General error propagation

May remember $c = a \pm b \Rightarrow \sigma_c^2 = \sigma_a^2 + \sigma_b^2$

$c = a \times\!\div b \Rightarrow \left(\frac{\sigma_c}{c}\right)^2 = \left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2$

More generally, have $c = f(a_1, a_2, \ldots, a_n)$

for $n=1$, simple-minded way: $\left.\begin{array}{l} c_+ = f(a+\sigma_a) \\ c_- = f(a-\sigma_a)\end{array}\right\} \sigma_c = \frac{|c_+ - c| + |c_- - c|}{2}$

Of course, really this is $\sigma_c = \sigma_a \left|\frac{df}{da}\right|$

But this assumes $\left|\frac{d^2f}{da^2}\right| \sigma_a << \left|\frac{df}{da}\right|$. Counterexample

$c = \ln(a),\ a=1,\ \sigma_a = 0.5$

for $n>1$, for uncorrelated parameters

$$\sigma_c = \left[\sum_{i=1}^{n} \left(\sigma_{a_i} \left|\frac{\partial f}{\partial a_i}\right|\right)^2\right]^{1/2}$$

Example; taking average $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$\Rightarrow \sigma_{\bar{x}} = \left[\sum_{i=1}^{n}\left(\frac{\sigma_i}{n}\right)^2\right]^{1/2} \underset{\text{if } \sigma_i = \sigma\ \forall i}{=} \left[\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2\right]^{1/2} = \frac{\sigma}{\sqrt{n}}$$ as expected

(but often done wrong)

Can estimate σ from data
"standard deviation" $\sigma = \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}\right]^{1/2}$

Why this form? For normal distribution $(e^{-\frac{1}{2}x^2/\sigma^2})$,
on average $\langle (x_i - \bar{x})^2 \rangle = \sigma^2$
So use all data points as estimates & average these
BUT: divide by $n-1$
WHY? Since we determined average from data too
(would be division by n if $\bar{x}$ was known independently)

χ^2 fitting

Above suggests general goodness of fit

$$\chi^2 = \sum_{i=1}^{n} \underbrace{\left(\frac{y_i - f(x_i; a_1, a_2, \ldots a_m)}{\sigma_i} \right)^2}_{\text{expectation of } \sim 1 \text{ for a good model}}$$

$\Rightarrow \langle \chi^2 \rangle = n$ if $f(x_i)$ has no free parameters

More generally $\langle \chi^2 \rangle = n - m$ = degrees of freedom

How to find minimum? Need $\dfrac{\partial \chi^2}{\partial a_m} = 0 \ \forall m$

Example: determine mean: $\chi^2 = \sum_{i=1}^{n} \left(\dfrac{y_i - a}{\sigma_i} \right)^2$

$$\frac{\partial \chi^2}{\partial a} = 2 \sum_{i=1}^{n} \frac{y_i - a}{\sigma_i^2} \times -1 = 0$$

$$\Rightarrow -\sum_{i=1}^{n} \frac{y_i}{\sigma_i^2} + a \sum \frac{1}{\sigma_i^2} = 0$$

$$\Rightarrow a = \frac{\sum_{i=1}^{n} y_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}$$ weighted mean w/ $w = 1/\sigma_i^2$

Check: suppose $\sigma_i = \sigma \ \forall i \Rightarrow a = \dfrac{1/\sigma^2 \sum y_i}{1/\sigma^2 \sum 1} = \dfrac{\sum y_i}{n} = \bar{y}$ ✓

What is expected uncertainty on a?

Above some $\chi^2 = \chi^2_{min} + \Delta\chi^2$ fit is not good anymore

$1\sigma \Rightarrow \Delta\chi^2 = 1$ ($2\sigma \Rightarrow \Delta\chi^2 = 4$, etc.)

Taylor expand around a_0: $\chi^2 = \chi^2_{min} + \underbrace{(a-a_0) \dfrac{\partial \chi^2}{\partial a}\Big|_{a=a_0}}_{=0} + \dfrac{1}{2}(a-a_0)^2 \dfrac{\partial^2 \chi^2}{\partial a^2}\Big|_{a=a_0} + \ldots$

Curvature $\dfrac{\partial^2 \chi^2}{\partial a^2} = 2 \sum \dfrac{1}{\sigma_i^2} \Rightarrow \Delta\chi^2 = \dfrac{1}{2}\left(2 \sum \dfrac{1}{\sigma_i^2}\right)(a-a_0)^2$

$\Delta\chi^2 = 1 \Rightarrow (a-a_0)^2 = \sigma_a^2 = \dfrac{1}{\sum 1/\sigma_i^2} \overset{\sigma_i = \sigma \forall i}{\Rightarrow} \sigma_a = \dfrac{\sigma}{\sqrt{n}}$ ✓

Quality of the fit

If fit is good (model an adequate description of the data)

Expect $\langle \chi^2 \rangle \simeq n - m$ or $\chi^2_{red} \equiv \frac{\chi^2}{n-m} \simeq 1$

Have statistical fluctuations w/ "χ^2 distribution" (see N.R.)

