Class 3: Model-fitting

In this class we will describe the use of the χ^2 statistic as a hypothesis test of a model, and in determining best-fitting parameters

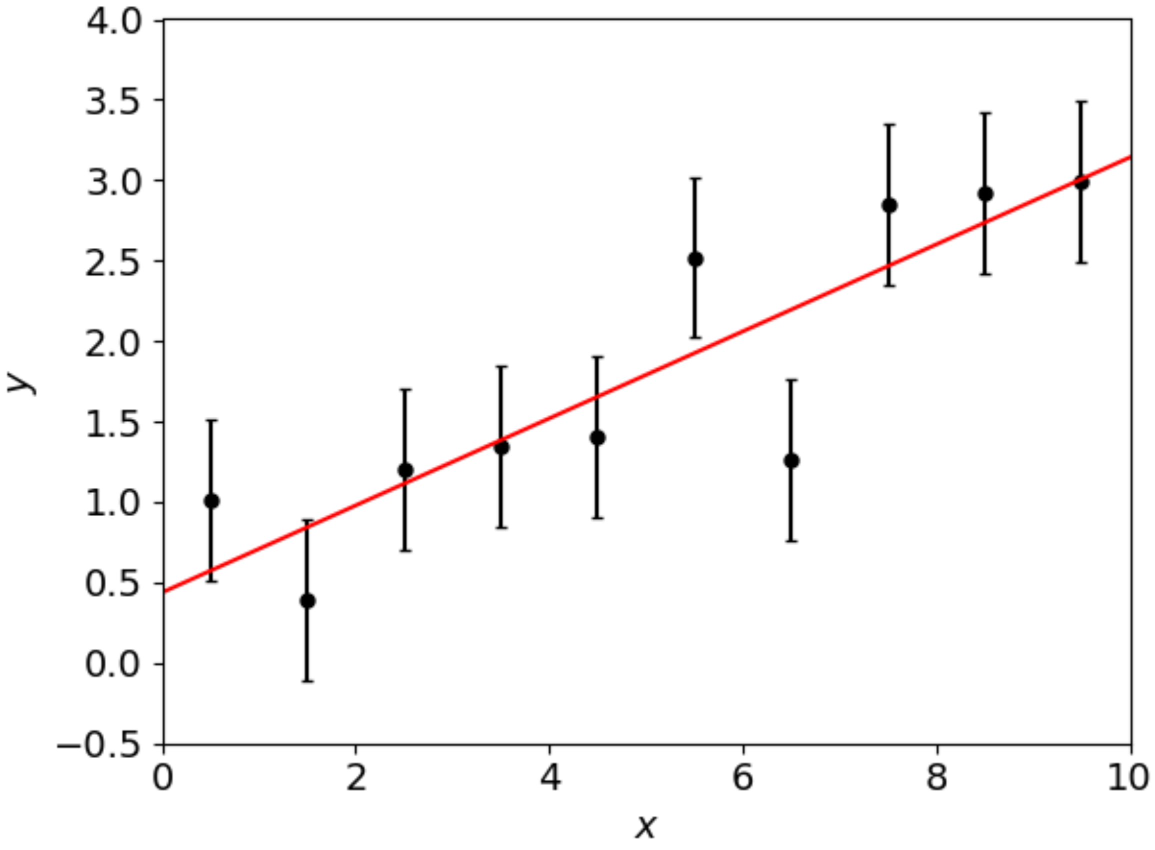
Class 3: Model-fitting

At the end of this class you should be able to …

- … apply the χ^2 statistic as a hypothesis test
- … understand the probability distribution of the χ^2 statistic and its interpretation as a p-value
- … apply the χ^2 statistic in parameter fitting
- … determine parameter errors and joint confidence regions using intervals of $\Delta\chi^2$

Comparing data and models

- A key task in statistics is to build models which describe our data:



Comparing data and models

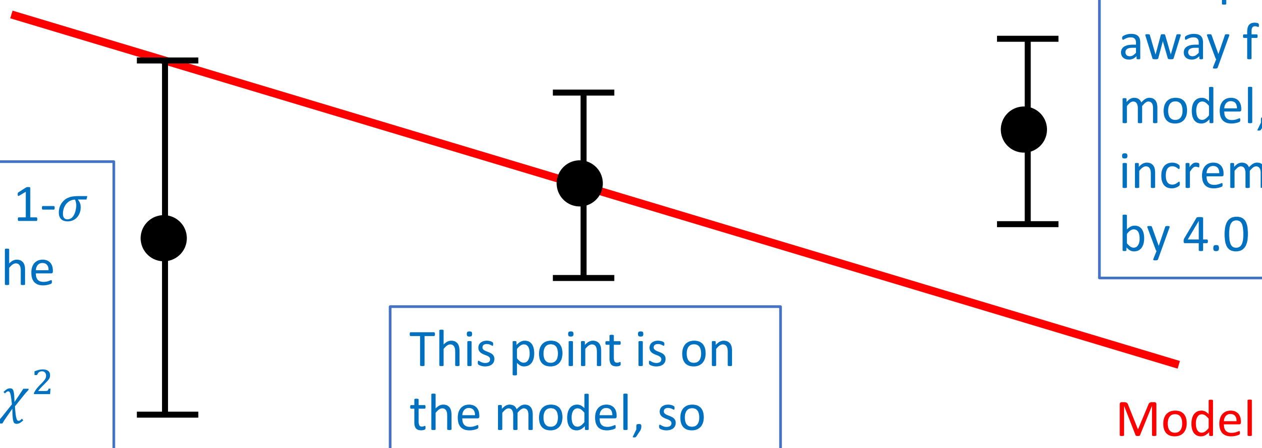
- *When comparing data and models, we are typically doing one of two things …*
- **Hypothesis testing**: we have a set of N measurements $x_i \pm \sigma_i$, which a theorist says should have values μ_i. *How probable is it that these measurements would have been obtained, if the theory is correct?*
- **Parameter estimation**: we have a parameterized model which describes the data, such as $y = ax + b$. *What are the best-fitting parameters and errors in those parameters?*

The χ^2 statistic

- The most important statistic to help with these tasks is the $\boldsymbol{\chi^2}$ **statistic** between the data $x_i \pm \sigma_i$ and model μ_i:

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

χ^2 is a sum over the data points



This point is 1-σ away from the model, so increments χ^2 by 1.0

This point is on the model, so increments χ^2 by 0.0

This point is 2-σ away from the model, so increments χ^2 by 4.0

The χ^2 statistic

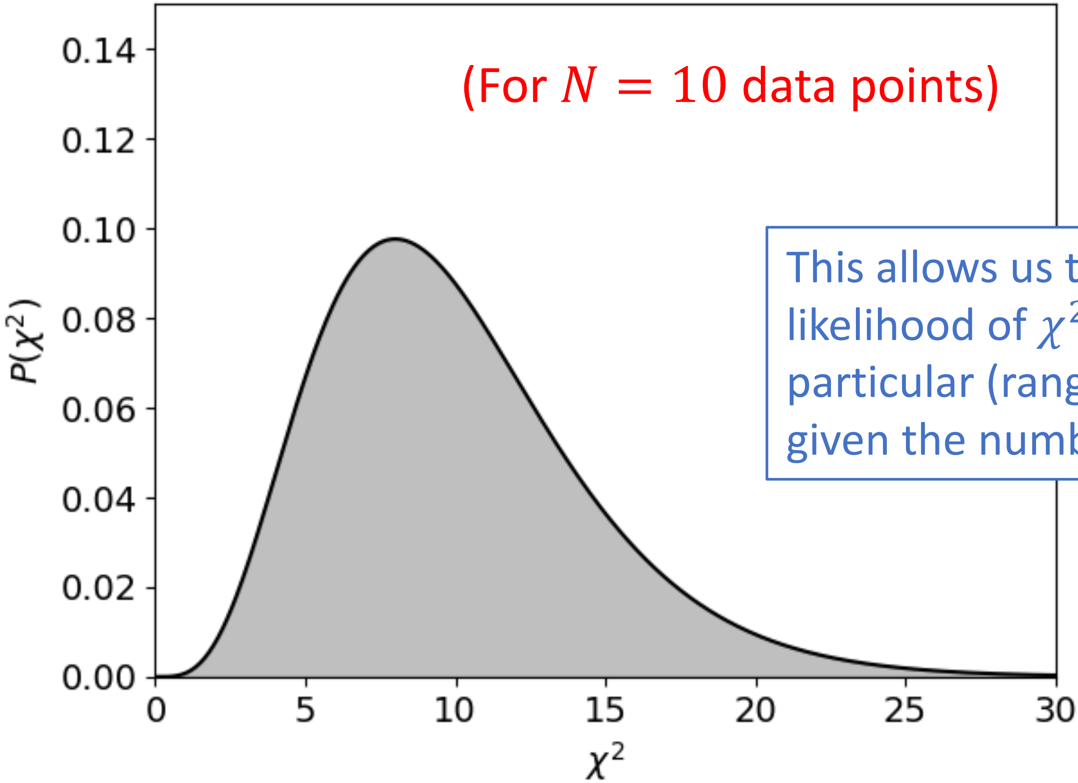
- The most important statistic to help with these tasks is the $\boldsymbol{\chi^2}$ **statistic** between the data $x_i \pm \sigma_i$ and model μ_i:

