Junior Lab Data Analysis Assignment - #2 - Data Reduction

Teaching Staff*
MIT Department of Physics
(Dated: September 14, 2004)

Fits, Method of Maximum Likelihood; χ^2 minimization, least squares method and fits, χ^2 distribution, goodness of fits. *Due September 27,28 - 2004*

1. READING IN BEVINGTON & ROBINSON

- Chapter 6 - Least Squares fit to a straight line
- Chapter 7 - Least Squares fit to a polynomial (skim only)
- Chapter 8 - Least Squares fit to an arbitrary function
- Chapter 11 - (only section 1) Testing the fit.

2. SUMMARY

Now, after we know how to assign a reliable statistical (and) systematic error to a point value, we can proceed to check theories of parameters by their functional dependencies.

Suppose you want to measure the lifetime of a radioactive nucleus or a particle or an excited atom. These are examples of truly indepenent "decays", i.e. it does not matter what their neighbors do nor how long ago they were created (this is different from us; an 80 year old person will die sooner than a 20 year old person). A measurement typically records the number of decays in time intervals of length t or $N(t_i) \pm \sqrt{N(t)}$

To get the best value of the lifetime τ, we must compare fit values with the expectation of $N(t) = N_0 e^{-t/\tau}$ starting from some arbitrary values of N_0' and τ'. Asking a computer to "fit" is to ask it to adjust N_0', and τ' so as to minimize

$$\chi^2 = \sum \left[\frac{N(t_i) - N_0' e^{-t_i/\tau'}}{\sigma(t_i)} \right]^2 \qquad (1)$$

This gives a good fit determining the best values N_0 and τ. Remember that for Gaussian errors, there is a 31.7% chance to find a value outside. Therefore ≈32% of your points will NOT touch the best fit curve through your data, even if the fit is fine!

How do we get the error on the best parameters? Let's look at it the other way around. Given $N_0 e^{-t/\tau}$, what is the probability to observe the values $N(t_i)$?

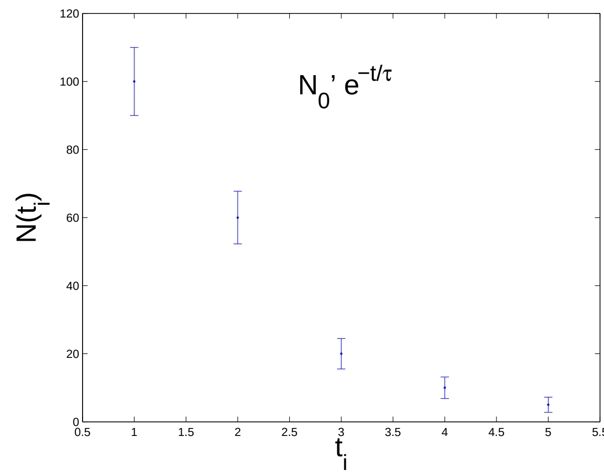

FIG. 1: Typical data for analysis.

$$P(N(t_i)) = \prod^N P_i = \prod^N \frac{1}{\sqrt{2\pi}\sigma(t_i)} e^{-\frac{\chi^2}{2}}. \qquad (2)$$

which is maximum when χ^2 is a minimum. This if we misadjust N_0', τ' such that χ^2 changes by 1/2, we have reached the error limits on our parameters. (Big trouble however, if they are correlated through the covariance matrix). If we had only 2 data points, obviously the procedure would not work. If we had 3 we could expect χ^2 to be small most of the time. For 4,5,6, we might expect larger values. That means χ^2 by itself has a distribution which depends on the number of datapoints minus the degrees of freedom or $N - \nu$.

It is listed here (see Bevington for derivation).

$$P(\chi^2, \nu) = \frac{(\chi^2)^{\frac{1}{2(\nu-2)}} e^{-\chi^2/2}}{2^{\nu/2}\Gamma(\nu/2)}, \qquad (3)$$

So if you find $\chi^2 = 6$ for $\nu = 2$;, either (1) the theory describing the data points is wrong or (2) the errors are underestimated. For $\chi^2 = 6$ with $\nu = 5$, the fit is okay.