$\sigma_{\chi^2} = \sqrt{2(n-m)}$ (eg $n-m = 50 \Rightarrow \chi^2 = 50 \pm 10$, expect $\chi^2_{red} = 1.0 \pm 0.1$)

What if $\chi^2_{red} > 1$? $\begin{cases} \text{Errors overestimated?} \\ \text{Bad model?} \end{cases}$

$\chi^2_{red} < 1$ Errors underestimated?

* If one has reason to believe errors are over/under estimated in a non-systematic way, one can rescale them

$\sigma_{new} = \sigma_{old} \sqrt{\chi^2_{red}}$, which will yield $\chi^2_{red, new} = 1$

(& rescale all parameter uncertainty estimates w/ $\sqrt{\chi^2_{red, old}}$)

But think carefully before blindly doing it! [Some fitting routines do this automatically!]

Eg, poor PSF

fit w/ Gaussian ⇒ poor fit. Absolute position: rescaling may be reasonable

Relative positions: better not to rescale

Fits are wrong in same way!

* Bad model? Try to increase complexity & test signif. of improvement with, e.g, F test. (Bayes better...)

Generally: much better to try to find good physical model

Fitting a line. Doing something else explicitly

Try $\chi^2 = \sum \left(\frac{y_i - a - bx_i}{\sigma_i}\right)^2$

$w \equiv 1/\sigma_i^2$

Minimum @ $\frac{\partial\chi^2}{\partial a} = -2\sum \frac{y_i - a - bx_i}{\sigma_i^2} = 0 \Rightarrow a\sum w + b\sum wx = \sum wy$ (1)

$\frac{\partial\chi^2}{\partial b} = -2\sum \frac{(y_i - a - bx_i)x_i}{\sigma_i^2} = 0 \Rightarrow a\sum wx + b\sum wx^2 = \sum wxy$ (2)

(2)$\times\sum w$ − (1)$\times\sum wx \Rightarrow b\left(\sum wx^2 \sum w - (\sum wx)^2\right) = \sum wxy\sum w - \sum wy\sum wx$

$$\Rightarrow b = \frac{\sum wxy \sum w - \sum wy \sum wx}{\sum wx^2 \sum w - (\sum wx)^2}$$

$$\Rightarrow a = \frac{\sum wy - b\sum wx}{\sum w}$$

a correlated with b ⇒

Problem: not very good choice of par. if $\sum wx$ large

Better to use $x_i' = x_i - \bar{x}$ $\quad \left(\bar{x} = \frac{\sum wx}{\sum w}\right)$

$$\Rightarrow \sum wx_i' = 0 \quad \Rightarrow \begin{cases} b' = b = \dfrac{\sum wx'y}{\sum wx'^2} \\ a' = \dfrac{\sum wy}{\sum w} \end{cases}$$

[can also give covariance, which gives same information, but in less useful way]

Real-world examples: star with spectrum ⇒ know intrinsic abs. mags

* Determine distance & reddening $\chi^2 = \sum_{\text{filters}} \left(\frac{m_i - (M_i + 5\log d + A_V \frac{A_i}{A_V})}{\sigma_i} \right)^2$

"y" "a" "b" "x"

$= (m-M)_0$

To minimize covariance, for globular clusters, generally $(m-M)_V$ & A_V are given

(unknown: less simple!)

* Fitting sine curve with known period $y_i - \cancel{a}\ b\sin(\omega t + \varphi)$

$= y_i - c - b\sin\omega t\cos\varphi - b\cos\omega t\sin\varphi$

"a" "b_1" "x_1" "b_2" "x_2"

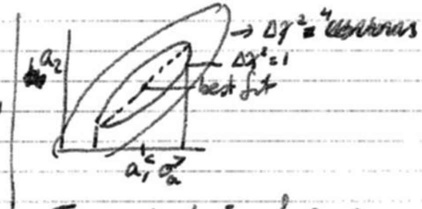
Again, want to choose $t=0$ in middle of sequence

Error estimates for $\nu > 1$ parameters

$\Delta\chi^2 = 1$ good if only one parameter is of interest,

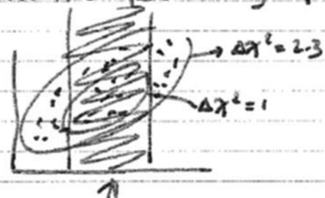
i.e., if one <u>marginalises</u> over all other parameters:

as you scan over the parameter of interest, you refit all others & continue until you reach $\Delta\chi^2 = 1$



For good choice of a_1, a_2, χ^2 surface is ~~circle~~ symmetric $\Rightarrow$ best-fit a_2 is independent of a_1

What does this mean, really:



1 par of interest: this BAND contains ~~68%~~ solution w/ 68.3% confidence

If you want a joint constraint, need a different $\Delta\chi^2$ level

$\Delta\chi^2 < 2.3$ contains solution for 2-parameters w/ 68.3% confidence

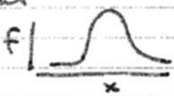
Stat mini #2

1

Optimal measurements: optimal aperture extraction radius

How to get best S/N? Examples are images & spectra
→ article by Horne about spectra was very useful for me as a graduate student

How to measure?



Just integrate, fit? How to optimize.