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

- We accumulate the statistic according to **how many standard deviations** each data point lies from the model
- χ^2 is a measure of the **goodness-of-fit** of the data to the model
- If the data are numbers taken as part of a **counting experiment**, we could use the Poisson error ${\sigma_i}^2 = \mu_i$
- [Small print: this equation assumes the data points are independent]

χ^2 probability distribution

- Sampling many realizations of N data points from a particular model, using Gaussian statistics, the χ^2 statistic has a **probability distribution**:



This allows us to quantify the likelihood of χ^2 taking on a particular (range of) values, given the number of data points

χ^2 probability distribution

- Sampling many realizations of N data points from a particular model, using Gaussian statistics, the χ^2 statistic has a **probability distribution**

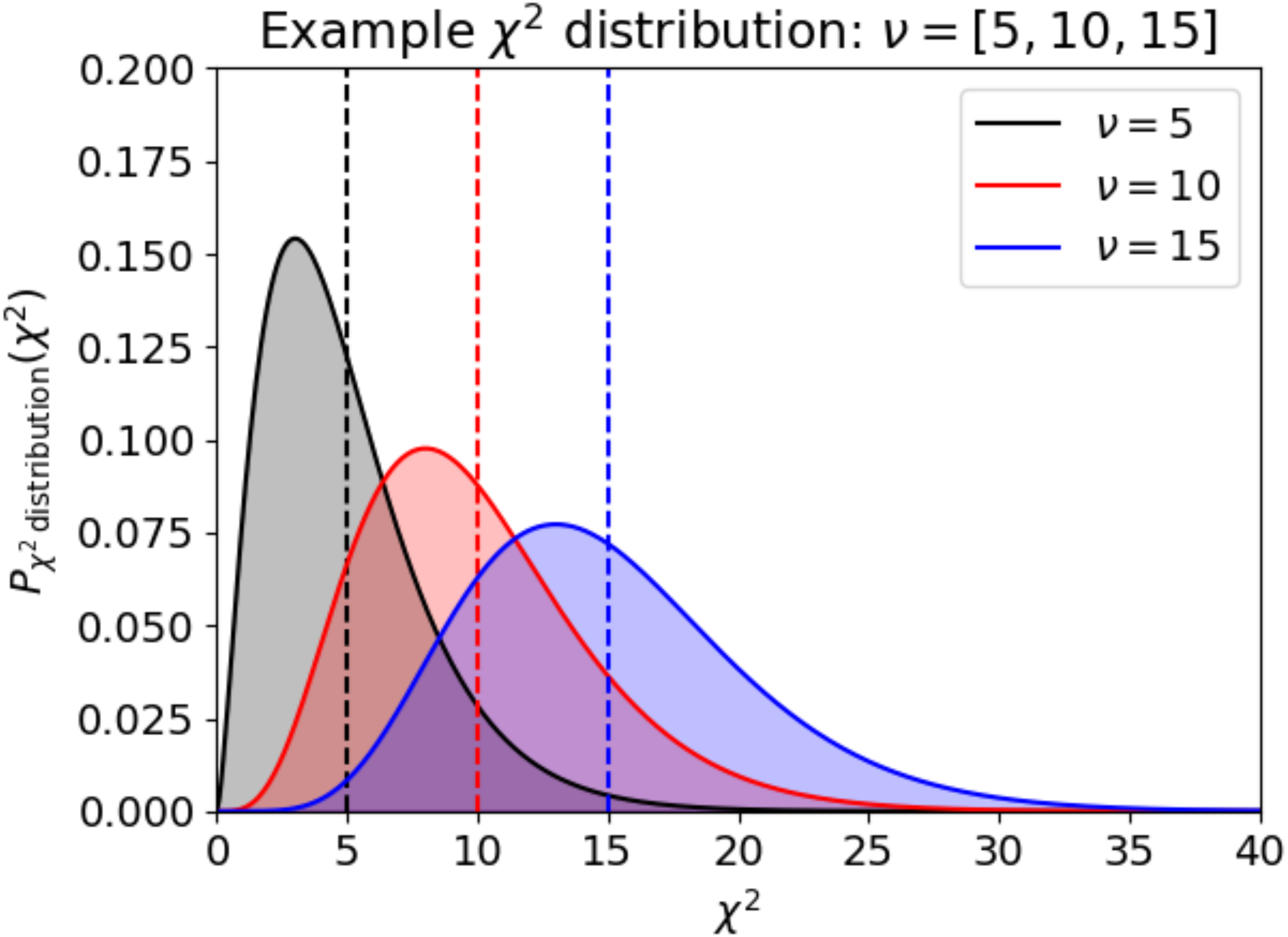
$$\boxed{P(\chi^2) \propto (\chi^2)^{\frac{\nu}{2}-1} e^{-\chi^2/2}}$$

- ν is the number of **degrees of freedom**

- If the model has no free parameters, then $\nu = N$

- If we are fitting a model with p free parameters, we can "force the model to agree exactly with p data points" and the degrees of freedom are reduced to $\nu = N - p$

χ^2 probability distribution



χ^2 probability distribution

- The χ^2 distribution:

$$\boxed{P(\chi^2) \propto (\chi^2)^{\frac{\nu}{2}-1} e^{-\chi^2/2}}$$

- The **mean** of the distribution: $\overline{\chi^2} = \nu = N - p$
- *This makes intuitive sense, because each data point should lie about 1σ from the model and hence contribute 1.0 to the χ^2 statistic*
- The **variance**: $\mathrm{Var}(\chi^2) = 2\nu$
- If the model is correct, we expect $\chi^2 \sim \nu \pm \sqrt{2\nu}$

