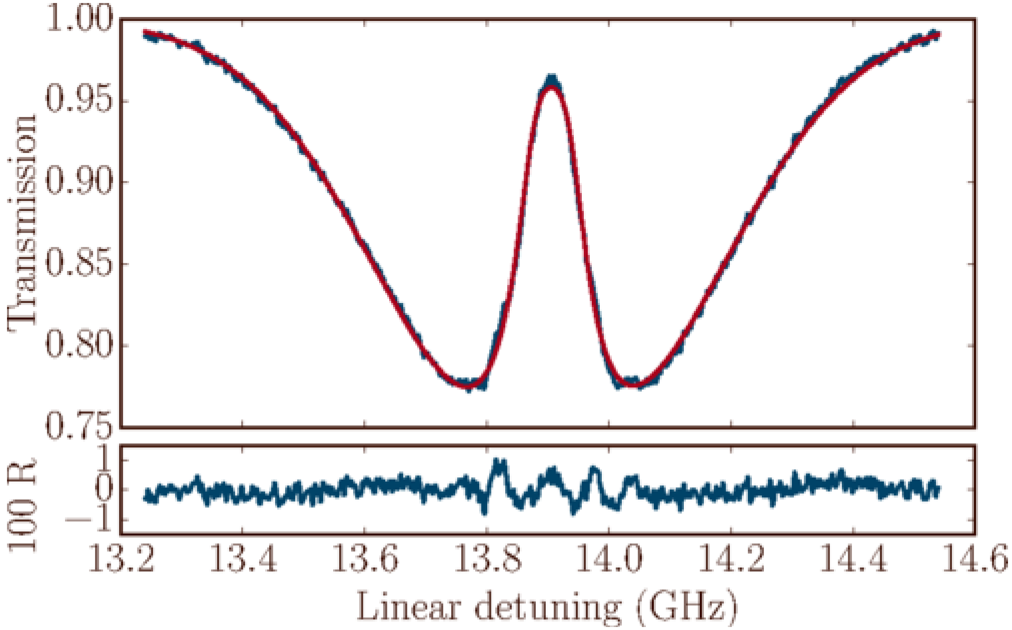
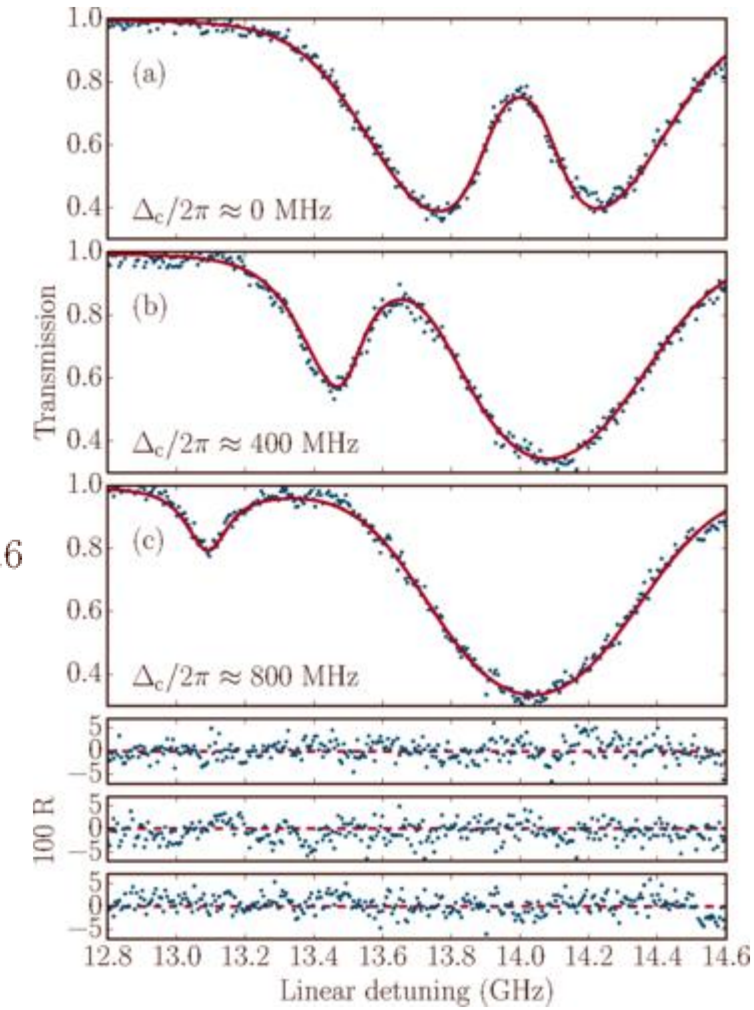
HYPOTHESIS TESTING –

How good are our models?



$$\chi(v) = \frac{4i\hbar d_{\mathrm{p}}^2/\epsilon_0}{\gamma_{\mathrm{p}} - i\delta_1(v) + \frac{\Omega_{\mathrm{c}}^2/4}{\gamma_{\mathrm{c}} - i\delta_2(v)}} N$$

$$T = \exp\left(- k_{\mathrm{p}} l \int_{-\infty}^{\infty} \mathrm{Im}[\chi(v)] n(v) dv \right)$$