Start with simplest, aperture photometry

Estimate flux with $F_{ap} = \dfrac{\sum_{x,y,<r} D_{x,y}}{Frac_{<r}}$ → this is $\sum_{x,y,<r} PSF_{x,y}$

Suppose limited by Poisson noise, background small] generally measured from bright stars

$\Rightarrow \sigma_{x,y} = \sqrt{D_{x,y}}$

$\Rightarrow S/N = \dfrac{\sum_{x,y,<r} D_{x,y}}{\sqrt{\sum_{x,y,<r} \sigma^2_{x,y}}} = \dfrac{1}{\sqrt{\sum_{<r} D_{x,y}}}$ ⇒ just make aperture as big as possible
→ must be so, since your adding more signal.

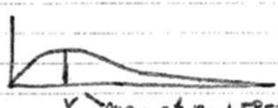
Now background limited, $\sigma_{x,y} = \sigma_b$

$$S/N = \frac{\sum_{<r} D_{x,y}}{\sqrt{\sigma_b^2 \sum_{<r} 1}} = \frac{\sum_{<r} D_{x,y}}{\sigma_b \sqrt{\pi r^2}}$$ → again makes sense.

For Gaussian PSF $\quad D_{x,y} = F \, e^{-\frac{1}{2}(x/\sigma)^2} e^{-\frac{1}{2}(y/\sigma)^2} \dfrac{1}{2\pi\sigma^2}$ ← normalisation of PSF

$$\Rightarrow \sum_{<r} D_{x,y} \simeq \frac{F}{2\pi\sigma^2} \int_0^r dr' \int_0^{2\pi} r' d\phi \; e^{-\frac{1}{2}(r'/\sigma)^2} = -e^{-\frac{1}{2}(r'/\sigma)^2}\Big|_0^r = 1 - e^{-\frac{1}{2}(r/\sigma)^2}$$

$$\Rightarrow \frac{S}{N} = \frac{F}{\sigma_b} \frac{1 - e^{-\frac{1}{2}(r/\sigma)^2}}{\sqrt{\pi}\, r}$$



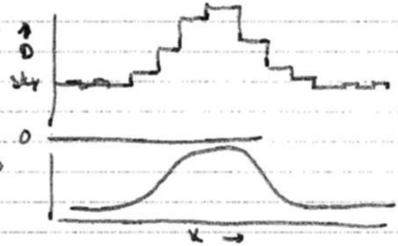
max at $r = 1.5852\,\sigma$ ≈ 0.7 FWHM

$$\max S/N = \frac{F}{\sigma_b \sqrt{\pi}\,\sigma} \frac{0.715}{1.5852} \simeq \frac{F}{\sigma_b \sqrt{15.4\sigma^2}}$$

⇒ equivalent to measuring all the flux in $15.4\,\sigma^2$ pixels

Optimal measurement → thinking more about it

Have image with stars with some PSF



What is PSF really?

→ probability that light falls at a given pixel x, y

hence, $D(x,y) = S + F * PSF(x,y)$

Conversely, if you know the PSF (& can centroid well) every pixel x, y gives an independent estimate of the flux:

$$F^{xy} = \frac{D(x,y) - S}{PSF(x,y)} \quad \text{with uncertainty} \quad \sigma(F^{xy}) = \frac{\sigma(x,y)}{PSF(x,y)}$$

Measuring the flux of a star is thus taking some weighted average of all these independent estimates

$$F = \frac{\sum_{x,y} w_{x,y} [D(x,y) - S] / PSF(x,y)}{\sum_{x,y} w_{x,y}}$$

What weights to choose? Best is: $w_{x,y} = \frac{1}{\sigma^2(F^{xy})} = \left(\frac{PSF(x,y)}{\sigma(x,y)}\right)^2$

$$\Rightarrow F_{opt} = \frac{\sum_{x,y} \frac{1}{\sigma^2_{(x,y)}} (PSF(x,y))^2 [D(x,y) - S] / PSF(x,y)}{\sum_{x,y} \frac{1}{\sigma^2_{(x,y)}} (PSF(x,y))^2}$$

$$\sigma_{opt} = \left[\frac{1}{\sum_{x,y} w_{x,y}}\right]^{1/2} = \left[\frac{1}{\sum_{x,y} (PSF/\sigma)^2_{x,y}}\right]^{1/2}$$

Uncertainties for optimal case?

$D \gg S$, $\cancel{\sigma_{x,y} = \sqrt{D_{x,y}}}$ $\Rightarrow$ $\cancel{F_{opt} = \dfrac{\sum_{x,y} \frac{(PSF(x,y))^2}{D(x,y)} \frac{D(x,y)}{PSF(x,y)}}{\sum_{x,y} \frac{PSF(x,y)^2}{D(x,y)}}}$

$\cancel{D(x,y) = F \cdot PSF(x,y)}$

$$\sigma_{x,y} = \sqrt{F\, PSF(x,y)} \Rightarrow F_{opt} = \frac{\frac{1}{F}\sum_{x,y} D_{x,y}}{\frac{1}{F}\underbrace{\sum_{x,y} PSF(x,y)}_{=1}} = \sum_{x,y} D_{x,y}$$