Reduced χ^2

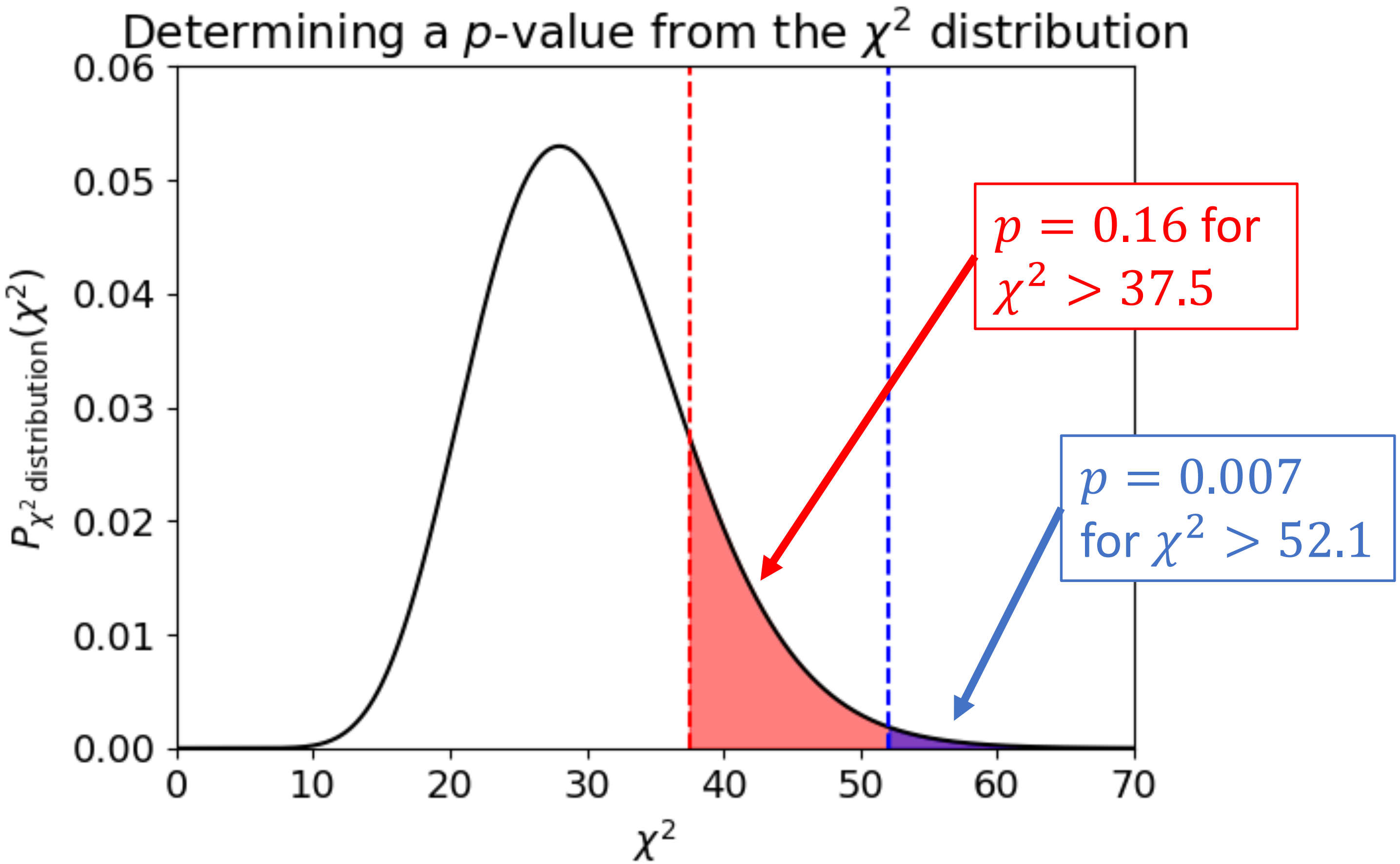
- As a way of summarizing the model fit, we can quote the **reduced χ^2 statistic**, ${\chi_r}^2 = \chi^2/\nu$
- For a good fit, ${\chi_r}^2 \sim 1$ (because $\overline{\chi^2} = \nu$)
- However, the true probability of the data being consistent with the model depends on **both** χ^2 and ν
- *Do not just quote the reduced χ^2 value*

Use of χ^2 statistic as a hypothesis test

- *We can use the χ^2 statistic to construct a hypothesis test describing the "goodness of fit" between data and model*
- **Null hypothesis**: the data are consistent with the model
- **Test statistic**: χ^2
- **Distribution of values**: The χ^2 probability distribution
- **Confidence statement**: What is the probability that this value of χ^2, or a larger one, could arise by chance?
- If the **p-value is not low**, the data are consistent with the model, which is *"ruled in"*
- If the **p-value is low**, the model is *"ruled out"*

Use of χ^2 statistic as a hypothesis test

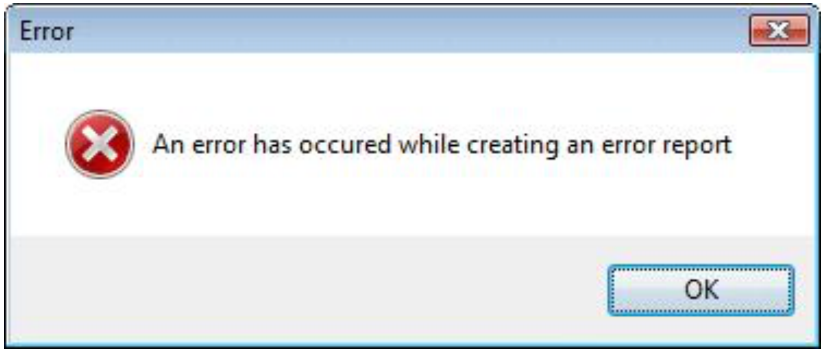
- *Suppose that* $\nu = 30$ *and we have 2 datasets with* $\chi^2 = 37.5$ *and* $\chi^2 = 52.1$. *What are the corresponding p-values?*



Cautionary words

- *Let's recall our discussion in Class 2 on the meaning of p*
- Suppose a χ^2 hypothesis test yields $p = 0.01$
- This means: **there is a 1% chance of obtaining a set of measurements at least this discrepant from the model, assuming the model is true**. It does not mean:
- "the probability that the model is true is 1%"
- "the probability that the model is false is 99%"
- "if we reject the model there is a 1% chance that we would be mistaken"
- **Frequentist statistics cannot assess the probability that the model itself is correct**

Cautionary words



- When using χ^2 we're assuming that the errors in the data are **Gaussian** and **reliable**. *This is not guaranteed!*
- If the errors have been **under-estimated**, then an **improbably high** value of χ^2 can be obtained
- If the errors have been **over-estimated**, then an **improbably low** value of χ^2 can be obtained
- Since errors are often approximate, a model is typically only rejected for **very low** values of p such as 0.001

Cautionary words

- An issue in using the χ^2 statistic is **binning of data**
- For example, suppose we have a sample of galaxy luminosities. To compare the data with a Schechter function model, we would bin it into a luminosity function
- Warning: if **the numbers in each bin are too small** the probabilities can become non-Gaussian
- As a rule of thumb, 80% of bins must have $N > 5$

Modification for correlated data

- If the data are **correlated**, the χ^2 equation must be modified:

$$\chi^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} (d_i - m_i)\, (C^{-1})_{ij}\, \left(d_j - m_j\right) = (\mathbf{d} - \mathbf{m})^T \mathbf{C}^{-1} (\mathbf{d} - \mathbf{m})$$

- Here, C is the covariance matrix of the data, such that

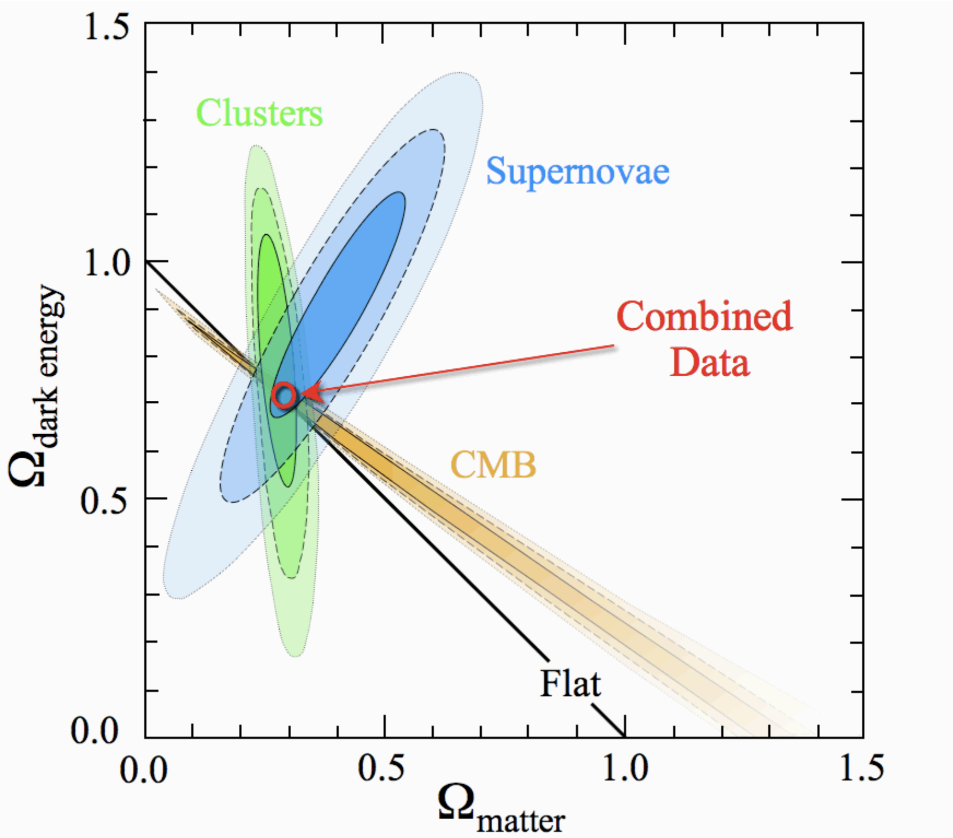
$$C_{ij} = \left\langle x_i\, x_j \right\rangle - \langle x_i \rangle \left\langle x_j \right\rangle$$