Whiting *et al. Phys. Rev. A* **93** 043854 (2016)

Degrees of freedom

If we have measured *N* independent data points and we are fitting a model with $\mathcal{N}$ parameters

then the **number of degrees of freedom** υ is defined as

$$\nu = N - \mathcal{N}$$

The more data points that are unconstrained, the more robust a statistical estimate of parameters such as the mean, variance and χ^2 become.

$$\chi^2 = \sum_i \frac{(y_i - y(x_i))^2}{\alpha_i^2}$$

Measurements and their Uncertainties, P. 102

The χ^2 probability distribution function

As the χ^2 statistic is a random variable it also has a normalised probability distribution function, given by

$$X\left(\chi^2;\nu\right)=\frac{\left(\chi^2\right)^{\left(\frac{\nu}{2}-1\right)}\exp\left[-\chi^2/2\right]}{2^{\nu/2}\,\Gamma\left(\nu/2\right)}$$

As with other probability distribution functions, the probability of obtaining a value of χ^2 between χ^2_{min} and ∞ is given by the cumulative probability function, $P\,(\chi^2_{min};\,\upsilon)$:

$$P\left(\chi^2_{\min}\le\chi^2\le\infty;\nu\right)=\int_{\chi^2_{\min}}^{\infty}X\left(\chi^2;\nu\right)\,\mathrm{d}\chi^2$$

Measurements and their Uncertainties, P. 104

Using χ^2 as a hypothesis test

We expect that if the proposed model is in good agreement with the data that χ^2_{min} will be close to the *mean* of the χ^2 distribution and so

$$\boxed{\chi^2_{\text{min}} \approx \nu}$$

- For a reasonable fit, the value of $P\left(\chi^2_{\text{min}};\nu\right) \approx 0.5$.
- If $P\left(\chi^2_{\text{min}};\nu\right) \to 1$ check your calculations for the uncertainties in the measurements, α_i.
- The null hypothesis is generally **not rejected** if the value of χ^2_{min} is within $\pm 2\sigma$ of the mean, ν, i.e. in the range $\nu - 2\sqrt{2\nu} \leq \chi^2_{\text{min}} \leq \nu + 2\sqrt{2\nu}$.
- The null hypothesis is **questioned** if $P\left(\chi^2_{\text{min}};\nu\right) \approx 10^{-3}$ or $P\left(\chi^2_{\text{min}};\nu\right) > 0.5$.
- The null hypothesis is **rejected** if $P\left(\chi^2_{\text{min}};\nu\right) < 10^{-4}$.

The reduced χ^2 statistic

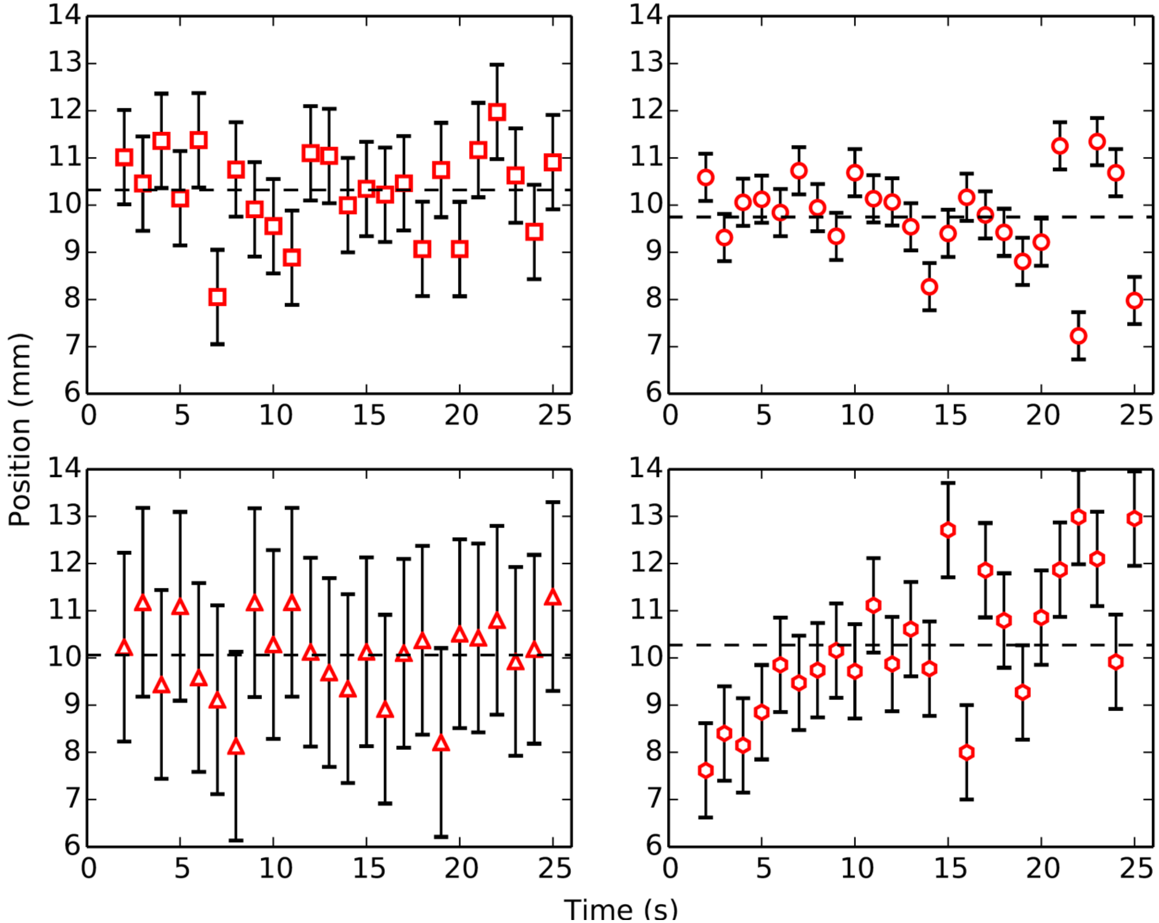
we can obtain a fast indication as to whether the null hypothesis should be rejected by considering the so-called **reduced chi squared statistic**, χ^2_ν, which is the value of χ^2_{min} divided by the number of degrees of freedom:

$$\boxed{\chi^2_\nu = \frac{\chi^2_{\min}}{\nu}}$$

- For a reasonable fit the value of $\chi^2_\nu \approx 1$.
- If $\chi^2_\nu \ll 1$ check your calculations for the uncertainties in the measurements, α_i.
- The null hypothesis is **questioned** if $\chi^2_\nu > 2$ for $\nu \approx 10$.
- The null hypothesis is **questioned** if $\chi^2_\nu > 1.5$ if ν is in the approximate range $50 \leq \nu \leq 100$.

Measurements and their Uncertainties, P. 107

- Here are 4 different data sets of a quantity which should not vary with time



What constitutes a good fit?

- Two-thirds of the data points should be within one standard error of the theoretical model.
- χ^2_ν is ≈ 1.
- $P\left(\chi^2_{\min}; \nu\right) \approx 0.5$.
- A visual inspection of the residuals shows no structure.
- A test of the autocorrelation of the normalised residuals yields $\mathcal{D} \approx 2$.
- The histogram of the normalised residuals should be Gaussian, centred on zero, with a standard deviation of 1.

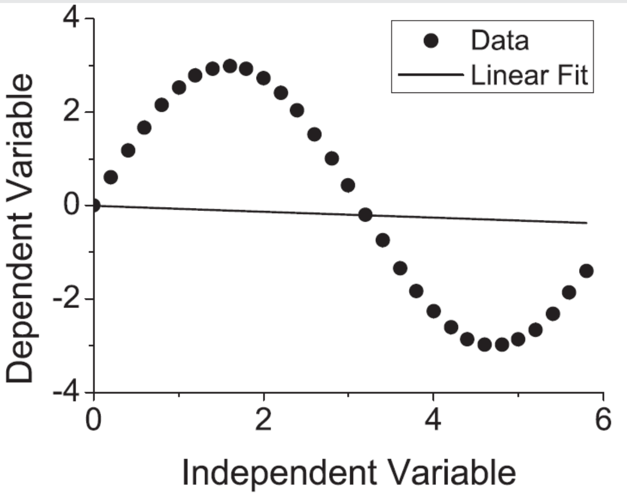


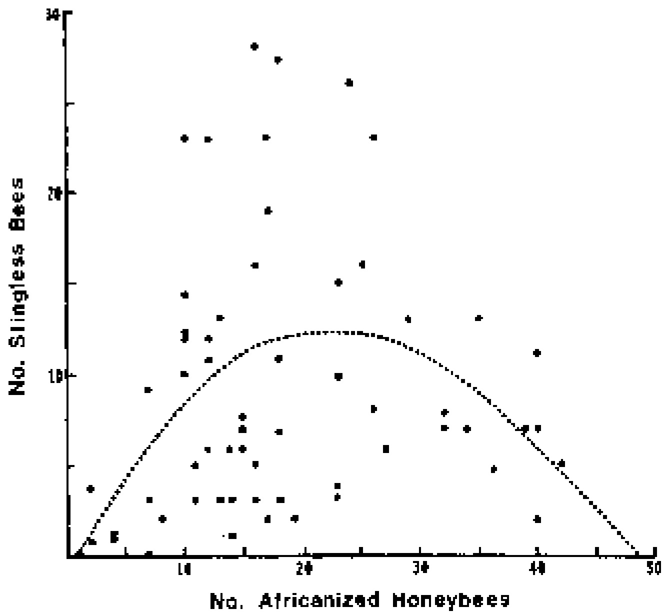
Fig. 5.9 The least-squares best-fit straight line to a sinusoidal variation. Error bars smaller than symbol size. **Health warning** It is possible to use the method of least squares to find the 'best-fit' straight line for *any* data set. This *does not* mean that a linear fit is an appropriate theoretical model.