So, integrating over aperture is optimal for Poisson-limited sources. Uncertainty $\sqrt{F}$

Background limited $\sigma_{x,y} = \sigma$

$$F_{opt} = \frac{\sum_{x,y} (PSF(x,y))^2 (D-S)/PSF}{\sum (PSF)^2}$$

$$\sigma_{opt} = \sigma \left[\frac{1}{\sum (PSF)^2}\right]^{1/2} = \sigma\sqrt{4\pi\sigma^2} = \sigma\left(\sqrt{4\pi}\,\sigma\right)$$

$$\frac{1}{(2\pi\sigma^2)^2}\int_0^\infty dr \int_0^{2\pi} r\,d\phi\; e^{-(r/\sigma)^2} = \frac{-1}{2\pi\sigma^2}\frac{1}{2}e^{-(r/\sigma)^2}\Big|_0^\infty = \frac{1}{4\pi\sigma^2} \qquad PSF = \frac{e^{-\frac{1}{2}(r/\sigma)^2}}{2\pi\sigma^2}$$

$\Rightarrow$ Equivalent to measuring full flux in 12.6 pixels

$\Rightarrow$ improvement in S/N of factor $\frac{12.6}{15} = 1.23$

effective integration time $\frac{2}{3}$ shorter

Biggest benefit: cosmic-ray removal!

Have independent estimates of total flux

$\Rightarrow$ can check their consistency & remove outliers

$\rightarrow$ Brian Lee's method

stat mini ~~#2~~

4

Equivalence to χ^2 fitting / Determining PSF

$$\chi^2 = \sum \Big(\frac{D - S - F\,PSF}{\sigma}\Big)^2$$

"y" − "a" − "b" × "x"

Note: this is again the same as fitting a line!

So could fit for Sky & Flux at the same time,

but instead let's assume S is fixed as before

$$\frac{\partial \chi^2}{\partial F} = -2 \sum \frac{(D-S - F \times PSF) \times PSF}{\sigma^2} = 0$$

$$\Rightarrow F = \frac{\sum \frac{1}{\sigma^2}(D-S)\,PSF}{\sum \frac{1}{\sigma^2} PSF^2} \Rightarrow \text{this is indeed identical}$$

How to determine PSF?

- Analytical model, fitting parameters to, say, bright stars
- Directly from bright stars (DAOPHOT: as residuals to analytic fit)
  - ↳ concept of ePSF: already integrated over pixels
- For spectra: <u>much</u> easier → can use spectrum itself, assuming slow changes with wavelength

What to use for uncertainties

Above, used $\sigma = \sqrt{D}$ as an estimate

But this is biased for faint sources

Check using simple weighted average:

$$\overline{D} = \frac{\sum_i \frac{1}{\sigma_i^2} D_i}{\sum_i \frac{1}{\sigma_i^2}} = \frac{N}{\sum_i \frac{1}{D_i}} = \frac{1}{\langle \frac{1}{D} \rangle}$$

say 9 ± 3

$$\Rightarrow \overline{D} \approx \frac{1}{(\frac{1}{6}+\frac{1}{12})/2} = \frac{1}{(\frac{3}{24})} = 8!$$

Problem is that statistical fluctuations towards low counts get higher weight than they should have had.

So, σ is not a good estimate. Instead should base the estimate of the uncertainty on expectation. Given a model predicted flux, $\sigma = \sqrt{D_{pred}}$.
This is unbiased.

Also makes CR rejection better.
Typically CR have many counts $\Rightarrow \sigma$ large
But if predicted counts low, this number of observed counts is very significantly too high

Actually, makes sense also in a more basic way:
What is the real measurement uncertainty if my CCD image has 10 123 counts? Could it really be 10122?
NO! $\Rightarrow$ it is 10123. It is just that any ~~flux~~ model that predict counts in the range $\sim 10123 \pm \sqrt{10123}$ is OK @ 68.3% confidence.

Similarly if you found, say, 1 planet among 100 stars.
Frequency is not 1 ± 1 %! After all, zero is excluded at high confidence... Rather find range of p for which prob. of finding 1 star is ~~some~~ above some desired confidence level.

Linear least-squares fitting

Saw before that perhaps surprising functions are easily solvable as a fit to a line

$$\chi^2 = \Sigma\left(\frac{D_i - S - F\,PSF_i}{\sigma_i}\right)^2$$

Generally, any function that depends linearly on its parameters can be solved by a simple matrix inversion

$$\chi^2 = \sum_i\left(\frac{D_i - \sum_m a_m f_i^m}{\sigma_i}\right)^2$$

w/ f_i^m an arbitrarily complex function of the indep. vari.'s

$$\frac{\partial \chi^2}{\partial a_m} = -2\sum_i\left(\frac{D_i - \sum_{m'} a_{m'} f_i^{m'}}{\sigma_i}\right) f_i^m$$

