Linear Algebra Solution to Straight line fit

As an example of a more realistic data-driven analysis, let's consider a simple three-parameter linear model which fits a straight-line to data with unknown errors. The parameters will be the the y-intercept α, the slope β, and the (unknown) normal scatter σ about the line.

For data $D = \{x_i, y_i\}$, the model is

$$\hat{y}(x_i|\alpha,\beta) = \alpha + \beta x_i,$$

and the likelihood is the product of the Gaussian distribution for each point:

$$\mathscr{L}(D|\alpha,\beta,\sigma) = (2\pi\sigma^2)^{-N/2} \prod_{i=1}^{N} \exp\left[\frac{-[y_i - \hat{y}(x_i|\alpha,\beta)]^2}{2\sigma^2}\right].$$

arXiv:1411.5018
Jake Van Der Plas
Frequentism and Bayesianism :
A Python Driven Primer

algebra. If we define the *parameter vector*, $\theta = [\alpha\ \beta]^T$; the *response vector*, $Y = [y_1\ y_2\ y_3\ \cdots\ y_N]^T$; and the *design matrix*,

$$X = \left[\begin{array}{ccccc} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_N \end{array} \right]^T,$$

it can be shown that the maximum likelihood solution is

$$\hat{\theta} = (X^T X)^{-1}(X^T Y).$$

The confidence interval around this value is an ellipse in parameter space defined by the following matrix:

$$\Sigma_{\hat{\theta}} \equiv \left[\begin{array}{cc} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{array} \right] = \sigma^2 (M^T M)^{-1}.$$

Here σ is our unknown error term; it can be estimated based on the variance of the residuals about the fit. The off-diagonal elements of $\Sigma_{\hat{\theta}}$ are the correlated uncertainty between the estimates. In code, the computation looks like this:

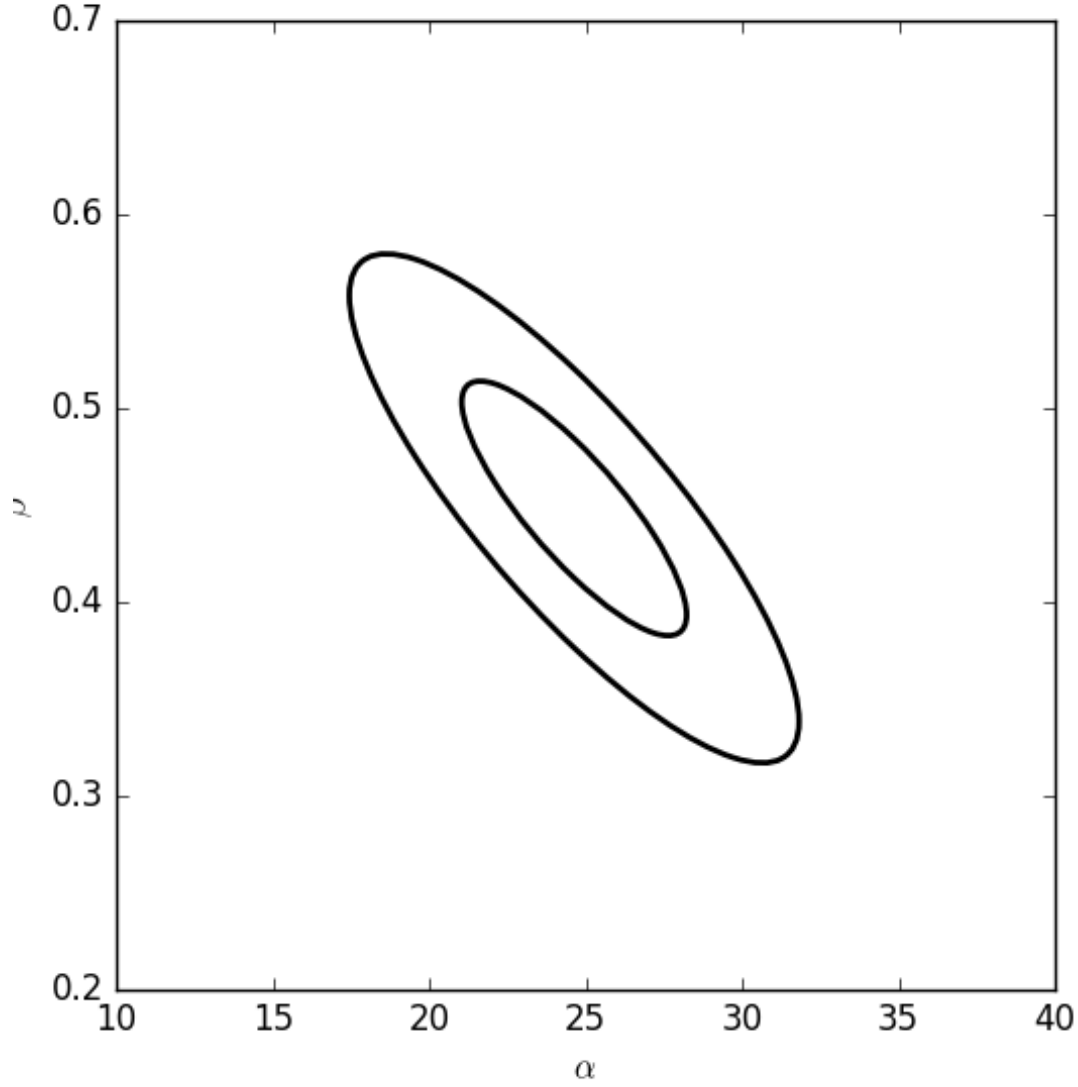
Code by Jake Van der Plas

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
from scipy import ndimage
np.random.seed(42)
theta_true = (25, 0.5)
xdata = 100 * np.random.random(20)
ydata = theta_true[0] + theta_true[1] * xdata
ydata = np.random.normal(ydata, 10)
# Compute the frequentist version
X = np.vstack([np.ones_like(xdata), xdata]).T
theta_freq = np.linalg.solve(np.dot(X.T, X),
                             np.dot(X.T, ydata))
y_model = np.dot(X, theta_freq)
sigma_y = np.std(ydata - y_model)
Sigma_freq = sigma_y ** 2 *
np.linalg.inv(np.dot(X.T, X))
```

Full source code available at https://tinyurl.com/y9hhded4

```python
def get_principal(Sigma):
    # See Ivezic, Connolly, VanderPlas, and Gray, section 3.5.2
    sigma_x2 = Sigma[0, 0]
    sigma_y2 = Sigma[1, 1]
    sigma_xy = Sigma[0, 1]

    alpha = 0.5 * np.arctan2(2 * sigma_xy, sigma_x2 - sigma_y2)
    tmp1 = 0.5 * (sigma_x2 + sigma_y2)
    tmp2 = np.sqrt(0.25 * (sigma_x2 - sigma_y2) ** 2 + sigma_xy ** 2)

    return np.sqrt(tmp1 + tmp2), np.sqrt(tmp1 - tmp2), alpha
```

```
print("Frequentist Version")
print("  theta = {0}".format(theta_freq))
print("  sigma_y = {0}".format(sigma_y))
print("  Sigma = {0}".format(Sigma_freq))

from matplotlib.patches import Ellipse
fig, ax = plt.subplots(figsize=(6, 6))
sigma1, sigma2, alpha = get_principal(Sigma_freq)
for nsigma in [1, 2]:
    ax.add_patch(Ellipse(theta_freq,
                          2 * nsigma * sigma1, 2 *
nsigma * sigma2,
                          angle=np.degrees(alpha),
                          lw=2, ec='k', fc='none'))

ax.plot([0, 0], [0, 0], '-k', lw=2)
ax.set_xlim(10, 40)
ax.set_ylim(0.2, 0.7)
ax.set_xlabel(r'$\alpha$')
ax.set_ylabel(r'$\beta$')