However, we do not know the correct parameter set $\{a_0\}$. We make a guess $\{\hat{a}\}$, so that the probability is now:

$$P(\hat{a}) = \prod^N P_i(y_i, \sigma_i, \hat{a}) = \prod^N \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\left(\frac{y_i - y(x,\hat{a})}{\sigma_i}\right)^2} = \mathcal{L}(\hat{a}) \qquad (4)$$

*Electronic address: ; URL: `http://web.mit.edu/8.13`

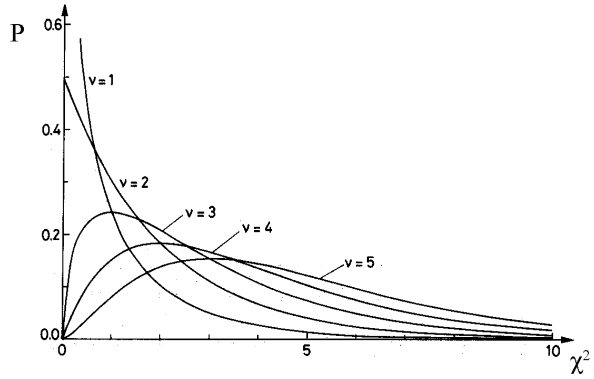

FIG. 2: Probability of getting χ^2 for $\nu = 1, 2, 3, 4$ and 5 degrees of freedom

which is usually smaller. The parameter set $\{a\}$ which maximizes $P(a)$ is called the "Maximum Likelihood" estimate. In the central limit theorem approximation (Gaussian), finding the 1σ level corresponds to increasing χ^2 by 1 or a decrease of $\mathcal{L}(a)$ by $e^{\frac{1}{2}}$. $\mathcal{L}(a)$ is the 'likelihood' function where for a set of x_i observations, $\mathcal{L}(a) = f(x_1|a)f(x_2|a)...f(x_3|a)$ For practical computing reasons, often the $\ln(\mathcal{L})$ is used in which case a change of 1/2 results which usually produces asymmetric $\pm$ errors. To find $(2, 3, \ldots, n) \cdot \sigma$ "confidence" levels, one looks for a $\ln(\mathcal{L})$ change of $n^2/2$. Note that the covariance matrix is connected to $\mathcal{L}$ by:

$$v_{ij} = \frac{\partial(\ln(\mathcal{L}))}{\partial a_i \partial a_j} \text{ with } \quad v_{ij} = V^{-1} \tag{5}$$

where V is the covariance matrix. When the best set found for $\{\hat{a}\}$ is inserted into the model, we have the "best fit." Maximizing $P(\hat{a})$ is equivalent to minimizing the exponent

$$\chi^2 = \sum \left[\frac{y_i - y(x, \hat{a})}{\sigma_i}\right]^2 \tag{6}$$

which is called the method of least squares. So one has a computer changing all parameters until χ^2 is a minimum. But there are "local" minima in the parameter space, which may end the fit prematurely.

When do we know the fit is "good enough"?

1. Plot data with error bars $\sigma_{y'}$ and superpose the best fit curve on the same graph. If 68% of the points are are intersected by the curve in their $\pm 1\sigma$ interval and the errors were Gaussian, it will be OK.

2. For a more quantitative approach, use the χ^2-test. If we repeated the entire measurement and then first fit with χ^2 an infinite number of times, we get the χ^2 distribution

$$P(\chi^2, \nu) = \frac{(\chi^2)^{\frac{1}{2(\nu-2)}} e^{-\chi^2/2}}{2^{\nu/2}\Gamma(\nu/2)}, \tag{7}$$

The mean of this distribution is ν where $\nu = N(\text{measurement points}) - m(\text{parameters})$, and is called "degrees of freedom" or DOF. The function Γ is the continuous version of the factorial function. For additional information on common probability distributions, see W. R. Leo "Techniques for Particle Physics Experiments," available from the Junior Lab e-library.