⇒ m equations $\sum_{m'} a_{m'} \sum_i f_i^{m'} f_i^m = \sum_i D_i f_i^m$ (ignoring weights)

$$\begin{pmatrix} \sum_i f_i^1 f_i^1 & \sum_i f_i^1 f_i^2 & \cdots & \vdots \\ \sum_i f_i^2 f_i^1 & \ddots & & \vdots \\ \vdots & & \ddots & \\ \vdots & \cdots & \cdots & f_i^m f_i^m \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum_i D_i f_i^1 \\ \vdots \\ \sum_i D_i f_{mi}^m \end{pmatrix}$$

$[\alpha]\cdot\vec{a} = [\beta]$

$[C] \equiv [\alpha]^{-1} = \longrightarrow$ inverse in the covariance matrix

$\vec{a} = [\alpha]^{-1}[\beta]$

$\sigma^2_{a_m} = C_{mm}$

So, only need matrix inversion. Quite a few tricks for that. See NR

Especially, SVD →

will give orthonormal set of base vectors

→ can get to know which properties of the model are really well-constrained

Example

$Y = Ae^{bx}$

~ linear → $\ln Y = \ln A - bx$

But beware

$Y = Ae^{bx+\phi}$

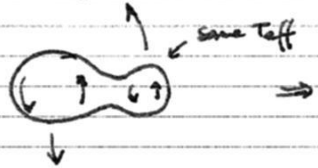
covariant

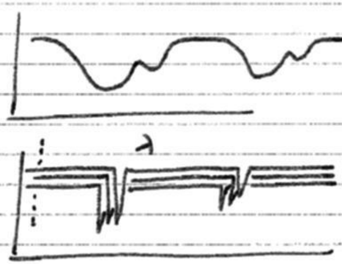
⇒ SVD will tell

Examples of linear least-squares
(some somewhat surprising)

* Broadening function

Suppose observe a contact binary (or galaxy)





How to measure vel. & width?

Could model, → non-linear (1)

Or decompose (2)

(1) $\chi^2 = \sum_\lambda \frac{1}{\sigma_\lambda^2}\left(y_\lambda - A_1 f_\lambda \otimes P(v\sin i_1, v_{rad_1}) - A_2 f_\lambda \otimes P(v\sin i, v_{rad_2})\right)^2$

(2) $\chi^2 = \sum_\lambda \frac{1}{\sigma_i^2}\left(y_\lambda - \sum_{\Delta v} a_{\Delta v} \underbrace{f\left(\lambda\left(1+\frac{\Delta v}{c}\right)\right)}\right)^2$

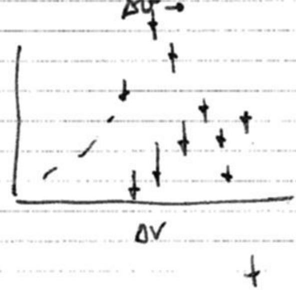
↳ somewhat complicated base function but independent of par → can simply used observed spectrum of slow rotator

amplitudes are the parameters

⇒ linear! Solve easily

Next, can fit this with, say, 2 gaussians

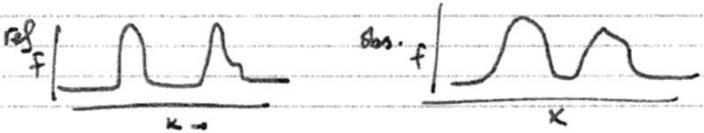
But: tricky to chose resolution right:
to closely spaced ⇒ get cancellation
(SVD helps here somewhat)



Example #2: Optimal image subtraction

Suppose one has time series of v. crowded field & want to look for variables (say μ-lensing).
Photometry $\Rightarrow$ tricky (diff seeing $\Rightarrow$ DAOPHOT won't find all stars etc)

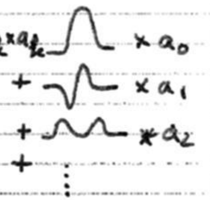
Would like to just subtract! But optimally ..



Take good image as reference

Want to find optimal match: $\chi^2 = \sum_i \frac{1}{\sigma_{obs}^2}(obs - ref \otimes PSF)^2$

Looks tricky, but maybe write PSF $= \sum_k$ gauss $\times$ pol$_k$ $\times a_k$



$$\Rightarrow \chi^2 = \sum_i \frac{1}{\sigma_i^2}\left(obs_i - \sum_k a_k (ref \otimes PSF_k)_i\right)^2$$

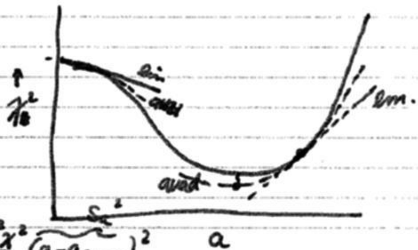
$\Rightarrow$ again linear! Can add $-b_0 - b_1 x - b_2 x^2$ for background variations

This can work amazingly well.