Measurements and their Uncertainties, P. 111

Example 3 - A Very Silly Fit

Two of the very highest prestige scientific journals in the world are *Nature* and *Science*. Here is a figure and caption from an article published by David W. Roubik in Science **201** (1978), 1030.

From *"Error Analysis in Experimental Physical Science" David Harrison University of Toronto*



The dashed line is a quadratic polynomial (given by $y = -0.516 + 1.08x - 0.23x^2$) which gave the best fit to the points.

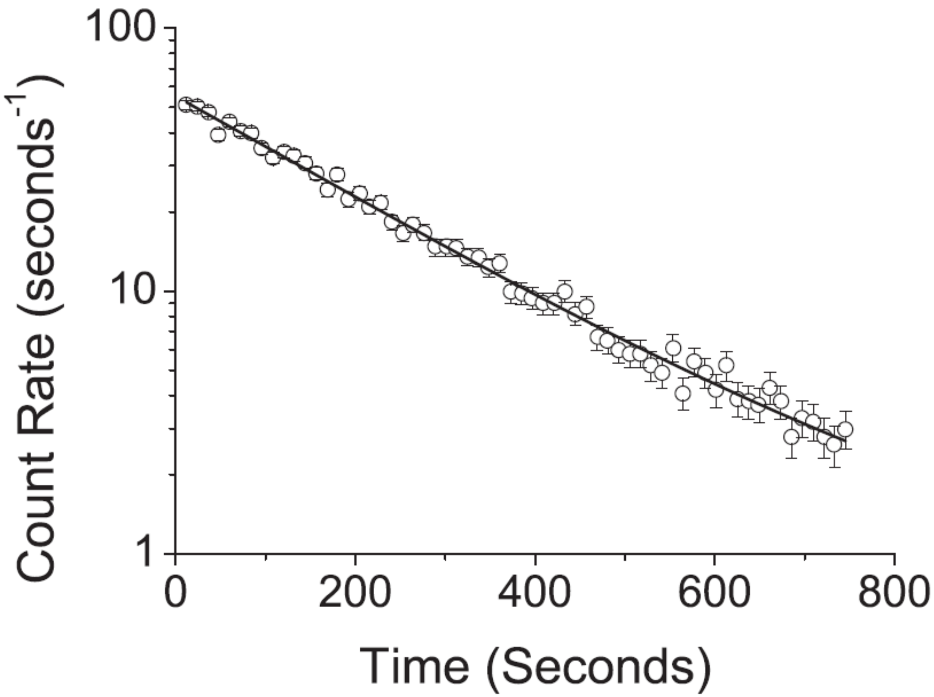
It seems fairly clear that this "best fit to the points" in fact has no relationship whatsoever to what the data actually look like. In fact, some think the data look more like a *duck*, with the beak in the upper-left and pointing to the left

Testing distributions using χ^2

AtMol. Durham Atomic & Molecular Physics

The procedure to perform a χ^2 test for a distribution is as follows:

- Construct a histogram of the sample distribution, O_i.
- For the same intervals construct a histogram of the expected occurrences, E_i.
- Combine sequential bins until $E_i > 5$ for all bins.
- Calculate χ^2 using $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$.
- A good fit will have: $\chi_\nu^2 \approx 1$ and $P\left(\chi^2; \nu\right) \approx 0.5$.
- A poor fit will have: $\chi_\nu^2 << 1$ or $\chi_\nu^2 > 3$, and $P\left(\chi^2; \nu\right) \to 1$ or $P\left(\chi^2; \nu\right) < 10^{-4}$.



of $\chi^2_{\min} = 53.5$. There are 62 data points and as there are three fit parameters the number of degrees of freedom is $\nu = 59$. The reduced χ^2 value is therefore $\chi^2_\nu = 0.9$. As this is less than 1.5 (Table 8.1), and not significantly less than 1, we do not reject the null hypothesis. We can further quantify the quality of fit by using eqn (8.4) to determine the probability of obtaining the value $\chi^2_{\min} = 53.5$, or larger, for 59 degrees of freedom. Using suitable look-up tables we find $P(53.5; 59) = 0.68$. As this probability is close to 0.5 there is, again,

nt rate (the number of
) as a function of time for
Ba isotope. The weighted
own as a solid line.