plt.show()
```


0.7
0.6
0.5
0.4
0.3
0.2
μ
10
15
20
25
30
35
40
α

Model Comparison

Week 4+5

Introduction to Model Comparison

Qt : Given a dataset:

1) how do we decide which among (two or more) models fit the data best
2) How do we quantify the significance of goodness of fit of the best model compared to disfavored model?

Multiple methods to address the above questions :

- Frequentist model comparison tests (based on difference in chi-square or log-likelihood ratio)
- Bayesian methods (to be discussed later)
- Information criterion based methods

Frequentist Model Comparison

http://jakevdp.github.io/blog/2015/08/07/frequentism-and-bayesianism-5-model-selection/

arXiv:1607.03549 by Louis Lyons (Sect 13)

arXiv:1901.07726 Kerscher & Weller

arXiv:1607.03845 (SD)

A simple application to test if measurements of Newton's Constant show sinusoidal dependence with time as proposed by some authors.

arXiv: 1706.01202 (Shalini Ganguly + SD)

Statistical Significance of spectral lag transition in GRB160625B

arXiv:1906.05726 Aditi Krishak, Aisha Dantuluri, SD model comparison of annual modulation in DM experiments

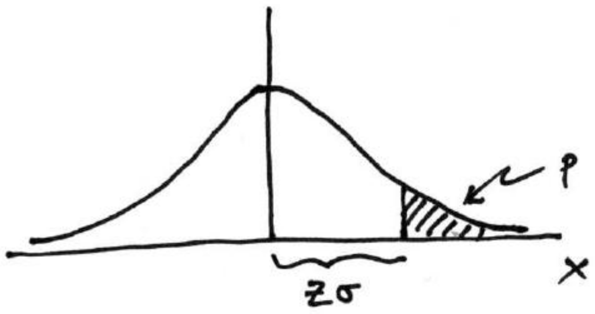
Frequentist Model Comparison Summary

- Calculate best fit $\chi 2$ for each model
- Calculate $\chi 2$ for each model and the one with larger value of $\chi 2$ GOF is the preferred value.
- Use the fact that if the two models are nested $\Delta\chi 2$ between the two models has a $\chi 2$ distribution with DOF = difference in number of free parameters between model 2 and model 1.
- Consider the model with fewer free parameters as the null hypothesis calculate the p-value that simply by chance we would see the more complicated model.
- p-value = 1-chi-square CDF for DOF=difference in no of free parameters and $\chi 2 = \Delta\chi 2$ (between the two models)