Non. linear models

not linear <u>in parameters</u>

E.g., Gaussian: $\chi^2 = \sum \left(\dfrac{y - a\, e^{-\frac{1}{2}[(x-x_c)/\sigma]^2}}{\sigma} \right)^2$



Need to find minimum iteratively.
No derivatives → Simplex method
Derivatives → much faster!
Example for 1 parameter

Expand $\chi^2 = \chi^2_{curr} + \dfrac{\partial \chi^2}{\partial a}\overbrace{(a - a_{curr})}^{\delta a} + \dfrac{1}{2}\dfrac{\partial^2 \chi^2}{\partial a^2}(a - a_{curr})^2$

Can simple take step down (steepest descent)

$$\delta a = -\text{CONSTANT} \times \frac{\partial \chi^2}{\partial a}$$

Near minimum, χ^2 surface ~ parabolic → can use 2^{nd} derivative

$$\Rightarrow \text{minimum @ } \delta a = -\frac{\partial \chi^2}{\partial a} \Big/ 2\frac{1}{2}\frac{\partial^2 \chi^2}{\partial a^2} = -\frac{\partial \chi^2/\partial a}{\partial^2 \chi^2/\partial a^2}$$

For > 1 parameter, need to solve: $\displaystyle\sum_{l=1}^{m} \left(\frac{\partial^2 \chi^2}{\partial a_k \partial a_l} \delta a_l \right) = -\frac{\partial \chi^2}{\partial a_k}$

What are deriv.s? $\displaystyle\frac{\partial \chi^2}{\partial a_k} = -2 \sum_i \frac{y_i - f(x_i;\vec{a})}{\sigma_i^2} \frac{\partial f(x_i;\vec{a})}{\partial a_k}$

$$\frac{\partial^2 \chi^2}{\partial a_k \partial a_l} = 2 \sum_i \frac{1}{\sigma_i^2}\left[\frac{\partial f(x_i;\vec{a})}{\partial a_k}\frac{\partial f(x_i;\vec{a})}{\partial a_l} - \underbrace{[y_i - f(x_i;\vec{a})]\frac{\partial^2 f(x_i;\vec{a})}{\partial a_l \partial a_k}}_{} \right]$$

close to min, where parabolic approx useful, this should be small ⇒ <u>ignore</u>

Define $\beta_k = -\dfrac{1}{2}\dfrac{\partial \chi^2}{\partial a_k}$

$$\alpha_{kl} = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial a_k \partial a_l} \simeq \sum \frac{1}{\sigma_i^2}\left[\frac{\partial f(x_i;\vec{a})}{\partial a_k}\frac{\partial f(x_i;\vec{a})}{\partial a_l}\right]$$

CURVATURE MATRIX

Levenberg-Marquardt

How to switch between steepest descent & inverse-Hessian

Steepest descent: what to use for constant? Matrix gives some info.

Derivative changes on scale $1/\frac{\partial^2 \chi^2}{\partial a^2}$

$\Rightarrow$ choose $\delta a_\ell = \frac{1}{\lambda \alpha_{\ell\ell}} \beta_\ell$

"CONSTANT" factor to allow smaller steps
NON-DIMENSIONAL

Can also write this as $\lambda \alpha_{\ell\ell}\, \delta a_\ell = \beta_\ell$

This looks like $\Sigma \alpha_{\ell k}\, \delta a_k = \beta_\ell$ for 2nd order

Combine $\alpha'_{\ell k} = (1+\lambda \delta_{\ell k})\alpha_{\ell k}$

$\Rightarrow \Sigma \alpha'_{\ell k}\, \delta a_k = \beta_\ell$

SOLVE ALWAYS

λ large $\Rightarrow$ diagonally dominant $\Rightarrow$ steepest descent (with relatively small step)

λ small $\Rightarrow$ parabolic

Scheme: Compute $\chi^2(\vec{a})$ start with, say, $\lambda = 10^{-4}$
* Solve for $\delta\vec{a}$ & evaluate $\chi^2(a+\delta a)$
if $\chi^2(a+\delta a) > \chi^2(a)$, increase λ (factor 10) & repeat
if $\chi^2(a+\delta a) < \chi^2(\vec{a})$, decrease λ, ~~replace~~ update $\vec{a}$ + repeat

Stop when $\Delta\chi^2 \leq 0.01$ (say; $\Delta\chi^2 \ll 1$ not meaningful)
then calculate proper $\alpha_{k\ell}$ & covariance matrix $= \alpha^{-1}$

Non - Gaussian errors

Much of astronomy based on detection of individual photons
$\Rightarrow$ Poisson statistics.
For large N, $\sigma \simeq \sqrt{N}$, so not so bad

But, esp. @ X-ray en. often just few photons.

Ex. $\rightarrow$ time series $\rightarrow$ find period
object detection $\rightarrow$ find centroid, flux
spectral data $\rightarrow$ fit model

Could bin such that one has $\gtrsim 20$ / bin, & $\sigma \simeq \sqrt{N} \sim$ Ok
But loose information! Also bin-size arbitrary. Will result depend on it?
Note: 3 photons is significant detection for Chandra!

—

So, should use actual probability: for some model f

$$P = \prod_i \frac{f_i^{n_i} e^{-f_i}}{n_i!}$$

Nice general property for "sufficiently large" n & "well-behaved" P

$$-2 \ln \frac{P}{P_{max}}$$ is distributed like χ^2! (w/ corr. # of par)
(correction $\sim n^{-1/2} \Rightarrow$ small already for $n = 9$ or 10)

So, for Poisson minimize $$C = -2 \ln P = 2 \sum_i \left(f_i - n_i \ln f_i + \underbrace{\ln n_i!}_{} \right)$$

does not depend on model & cancels on taking $\Delta C = C - C_{min}$
$\Rightarrow$ can be ignored

NOTE 1: solvers can easily be rewritten to minimize C instead of χ^2; simply use slightly different dependences of α (curvature matrix) & β on the derivatives.

