



Astro 500



Techniques of Modern Observational Astrophysics

*Matthew Bershady
University of Wisconsin*

Lecture Outline

- Random vs systematic error
- Modeling data uncertainties: distributions
- Characterizing distributions: moments
- Errors, weighting and propagation
- Goodness of fit, χ^2 and robust estimation
- Regressions, error models and intrinsic scatter

Take away from last lecture:

- The notes contain a suitable reference for almost all of your future needs with optical-NIR photometry.
- Band-passes and calibration requires detailed information about filters, detector, telescope (optics), and atmosphere.
- Well-calibrated photometry is difficult to achieve.
 - 10%? You're not trying hard enough.
 - 5%? Ok.
 - 3%? Good. You are pressing the limit of the absolute calibration.
 - 1%? Don't fool yourself; you're limited by the absolute calibration.

Which brings up the topic of precision vs accuracy...

Random vs systematic error

- Same as the difference between precision vs accuracy
 - Precision: How well can you measure a quantity (what's the variance of repeat measurements)?
 - Accuracy: How well do your measurements (in the mean) reflect the value you are trying to measure?
 - *Know the difference – fundamental.*
- We're lucky when an astronomical result includes a quote of random error. Systematic errors in astronomy are rarely addressed, but this is changing; it has long been commonplace in Physics.
- In error handling and reporting, be a physicist in rigor with the intuition of an astronomer.

Modeling Uncertainties: Distributions

- Binomial – coin-toss
 - Poisson – generalized counting statistics (\sqrt{N})
 - Gaussian – “normal” distribution, appropriate for most error distributions
 - Lorentzian – suitable to describe data related to resonances
-
- Processes in the universe do not “obey” these distributions, they are merely human-generated, mathematical contrivances.
 - Amazingly, a number of processes in the universe do appear to be well described by these distributions

Characterizing distributions

- Moments

- 1st: Mean, \bar{x} (location)

- Other 1st-moment indicators:

- *median* (robust estimator)
 - *mode*

- 2nd: Standard deviation, σ (width)

- Other 2nd-moment indicators:

- *Average deviation* (robust estimator): $AD = \frac{1}{N} \sum_{i=1}^N |x_i - \bar{x}|$
 - *full-width half-maximum (FWHM)*

- 3rd: Skew, s (symmetry)

- 4th: Kurtosis, k (shape)

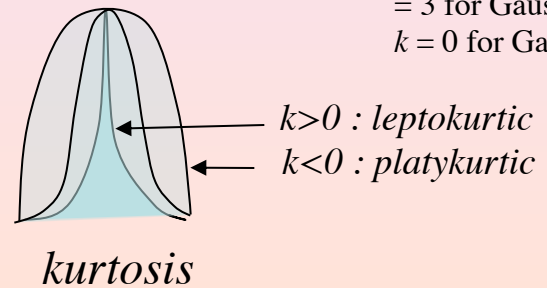
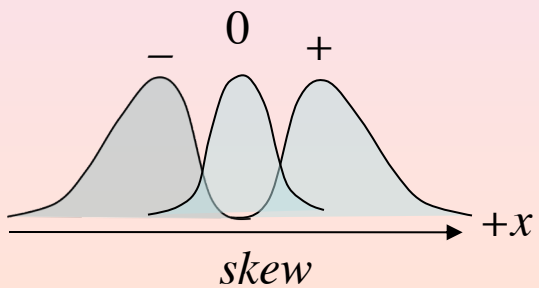
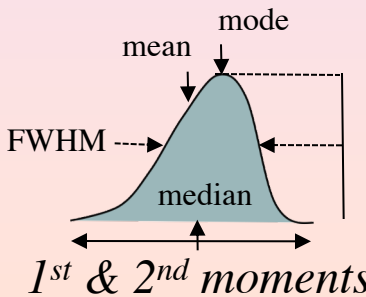
$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

$$s = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$$

$$k = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^4 - 3$$

$\underbrace{\hspace{10em}}$
 = 3 for Gaussian
 k = 0 for Gaussian



Gaussian distribution

- By adopting a distribution function you can relate moments to distribution function parameters
- You can also invert the process, compare moments and determine the suitability of model distribution choice
- For a Gaussian distribution the following holds:
 - 1st: mean=mode=median
 - 2nd: $2.354 \sigma = FWHM$
 - 3rd: Skew, $s = 0$ (*symmetric*)
 - 4th: Kurtosis, $k = 0$ (*normal*)
- This is true strictly *only* for a Gaussian.