Measurements and their Uncertainties, P. 109

Worked example 2 – testing different models to a data set

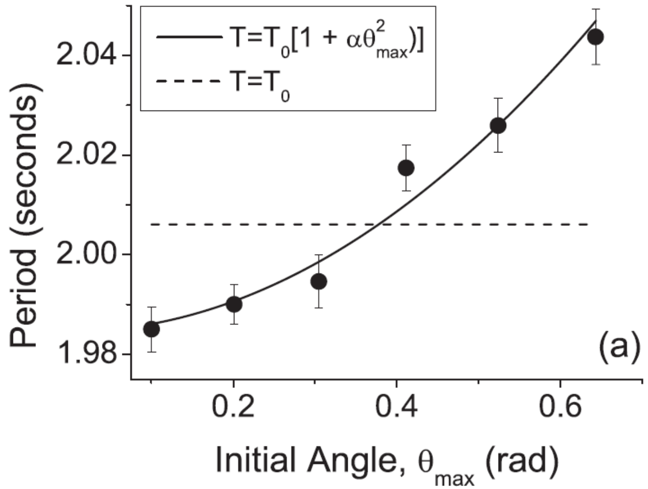


Fig. 8.7 Experimental data (filled circles) for the dependence of the period of a pendulum on the initial angular displacement. In (a) the best-fit constant period is shown as a dashed line, and a model comprising a constant with a quadratic correction shown as the solid line. In (b) a straight-line fit is made to the data.

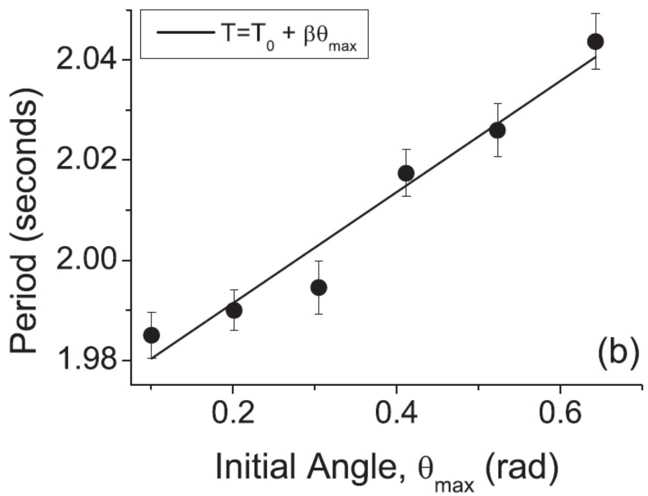


Table 8.2 Three different models are used for the dependence of the period of oscillation of a pendulum on the initial angular displacement.

Model	Degrees of freedom	$\chi^2_{\min}$	χ^2_{ν}	$P(\chi^2_{\min};\nu)$
$T = T_0$	5	107.2	21.4	1.6×10^{-21}
$T = T_0\left[1 + \alpha\,\theta^2_{\max}\right]$	4	3.39	0.9	0.49
$T = T_0\left[1 + \beta\,\theta_{\max}\right]$	4	4.39	1.1	0.36

Measurements and their Uncertainties, P. 109

Worked example 3 – testing a discrete distribution

Table 8.3 Comparing experimental radioactive decays with a Poisson model. The first column is the number of counts, the second is the occurrence of each count. Column three is the Poisson probability for obtaining a given number of counts using the mean count of the data. The fourth column gives the expected number of occurrences, and χ^2 is calculated in the fifth column.

Number of counts	O_i		Prob. (%)	E_i		$\frac{(O_i - E_i)^2}{E_i}$
0	0		0.05	0.03		
1	0		0.40	0.23		
2	0	7	1.50	0.87	7.44	0.026
3	1		3.77	2.19		
4	6		7.12	4.13		
5	5		10.75	6.23		0.24
6	8		13.53	7.85		0.003
7	13		14.60	8.47		2.43
8	10		13.78	7.99		0.51
9	2		11.56	6.71		3.30
10	4		8.73	5.06		0.22
11	4		5.99	3.48		
12	2		3.77	2.19		
13	2		2.19	1.27		
14	0	9	1.18	0.69	8.25	0.07
15	1		0.60	0.35		
16	0		0.28	0.16		
> 17	0		0.21	0.12		
	$\sum 58$		$\sum 100$	$\sum 58$		$\chi^2 = \sum 6.80$