NOTE 2: Unlike χ^2, no "goodness of fit", but can compare C for different models.

NOTE 3: result independent of binning

NOTE 4: $f_i < 0$ unphysical; $f_i = 0$ only for $n_i = 0$

Application of Cash statistic #1

Source on zero background : $f = A\,\dfrac{e^{-\frac{1}{2}((x-x_c)/\sigma)^2}}{\sqrt{2\pi\sigma^2}}$

$$C = +2\sum_i \left(f_i - n_i \ln f_i\right) = 2\sum_i \Big\{ A\frac{e^{\cdots}}{\sqrt{2\pi\sigma^2}} - 2\sum_i n_i \ln A + 2\sum_i \left(-\tfrac{1}{2}\right) n_i \left(\frac{x-x_c}{\sigma}\right)^2$$

$$+2\sum_i n_i \ln\sigma$$

no dep. on n_i!
Sum over whole image
$\Rightarrow\ = 2A$

minimize : $\dfrac{\partial C}{\partial A} = 2 - 2\sum \dfrac{n_i}{A} = 0 \Rightarrow A = \sum n_i$

This is general : best fit ALWAYS reproduces total number of observed counts

$$\frac{\partial C}{\partial x_c} = \underbrace{2\sum_i A\frac{e^{-\frac{1}{2}\left(\frac{x_i-x_c}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}}\cdot\frac{(x_i-x_c)}{\sigma^2}}_{=0} + \sum_i n_i\, 2\,\frac{x_i-x_c}{\sigma^2}\cdot -1 = 0$$

$$\Rightarrow x_c = \frac{\sum n_i x_i}{\sum n_i} \quad \text{makes sense}$$

$$\frac{\partial C}{\partial \sigma} = \underbrace{\cancel{2\sum_i A\frac{e^{-\frac{1}{2}\left(\frac{x_i-x_c}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}}\cdot\frac{(x_i-x_c)^2}{\sigma^2}\cdot\frac{1}{\sigma}}}_{=0} + \sum_i n_i \frac{(x_i-x_c)^2}{\sigma^2}\,\frac{-2}{\sigma} + 2\sum n_i \frac{1}{\sigma}$$

$$= 2\sum_i n_i \frac{1}{\sigma} - 2\,\frac{\sum_i n_i (x_i-x_c)^2}{\sigma^3} = 0$$

$$\Rightarrow \sigma^2 = \frac{\sum n_i (x_i-x_c)^2}{\sum n_i} \quad \text{makes sense}$$

Good : 2-dimensional splits & get same results for x & y
But : not so easy any more if you add background ...

Multiplicative terms easy $\Rightarrow$ can be minimized separately
Additive terms difficult ...

Stat mini #4

③

(maybe skip)

Application of Cash Statistics #2

Time series w/ modulation: $f = A(1+\alpha\cos\omega t_i + \beta\sin\omega t_i)$

Assume ω known

$$C = 2\sum_i A(1+\alpha\cos\omega t_i + \beta\sin\omega t_i) - 2\sum_i n_i \ln A - 2\sum_i \ln(1+\alpha \ldots + \beta \ldots)$$

$$\frac{\partial C}{\partial A} = 2\underbrace{\sum_i (1+\alpha\cos\omega t_i + \beta\sin\omega t_i)}_{2\, t_{tot}} - 2\,\frac{\sum n_i}{A}$$

$$\Rightarrow A = \frac{\sum n_i}{t_{tot}} = \frac{N}{t_{tot}}$$ makes sense

$$\frac{\partial C}{\partial \alpha} = 2\underbrace{\sum_i A\cos\omega t_i}_{=0} - 2\sum \frac{n_i\cos\omega t_i}{1+\alpha\cos\omega t_i + \beta\sin\omega t_i} = 0$$

$$\alpha, \beta \ll 1 \Rightarrow \sum n_i\cos(\omega t_i)\cdot(1-\alpha\cos\omega t_i - \beta\sin\omega t_i) = 0$$

$$= \sum n_i\cos\omega t_i - \sum_i n_i\alpha\left(\tfrac{1}{2} - \tfrac{1}{2}\cos 2\omega t_i\right) - \sum n_i\beta\,\tfrac{1}{2}\sin 2\omega t_i = 0$$

$\cos 2\omega t_i$ term $= 0$, $\sin 2\omega t_i$ term $= 0$

$\neq 0!$ Since n_i likely larger at some phases

$$\Rightarrow \alpha = 2\,\frac{\sum n_i\cos\omega t_i}{\sum n_i} = 2\,\frac{\sum\cos\omega t_i}{N} = 2\langle\cos\omega t_i\rangle$$

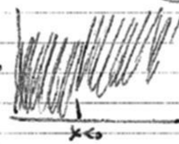
sparse time series $n_i = 0, 1$

Similarly $\beta = 2\langle\sin\omega t_i\rangle$

Stat mini #4

4

Risks of binning



Extreme example: 10000 points $y_i = 5 \pm 20$ ($x<0$)