- Note that $C_{ii} = \langle {x_i}^2 \rangle - \langle x_i \rangle^2$ is the variance ${\sigma_i}^2$

- The **number of degrees of freedom is unchanged** (for anything less than complete correlation)

Use of χ^2 statistic for parameter fitting

- *A model typically contains free parameters. How do we determine the most likely values of these parameters and their error ranges?*



Use of χ^2 statistic for parameter fitting

- *A model typically contains free parameters. How do we determine the most likely values of these parameters and their error ranges?*

- Suppose we are fitting a model with 2 free parameters (a, b)

- The most likely ("best-fitting") values of (a, b) are found by **minimizing the χ^2 statistic**

- The joint error distribution of (a, b) can be found by **calculating the values of χ^2 over a grid** of (a, b) and enclosing a particular region $\chi^2 < \chi^2_{\min} + \Delta\chi^2$

Joint confidence regions

- We plot 2D contours of constant $\chi^2 = \chi^2_{\min} + \Delta\chi^2$
- A **joint confidence region** for (a, b) can be defined by the zone which satisfies $\chi^2 < \chi^2_{\min} + \Delta\chi^2$
- The values of $\Delta\chi^2$ depend on the number of variables and confidence limits:

2 parameters varying

$\Delta\chi^2$ as a Function of Confidence Level and Degrees of Freedom

	ν					
p	1	2	3	4	5	6
68.3%	1.00	2.30	3.53	4.72	5.89	7.04
90%	2.71	4.61	6.25	7.78	9.24	10.6
95.4%	4.00	6.17	8.02	9.70	11.3	12.8
99%	6.63	9.21	11.3	13.3	15.1	16.8
99.73%	9.00	11.8	14.2	16.3	18.2	20.1
99.99%	15.1	18.4	21.1	23.5	25.7	27.8

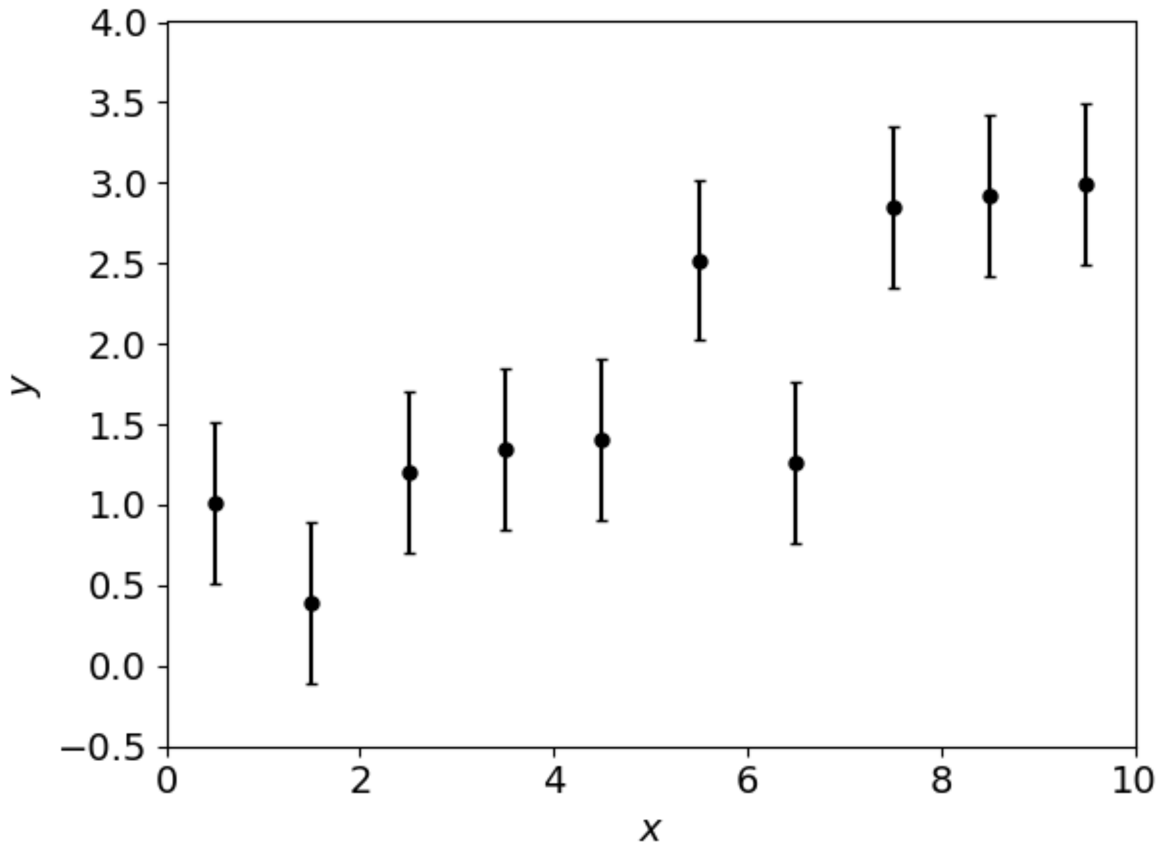
1-σ confidence

(Table taken from *Numerical Recipes* Chapter 15)

- [Small print: assumes the variables are Gaussian-distributed]

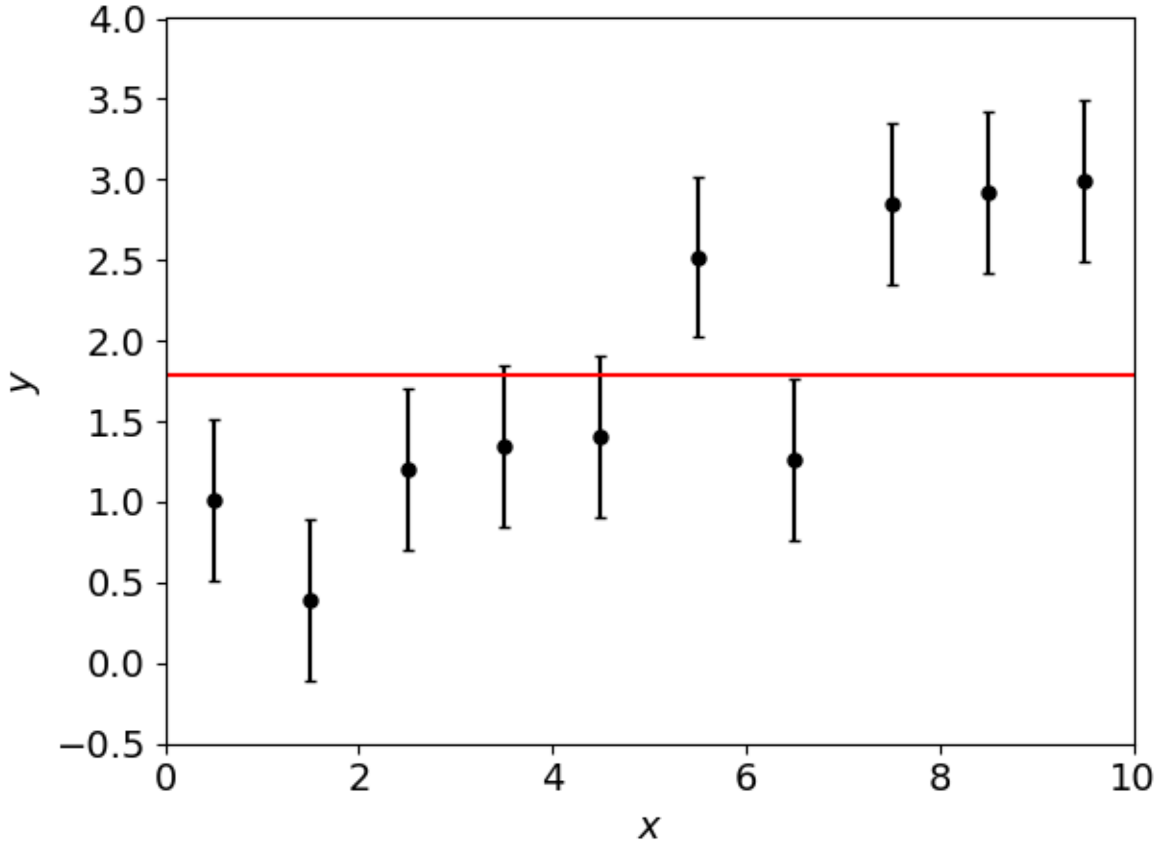
Use of χ^2 statistic for parameter fitting