om in this
Table 8.3
robability
0.5, there
clude that
son parent

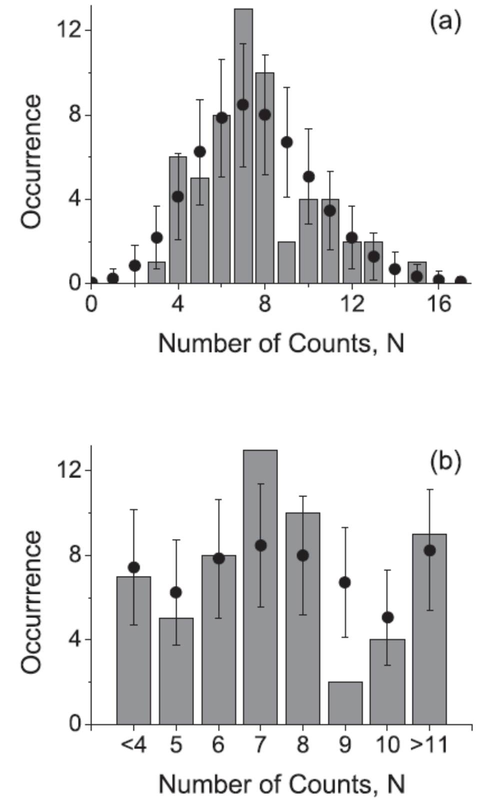


Fig. 8.8 Comparing experimental radioactive decay events with a Poisson distribution. In (a) a histogram of the observed occurrence of counts is plotted (bars) and compared with a Poisson distribution (points) and expected fluctuations. Part (b) shows a re-binned histogram with bin-widths chosen such that the number of expected occurrences is always greater than 5. Note that in the re-binned histogram six out of eight observed occurrences fall within the expected range taking into account the Poisson noise.

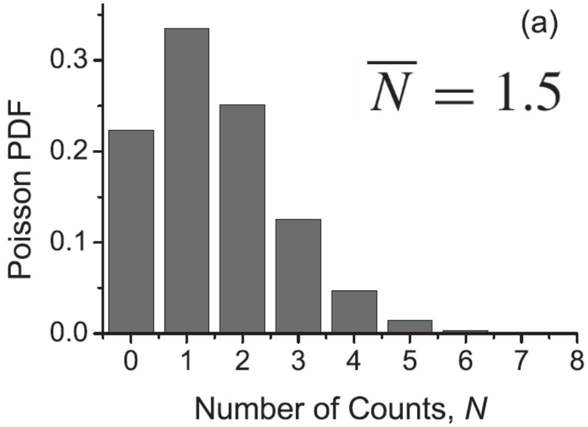
Poisson distribution

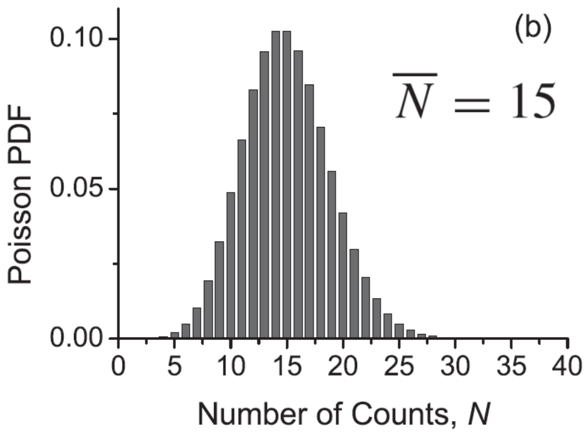
- **Counting rare events**
- **All events are independent**
- **Average rate does not change as a function of time**

$$P\left(N;\overline{N}\right)=\frac{\exp\left(-\overline{N}\right)\overline{N}^{N}}{N!}.$$

$\text{Mean} = \overline{N}$

$\text{Standard Deviation} = \sqrt{\overline{N}}$





χ^2 for data with Poisson errors

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

i is the number of counts (integer)

O_i is the observed number of occurrences (integer)

E_i is the expected number of occurrences

O_i is the observed number of occurrences (integer)

Note that χ^2 is (again) dimensionless

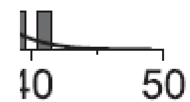
Measurements and their Uncertainties, Pg. 68

Worked example 4 – testing a continuous distribution

AtMol. Durham Atomic & Molecular Physics

Bin	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
<16	4	6.3	0.85
17–18	7	5.7	0.33
19–20	7	8.5	0.28
21	8	5.4	1.25
22	8	6.1	0.60
23	5	6.7	0.41
24	4	7.1	1.33
25	5	7.3	0.72
26	13	7.3	4.48
27	6	7.1	0.16
28	2	6.7	3.26
29	7	6.1	0.14
30	8	5.4	1.25
31–32	11	8.5	0.71
33–34	4	5.6	0.47
>35	7	6.3	0.08
$\sum$	**106**	**106**	**16.32**

Table 8.4 Comparing the distribution of the means of the 106 UK National Lottery draws with a Gaussian distribution. The first column is the mean value, the second is the occurrence of each mean. Column three gives the expected number of occurrences using a Gaussian model with the same mean and standard deviation as the sample data, and χ^2 is calculated in the final column. Bin-widths have been chosen such that E_i is greater than 5 in each bin. Note that all the entries in the last column are of order 1.



ns of 106
h draw six
1–49 and
lue of the
balls drawn per event. The continuous line is a Gaussian constrained to have the same mean and standard deviation as the sample data and is scaled by the total number of events.