In Python use p-value = 1-stats.chi2(ν).cdf($\Delta\chi^2$)

Significance from *p*-value

Often define significance Z as the number of standard deviations that a Gaussian variable would fluctuate in one direction to give the same p-value.



Glenn Cowan lecture Notes

Significance also known as Z-score

$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1 - \Phi(Z)$$

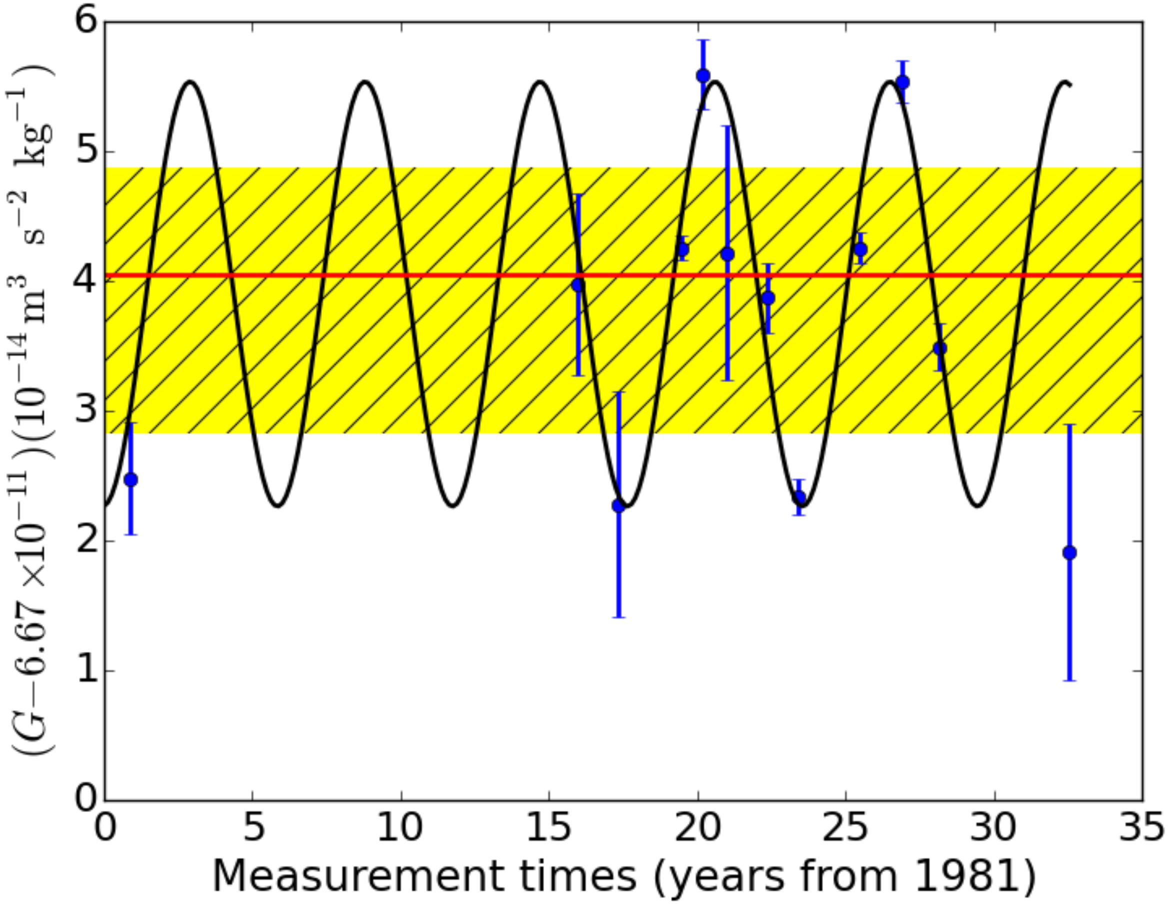
`TMath::Prob`

$$Z = \Phi^{-1}(1-p)$$

`TMath::NormQuantile`

In Python : Use stats.norm.isf(p-value)

Model Comparison Tests of G variation



arXiv:1607.03845

Model Comparison Tests of G variation

$$y_i = A \sin[\phi_0 + 2\pi(t_i/P)] + \mu_G,$$

H1. Data is consistent with a constant offset + measured uncertainties
H2. Same as H1, but an additional unknown systematic offset.
H3. Data is described by above equation (showing sinusoidal variation)
H4. Same as H3, but an additional unknown systematic offset.

Hypothesis	μ	σ_{sys}	A	$P(yrs)$	ϕ_0	DOF	χ^2/DOF	P(χ^2,ν)
H1	6.766×10^{-11}	-	-	-	-	11	28.04	6.8×10^{-60}
H2	6.674×10^{-11}	10^{-14}	-	-	-	10	1.27	0.059
H3	6.674×10^{-11}	-	1.64×10^{-14}	5.9	-0.07	8	2.93	0.0011
H4	6.571×10^{-11}	10^{-12}	1.9×10^{-14}	7.57	0.0011	7	1.71	0.032

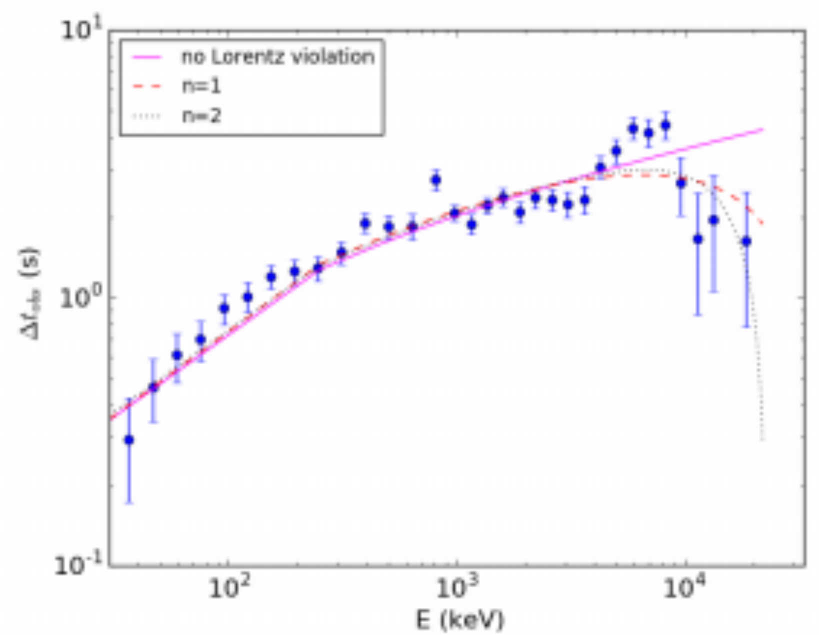


Figure 1 : Summary of the best fit LIV models for $n = 1$ and $n = 2$ along with no Lorentz violation superposed on top of the spectral lag data from GRB 160625B. We note that one data point at (E,Δt) = (15708 keV, -0.223 sec) has been omitted for brevity. All the spectral lag data points have been obtained from Table 1 of W17.

	No LIV[a]	**(n=1)**[b]	**(n=2)**[c]
Frequentist			
DOF	35	34	34
χ^2/DOF	2.6	2.37	2.23
χ^2GOF	2.2×10^{-7}	3.7×10^{-6}	1.5×10^{-5}
p-value		0.0014	9.2×10^{-5}
significance		3.05σ	3.74σ
Δ AIC		8.2	12.9