- Here is an example dataset containing $N = 10$ points:



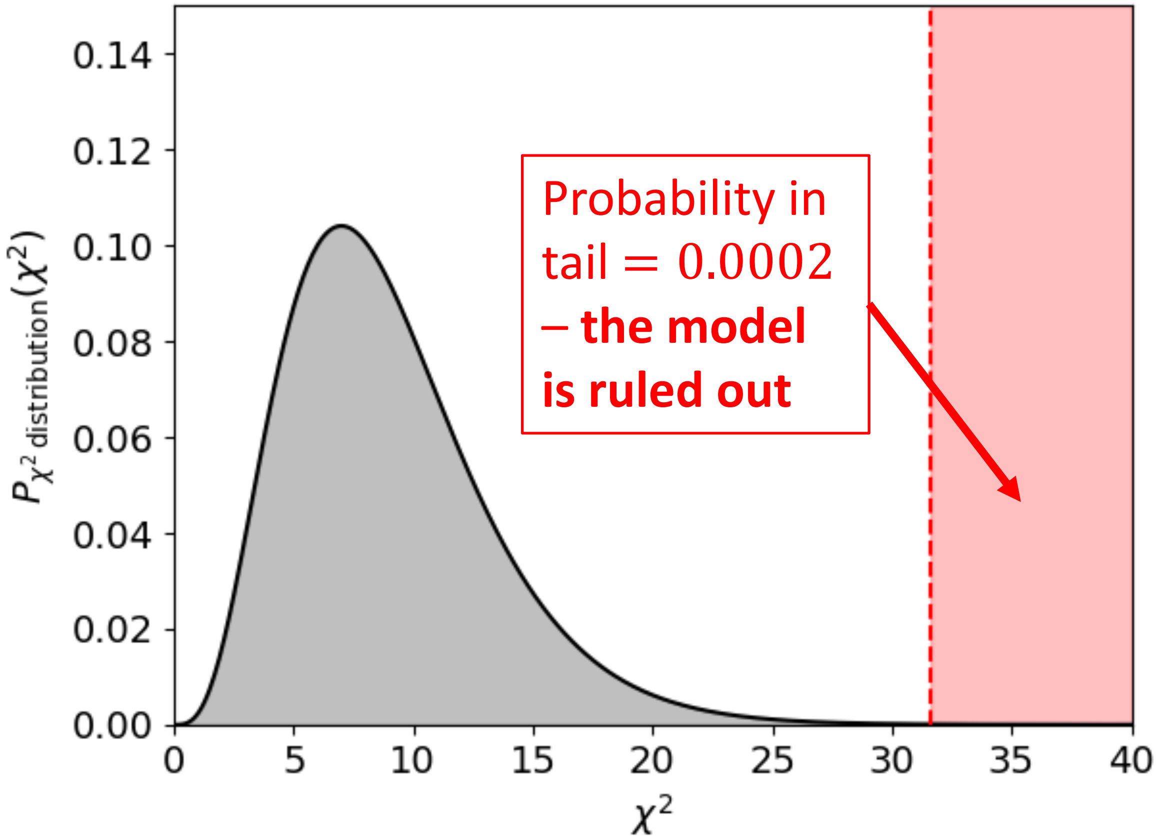
Use of χ^2 statistic for parameter fitting

- *Could these points be fit by a constant $y = b$?* Minimizing χ^2, we find $\chi^2_{\min} = 31.6$ for $b = 1.79$



Use of χ^2 statistic for parameter fitting

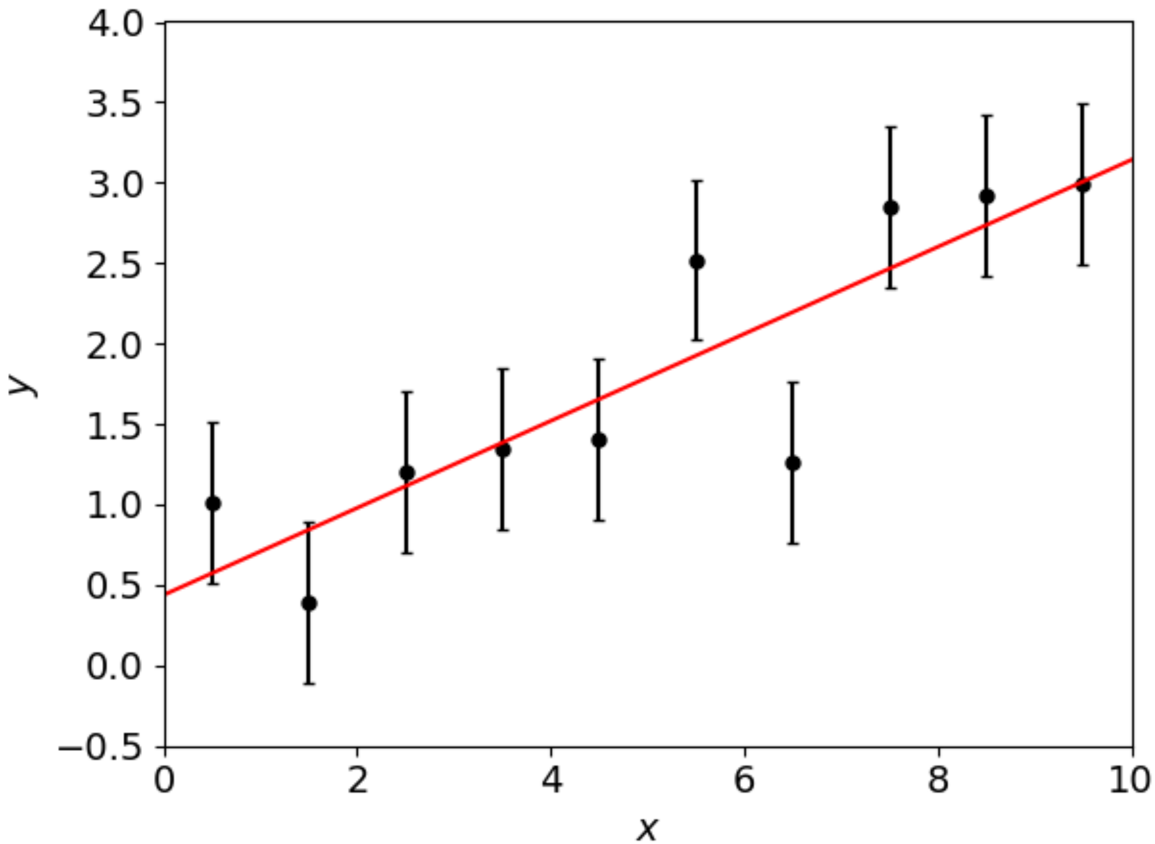
- *Is the minimum χ^2 likely given the model?* Consider the χ^2 probability distribution for $\chi^2 > 31.6$ and $\nu = N - 1 = 9$:



Use of χ^2 statistic for parameter fitting

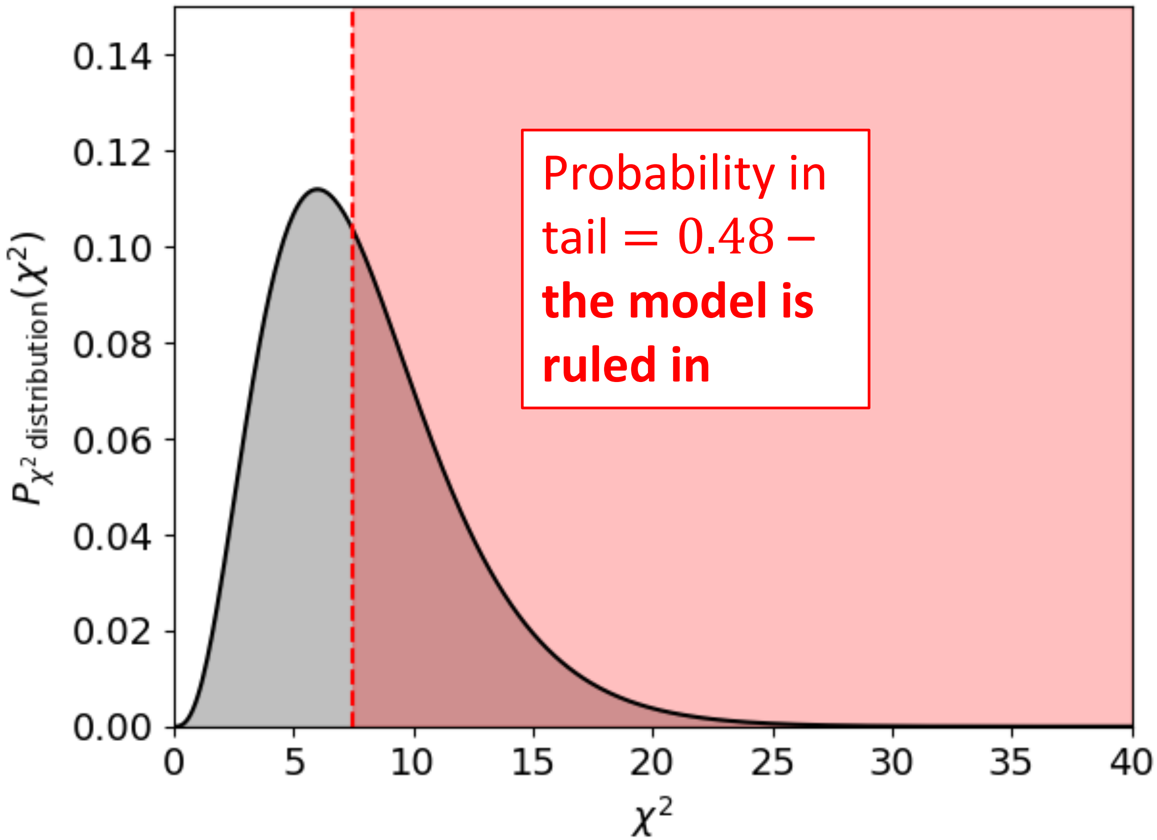
- *Could these points be fit by a straight line $y = ax + b$?*

  Minimizing χ^2, we find $\chi^2_{\min} = 7.5$ for $a = 0.27$ and $b = 0.44$



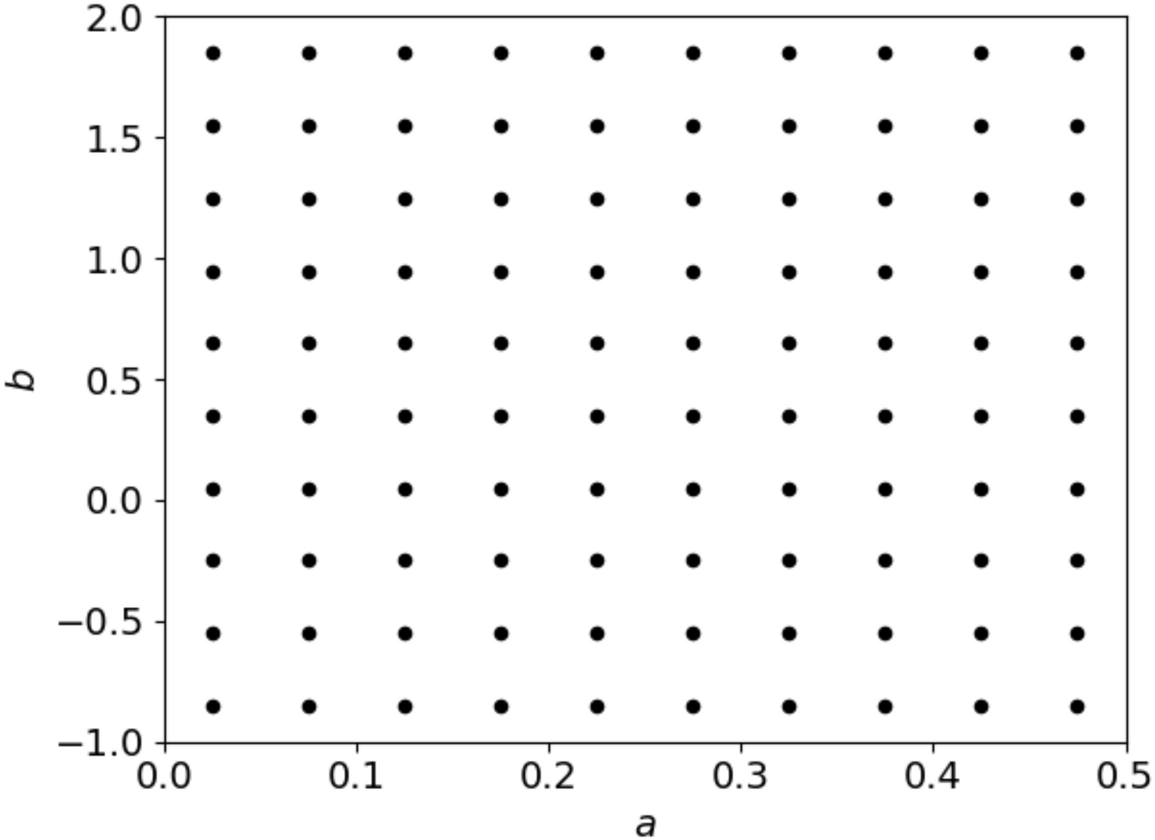
Use of χ^2 statistic for parameter fitting

- *Is the minimum χ^2 likely given the model?* Consider the χ^2 probability distribution for $\chi^2 > 7.5$ and $\nu = N - 2 = 8$:



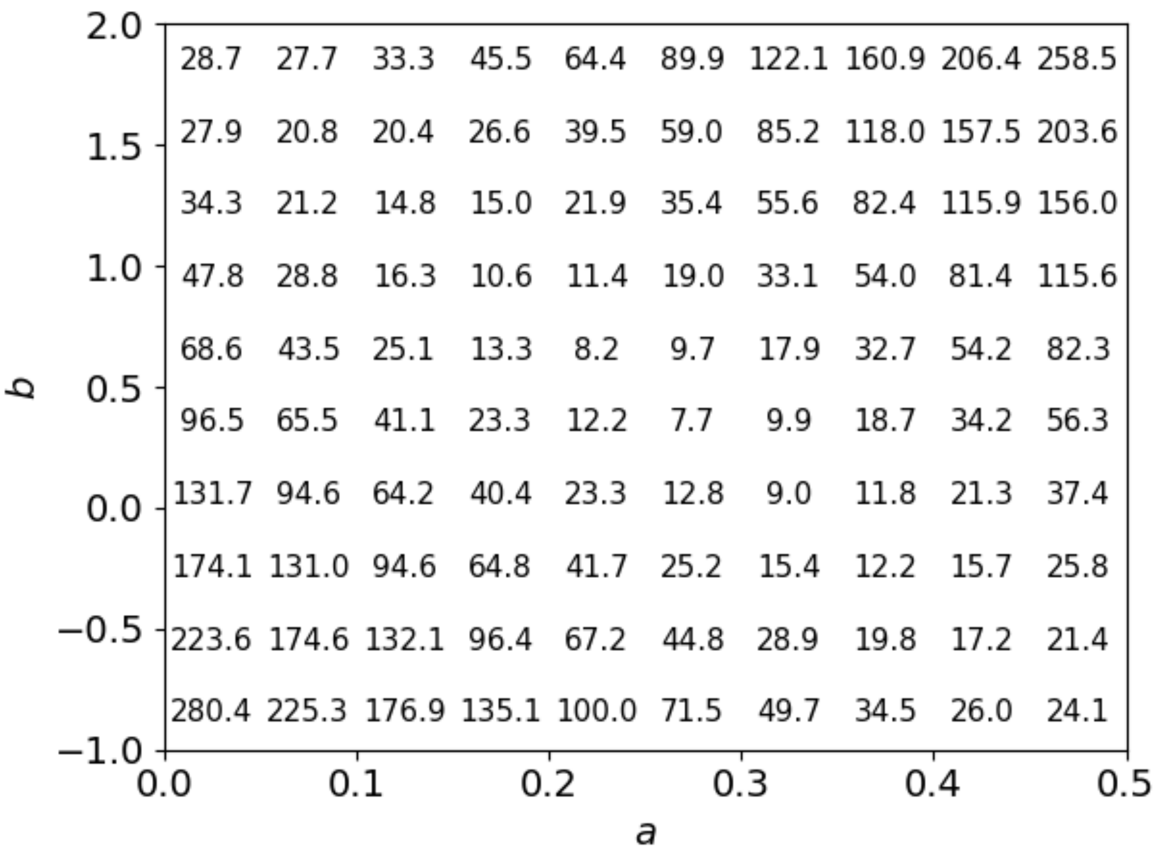
Use of χ^2 statistic for parameter fitting

- *Now let's determine the error ranges.* What is the distribution of χ^2 values across the (a, b) grid?



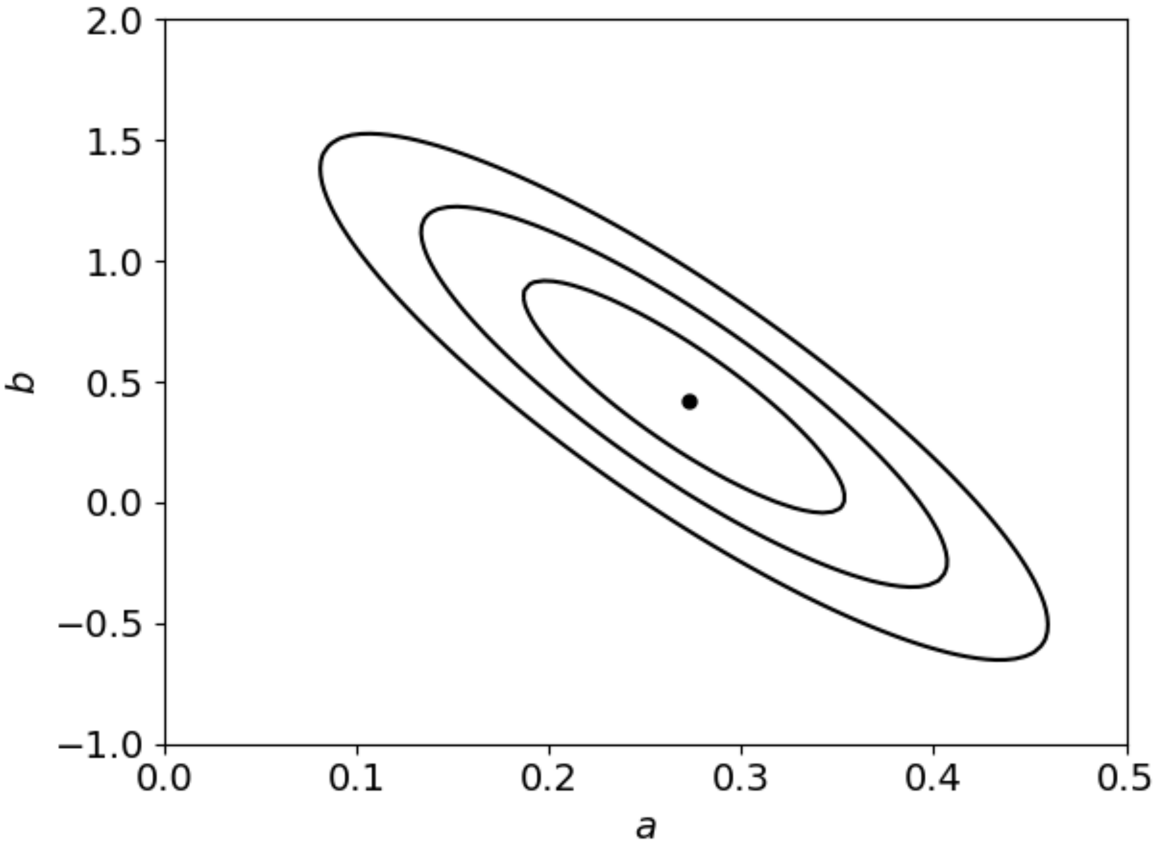
Use of χ^2 statistic for parameter fitting

- *Now let's determine the error ranges.* What is the distribution of χ^2 values across the (a, b) grid?



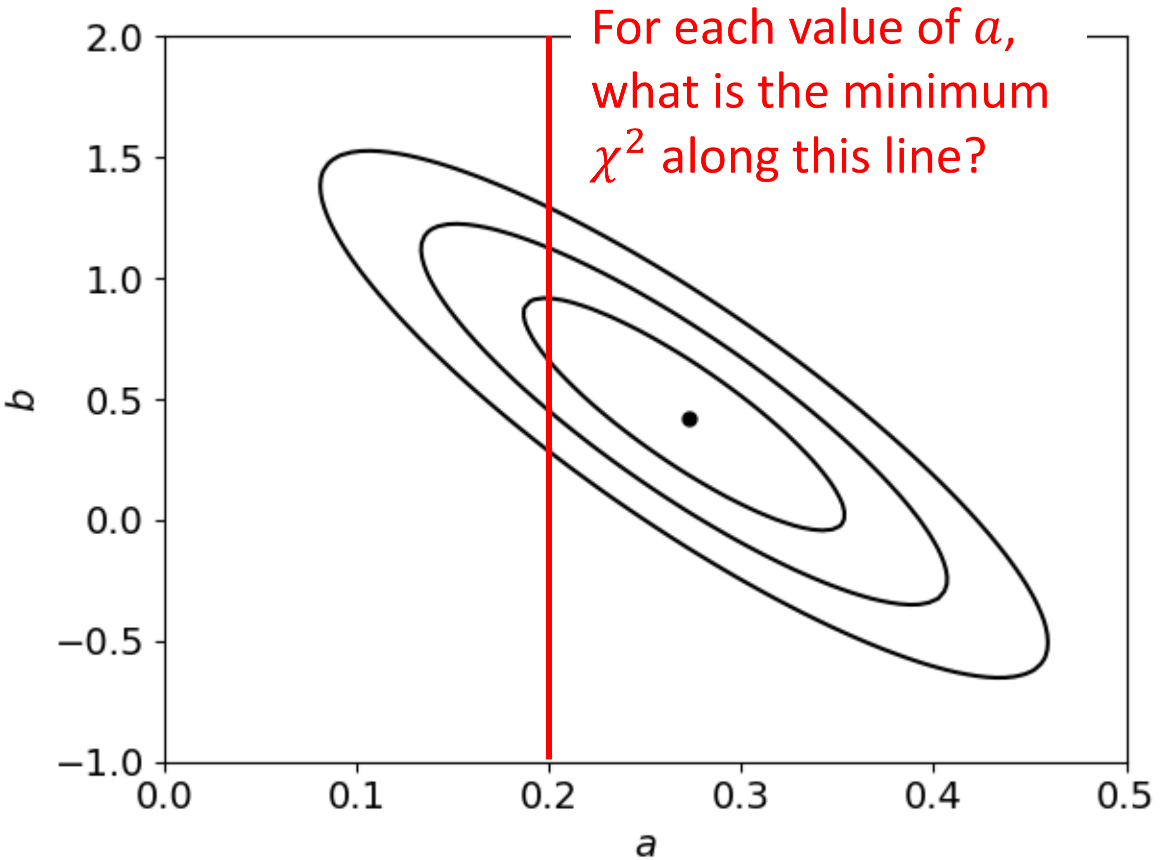
Use of χ^2 statistic for parameter fitting