Example:

In a given data set, 7 points are fit by a line $y = a + bx$, then $\nu = 7 - 2 = 5$. But as seen from Figure 2. the curve above $P(\chi^2)$ of 3, 4, 5 are approximately equal. So even though $\chi^2 = 3$ is less than a different data set minimizing at $\chi^2 = 4$ the probability $P(\hat{x})$ does not support that one parameter set is better than the other.

On the other hand, if $\int_{\chi_2}^{\infty} P(\chi^2, \nu)$ is $< 1\%$ (your judgement) it means that there was only a 1% chance to get a match this way so with 99% the hypothesis is bad.

Variance:

Recall

$$\sigma_x^2 = \langle (x_i - \langle x \rangle)^2 \rangle = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \langle x \rangle)^2 \tag{8}$$

for a distribution in x. If we have a function in two variables $y = f(x, y)$, we introduce the

Covariance:

$$\sigma_{xy}^2 = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle = \frac{1}{N} \sum^{N} (x_i - \langle x \rangle)(y_i - \langle y \rangle) \tag{9}$$

which can be positive, negative or zero. If $\sigma_{xy} = 0$, the x's and y's fluctuate up and down independently and cancel. If there is any correlation between x and y the balance is perturbed and the covariance exists.

So if x, y are independent: $\sigma_{xy} = 0$ and simply:

$$P(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{\Delta x^2}{\sigma_x^2} + \frac{\Delta y^2}{\sigma_y^2}\right)} \tag{10}$$

If x and y are not independent, we have:

$$P(x, y) = \frac{1}{2\pi\sigma^2(1 - \frac{\sigma_{xy}^2}{\sigma^2})} e^{-\frac{1}{2}\frac{\Delta x^2 + \Delta y^2 - 2\frac{\sigma_{xy}^2}{\sigma^2}}{2\sigma^2\left[1 - \left(\frac{\sigma_{xy}^2}{\sigma^2}\right)^2\right]}} \tag{11}$$

The contours of constant $P(x, y)$ are ellipses in (x, y) plane which are often plotted at the $1\sigma, 2\sigma$, and 3σ level in modern publications. 8.13 uses them very seldom.

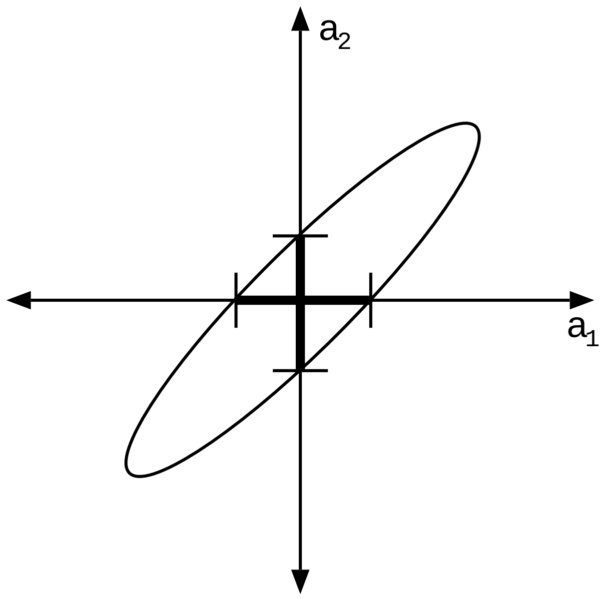

FIG. 3: Contours of constant χ^2 for 1σ (likelihood= 1/2) confidence levels for two connected parameters a_1 and a_2. Without correlation the errors would be the intercepts of the ellipse on the axes.

3. PROBLEMS

1. Old and New Bevington and Robinson problem 7.2
2. Hand in your solutions to the "Statistical Exercises" listed at the end of the labguide of your first introductory experiment.