" " $y_i = 7 \pm 20$ ($x>0$)

Average appropriately: $\bar{Y}_{x<0} = 5.0 \pm 0.2$

$\bar{Y}_{x>0} = 7.0 \pm 0.2$

fit constant $\Rightarrow \chi^2 = \left(\frac{5-6}{0.2}\right)^2 + \left(\frac{7-6}{0.2}\right)^2 = 50 \Rightarrow$ bad fit

$\langle (x_i - 6)^2 \rangle = \sigma^2 + 1$

w/ the 10000 × 2 points: $\chi^2 = \sum \left(\frac{x_i - 6}{\sigma}\right)^2 = \sum_i \frac{\sigma^2+1}{\sigma^2} = 20000 + \frac{20000}{\sigma^2} = 20050$

expected variance on $\chi^2 = \sqrt{2\chi^2} = \sqrt{40000}$

$\Rightarrow$ looks OK

But: if you were to try two-constant model,

you'd get $\Delta\chi^2 = 50$, which is signif.

BUT: converse: loss of information ...

For lightcurves, Gregory & Loredo nice Bayesian analysis

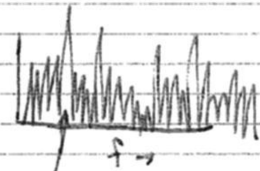
Number of trials

If you look for something 20 times, you're likely to have found a 2σ outlier!

Examples: { cross-identification
searches for periodicity

↓

Gregory & Loredo nice Bayesian overview



significant if you know the source varies at this frequency, not if your looking for variations

Source ID, error circle

How likely?

Equiv. to

$-\sigma_{\Delta x}^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2$

But should not just add $r^2 = \sigma_{\Delta x}^2 + \sigma_{\Delta\delta}^2$ & assume r/σ_Δ is normal

Far more area at larger sep!

This is Prior!

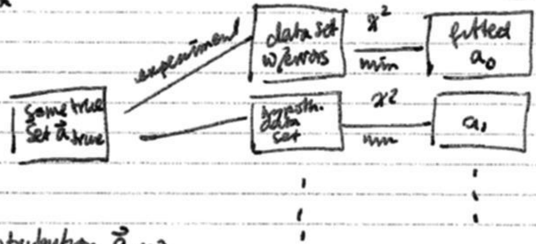
$$P_{\text{inside}} = \frac{1}{2\pi\sigma_{1D}^2}\int_0^r r\, e^{-\frac{1}{2}(r'/\sigma_{1D})^2}\, dr' = 1 - e^{-\frac{1}{2}(r/\sigma_{1D})^2}$$

Conversely r_{err} (some P) $= -2\sigma_{1D} \ln(1-P)$
say 95% conf. radius

r^2 (68.3%) = 2.3 σ_{1D}^2
95.4% 6.2 σ_{1D}^2
99.7% 11.6 σ_{1D}^2

Monte Carlo simulation

General case



get some probability distribution $\vec{a}_{(i)}$
of which a_0 is one member

Would like to know not PD of $\vec{a}$, but rather of $\vec{a}_{(i)} - \vec{a}_{true}$

How can we estimate this w/o knowing $\vec{a}_{true}$?

M.C. Suppose Nature was such that $\vec{a}_0$ = true ~~model~~ set

Hopefully, not too wrong. ASSUME PD$[a_{(i)} - a_0]$ = PD$[a_{(i)} - a_{true}]$

not assuming $a_0 = a_{true}$
just that random error enters in a way that does not vary
rapidly w/ a_{true}, so a_0 can serve as substitute

Now simulate data sets based on a_0, "measure" them & fit
Many times $\Rightarrow$ can get PD $(a_{(i)} - a_0)$

NOTE: one is free to choose weighting scheme & fitting technique
to optimize PD$(a_{(i)} - a_0)$!! (say, median vs least-squares)

Quick & Dirty: Bootstrap: take real data randomly,
with replacement

But data points need to be independent
(no measurement sequences where order is important)

No "clumpy" data

Stat mini #4

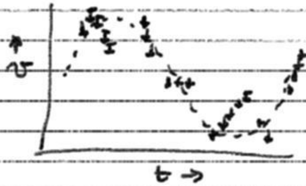
6

Monte-Carlo error estimating

How-to determine unc. for complicated cases?

Example: Vela X-1, measuring RV curve, I got

Correlated deviations



χ^2 does not work well

Can not just scale all errors, because data correlated.

Run M.C. w/ orbit + corr. deviations for 1 night long w/ random input amplitude, freq, phase (set by data)

What is difference?

Assume differences are random ⇒ can just scale errors

Ignoring correlation: think you have $N_{night} \times 5$ data points

With ": really only N_{night} data points

say

Normally use $\chi^2_{min}+1$ to infer errors

to force $\chi^2_{red}=1$, equivalent to scaling & using $\chi^2\left(1+\frac{1}{N_{dof}}\right)$

So, ignoring correlation, one underestimates unc. by $\sim \sqrt{5}$