Electron EDM measurement

In total, we collected 1024 blocks (360.3 hours) of

Selected for a Viewpoint in *Physics*

PRL **119,** 153001 (2017) PHYSICAL REVIEW LETTERS week ending 13 OCTOBER 2017

Precision Measurement of the Electron's Electric Dipole Moment Using Trapped Molecular Ions

William B. Cairncross,[*] Daniel N. Gresh, Matt Grau,[†] Kevin C. Cossel,[‡] Tanya S. Roussy, Yiqi Ni,[§] Yan Zhou, Jun Ye, and Eric A. Cornell

JILA, NIST and University of Colorado, Boulder, Colorado 80309-0440, USA
and Department of Physics, University of Colorado, Boulder, Colorado 80309-0440, USA

(Received 21 April 2017; published 9 October 2017)

We describe the first precision measurement of the electron's electric dipole moment (d_e) using trapped molecular ions, demonstrating the application of spin interrogation times over 700 ms to achieve high sensitivity and stringent rejection of systematic errors. Through electron spin resonance spectroscopy on $^{180}\mathrm{Hf}^{19}\mathrm{F}^{+}$ in its metastable $^3\Delta_1$ electronic state, we obtain $d_e = (0.9 \pm 7.7_{\mathrm{stat}} \pm 1.7_{\mathrm{syst}}) \times 10^{-29}$ e cm, resulting in an upper bound of $|d_e| < 1.3 \times 10^{-28}$ e cm (90% confidence). Our result provides independent confirmation of the current upper bound of $|d_e| < 9.4 \times 10^{-29}$ e cm [J. Baron *et al.*, New J. Phys. **19**, 073029 (2017)], and offers the potential to improve on this limit in the near future.

DOI: 10.1103/PhysRevLett.119.153001

The reduced chi-squared statistic for fitting a weighted mean to the eEDM data set is $\chi_r^2 = 1.22(5)$. This overscatter is present in all frequency channels, and is attributable to

The "First-Digit Phenomenon"

$$P(d) = \log_{10}\left(1 + \frac{1}{d}\right)$$

- Simon Newcomb (1881) & Frank Benford (1938)

- Base- and Scale-Invariant
- Many, many examples in nature
  - Seemingly arbitrary, unconnected data sets
  - Some exceptions

Occam’s Razor

A Franciscan friar ✓

William of Occam

entia non sunt multiplicanda praeter necessitatem

(entities must not be multiplied beyond necessity)

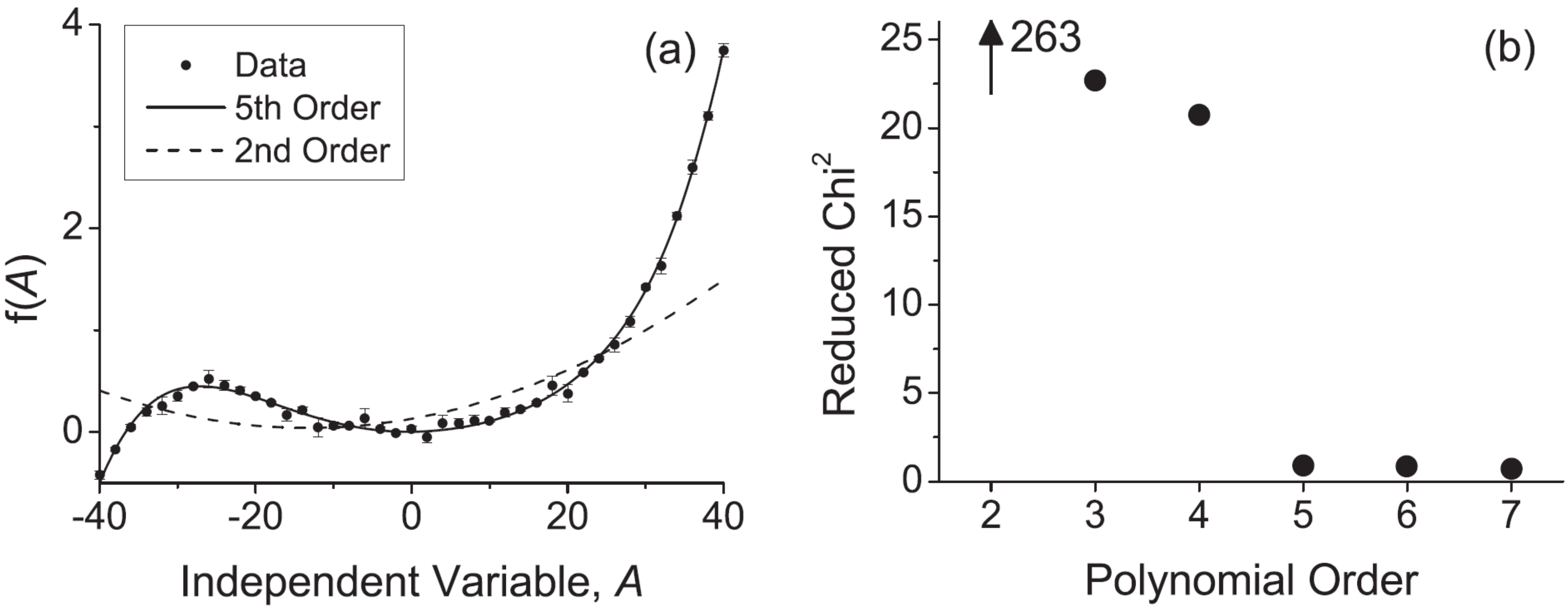


Fig. 8.10 Different polynomial fits to a data set with 41 points. Low-order polynomial fits such as the second-order fit shown in (a) systematically fail to account for the trends in the data set. Higher order polynomials fit the data set better. The evolution in the quality of fit, χ^2_ν, is shown as a function of polynomial order in (b). Fits with polynomial orders greater than five do not significantly improve the quality of the fit.

what comes next in this sequence: 2; 4; 6; 8;

Measurements and their Uncertainties, P. 114

What do we do if our value of χ^2_{min} >1?