- *Set confidence regions using $\Delta\chi^2$ intervals:* for 2 parameters, the $(68,95,99)\%$ regions are $\Delta\chi^2 = (2.30, 6.17, 11.8)$



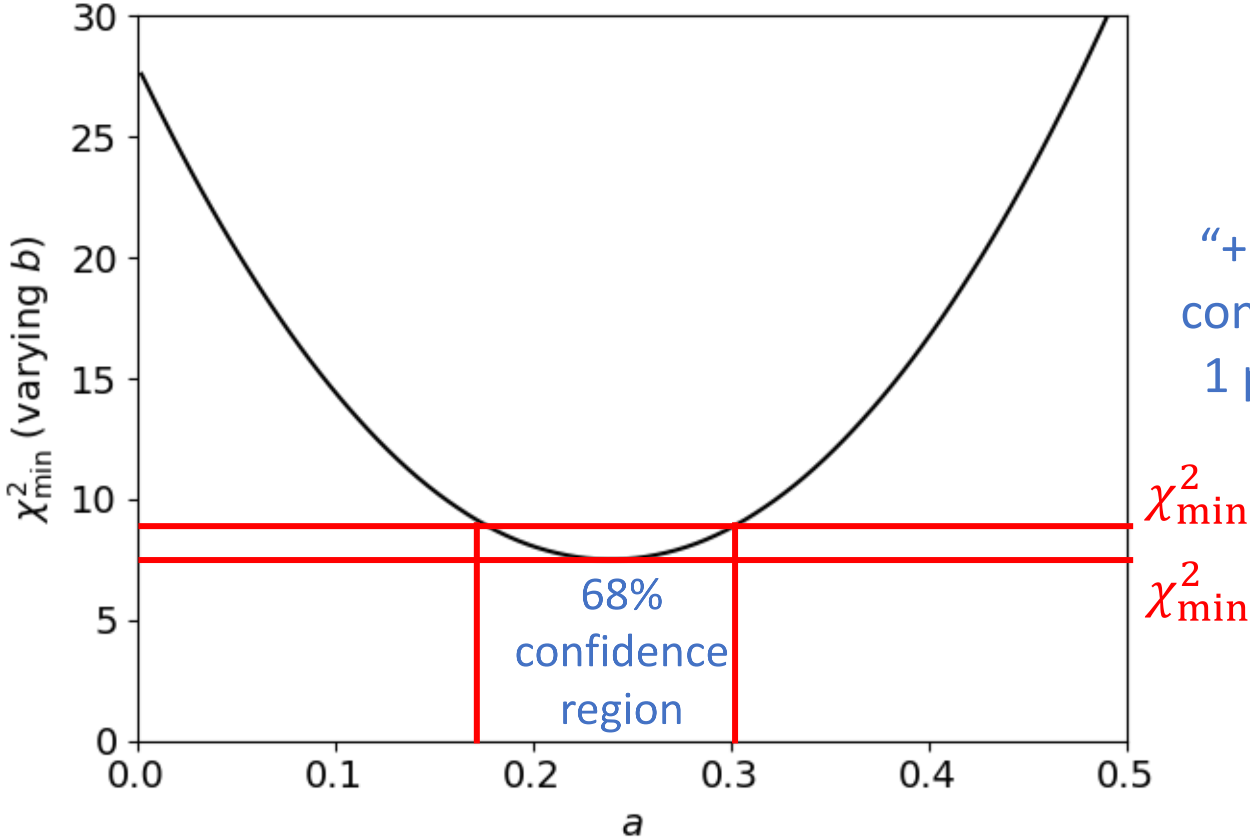
Use of χ^2 statistic for parameter fitting

- *What about errors in individual parameters?* For each value of parameter a, find the minimum χ^2 varying parameter b:



Use of χ^2 statistic for parameter fitting

- *What about errors in individual parameters?* For each value of parameter a, find the minimum χ^2 varying parameter b:



"+1" for 68% confidence for 1 parameter

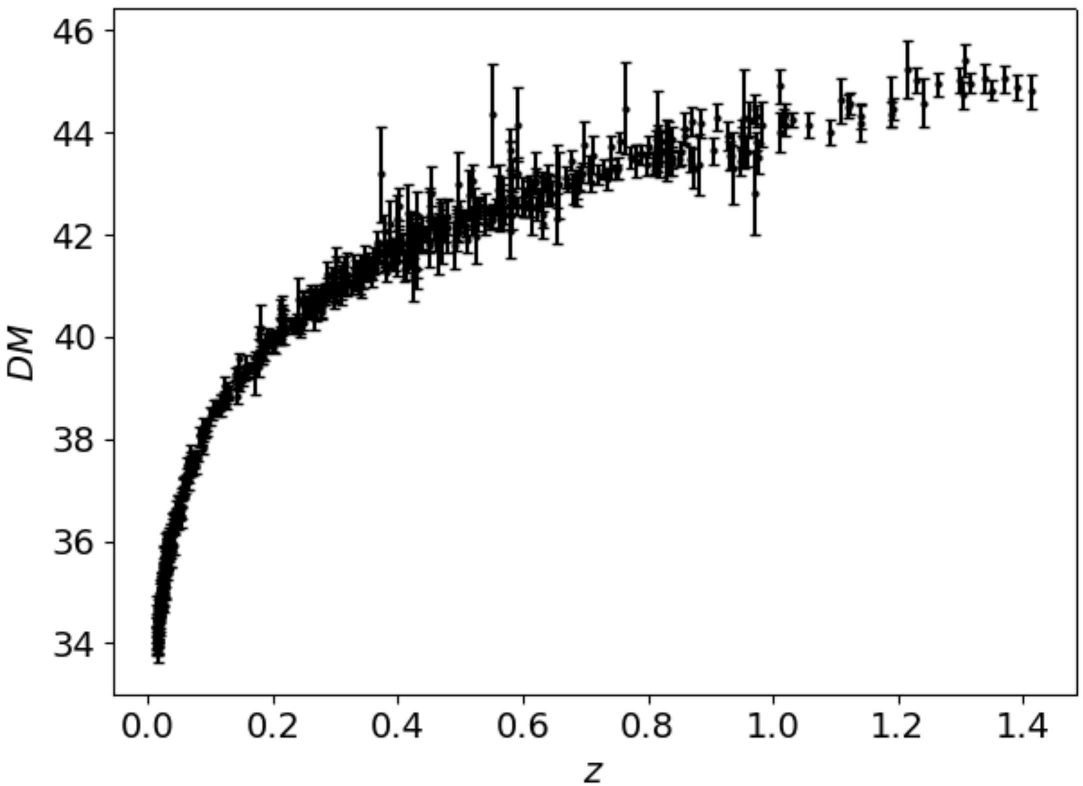
Supernova cosmology

- In this Activity we will use a recent **supernova distance-redshift dataset** to determine the parameters $(\Omega_m, \Omega_\Lambda)$

- Minimize χ^2 and find the best-fitting $(\Omega_m, \Omega_\Lambda)$

- Construct the 2D confidence regions in $(\Omega_m, \Omega_\Lambda)$

- Determine the individual errors in Ω_m and Ω_Λ



Summary

At the end of this class you should be able to …

- … apply the χ^2 statistic as a hypothesis test
- … understand the probability distribution of the χ^2 statistic and its interpretation as a p-value
- … apply the χ^2 statistic in parameter fitting
- … determine parameter errors and joint confidence regions using intervals of $\Delta\chi^2$