Δ BIC		6.9	11.7

[a] *No Lorentz Invariance*
[b] *Lorentz Invariance up to linear (n=1) order*
[c] *Lorentz Invariance up to quadratic (n=2) order*

Shalini Ganguly , SD (arXiv:1706.01202)

Information Criterion based tests

Akaike Information Criterion : based on Kullbeck-Leibler information entropy
Bayesian Information Criterion : based on Bayesian evidence

arXiv:1207.5875 Shi, Huang and Lu astro-ph/0701113 Liddle

$$\mathrm{BIC} \equiv -2 \ln \mathcal{L}_{\mathrm{max}} + k \ln N \, ,$$

$$\mathrm{AIC} \equiv -2 \ln \mathcal{L}_{\mathrm{max}} + 2k \, ,$$

$$\mathrm{AIC_c} = \mathrm{AIC} + \frac{2k(k+1)}{N-k-1} \, .$$

N = no of data points
K = no of free parameters

Strength of Evidence Tests for AIC/BIC

For the AIC, Burnham & Anderson (2003) featured the following "strength of evidence" in the form of $\Delta\text{AIC} = \text{AIC}_i - \text{AIC}_{min}$:

ΔAIC	Level of Empirical Support For Model i
$0 - 2$	Substantial
$4 - 7$	Considerably Less
> 10	Essentially None

arXiv:1207. 5875
by Shi et al

For the BIC, Robert & Adrian (1995) featured the following "strength of evidence", where $\Delta\text{BIC} = \text{BIC}_i - \text{BIC}_{min}$:

ΔBIC	Evidence against Model i
$0 - 2$	Not Worth More Than A Bare Mention
$2 - 6$	Positive
$6 - 10$	Strong
> 10	Very Strong

Other Advanced Model-Comparison Tests

- Deviance Information Criterion
- Takeuchi Information Criterion
- F-test
- Likelihood ratio tests
- Cross-validation
- Bayesian Methods : Bayes Factor, Posterior Odds Ratio

For more advanced details see the following note by Louis Lyons
https://www-cdf.fnal.gov/physics/statistics/notes/H0H1.pdf

For more examples of model comparison in Cosmology/Astrophysics, see

Liddle : astro-ph/0701113
Liddle, Mukherjee, Parkinson : astro-ph/0608184
Kerscher and Weller arXiv:1901.07726