There are two options frequently used:

- Use the fit to extract the common uncertainty
- **Scale the Uncertainties**. Be very careful!

common uncertainty, α_C, in the measurements by setting $\chi^2_\nu = 1$. If the data set is homoscedastic one can use eqn (6.1), to give

It is also possible to use the value of $\chi^2_{\min}$ to scale the errors in the fit parameters when one has complete confidence in the theoretical model being used to describe the experimental data. Let S be the **scale factor** defined as

$$S = \sqrt{\frac{\chi^2_{\min}}{\nu}} = \sqrt{\chi^2_\nu}. \tag{8.9}$$

Amsler, C. *et al.* (Particle Data Group) (2008). *Phys. Lett.*, **B667**, 1–1340.

Measurements and their Uncertainties, P. 116

Scaling uncertainties

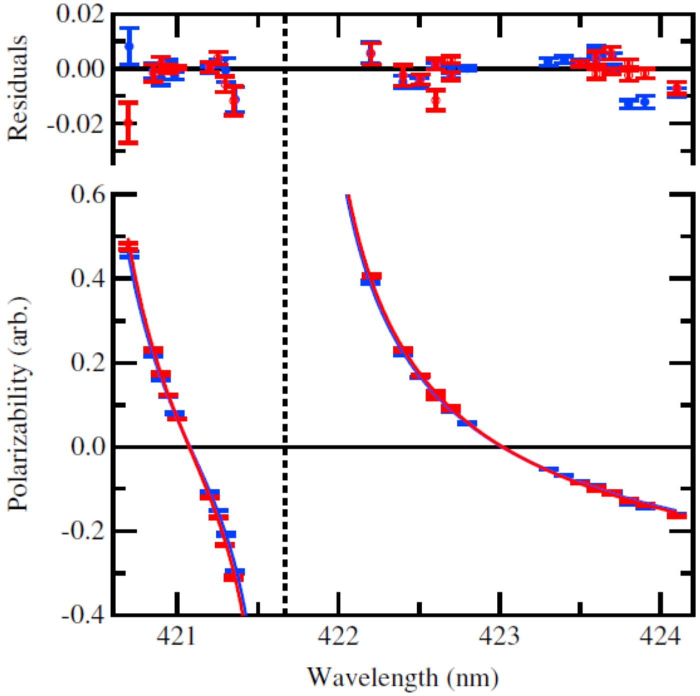


FIG. 4 (color online). Measured polarizability (arbitrary units) versus wavelength for two linear, orthogonal input polarizations S (red, open circles) and P (blue, filled circles). Each point is an average of up to 20 measurements, shown with 1σ statistical error bars. The solid lines (nearly indistinguishable) are fits to the expected form [Eq. (1)] with the $6p$ matrix elements as free parameters; reduced χ^2 for the S and P fits is 4 and 11, respectively [37]. Additionally, we allow a separate amplitude about each zero to account for different laser power. Fit residuals are shown at the top.

RS week ending 14 DECEMBER 2012

a Light Shift Cancellation

ıd J. V. Porto

ırk, Maryland 20742, USA

ırk, Delaware 19716, USA

r 2012)

That reduced χ^2 for the fits in Fig. 4 is larger than 1 is accounted for by the additional uncertainty in alignment drift, which we expect also explains the disparity between S and P fits; rescaling the statistical uncertainty to give a reduced χ^2 of 1 produces similar total uncertainty.

Scaling uncertainties

throughout this thesis uncertainties have been extracted via a chi-squared (χ^2) minimisation, as described in detail in [238], using the graphing software *Origin*. The author strongly recommends that the option to `Scale Errors with sqrt(reduced chi^2)` is turned off when fitting using *Origin*. Beyond this it is still essential to check the magnitude of the errors and, using a common sense approach, decide whether they have been significantly underestimated or overestimated. Where this has occurred within *Origin*, the author has undertaken a manual χ^2 minimisation using *Microsoft Excel* or *MATLAB*.

"…error analysis is a participation, rather than a spectator, sport."

Please submit homework before 2pm Monday

11th November. This is SUMMATIVE.

For MiSCaDA students via
https://notebooks.dmaitre.phyip3.dur.ac.uk/miscada-da/hub/login

For PhD students either the notebook server OR email me a document