Standard deviation vs standard error

- σ is the *standard deviation*. It describes the width of a distribution, but is only proportional to the error in the estimate of the mean.
- $\sigma_{\bar{x}}$ is the *standard error in the mean*, \bar{x} . It depends on the distribution width and how well it is sampled, i.e., N .

- In general:

$$\sigma_{\bar{x}} = \sigma / \sqrt{N}$$

regardless of the distribution shape.

Weighting your data

- Moment estimates of a distribution can be dramatically improved in *precision* by weighting the data according to the uncertainties specific to each datum (heteroscedastic errors)
- $w_i \equiv 1/\sigma_i^2$
- σ_i is the variance on each datum x_i
- Similar generalizations for s and k .
- And it is still true:

$$\sigma_{\bar{x}_w} = \sigma_w / \sqrt{N}$$

$$\bar{x}_w = \frac{\sum_{i=1}^N (w_i x_i)}{\sum_{i=1}^N w_i}$$
$$\sigma_w^2 = \frac{N}{N-1} \frac{\sum_{i=1}^N [w_i (x_i - \bar{x})^2]}{\sum_{i=1}^N w_i}$$

- Look familiar?

$$\lambda_{\text{eff}} \equiv \frac{\int \lambda S_{\lambda} f_{\lambda} d\lambda}{\int S_{\lambda} f_{\lambda} d\lambda},$$

w_i

Error propagation

- Gaussian quadrature or N-dimensional hypotenuse
- When errors on individual parameters, or components, x_i , of your observable, f , are uncorrelated (orthogonal):

$$f = f(x_1, \dots, x_n)$$
$$df = \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} dx_i \right)^2}$$
$$dx_i = \sigma_{x_i}$$

- With covariance:

$$df = \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} dx_i \right)^2 + 2 \sum_{j \neq i} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} dx_{ij}}$$
$$dx_{ij} = \sigma_{ij}$$

- Best to work in terms of covariance matrix

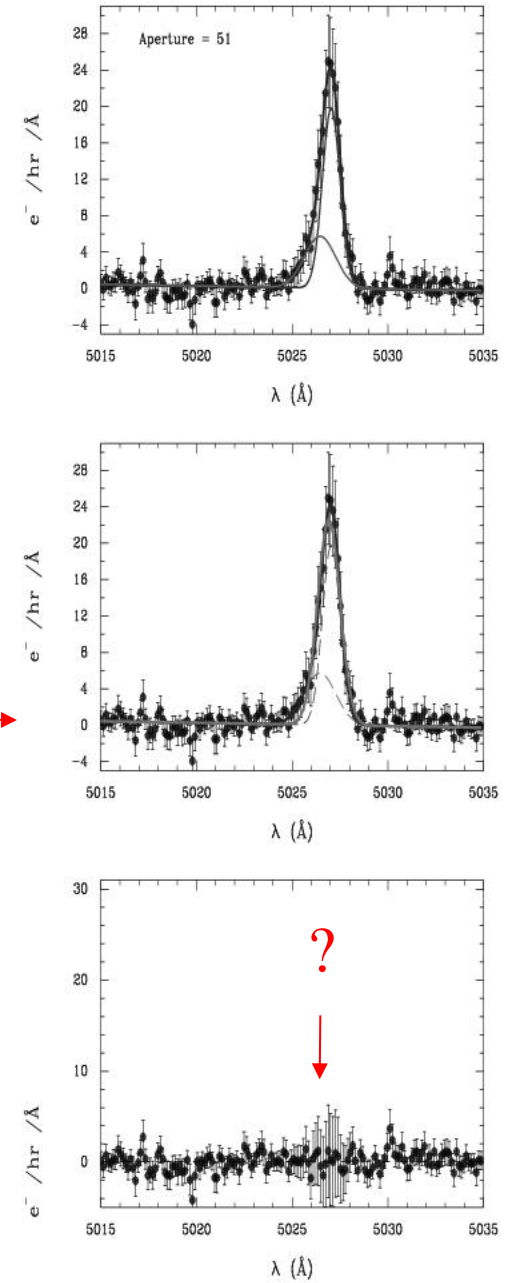
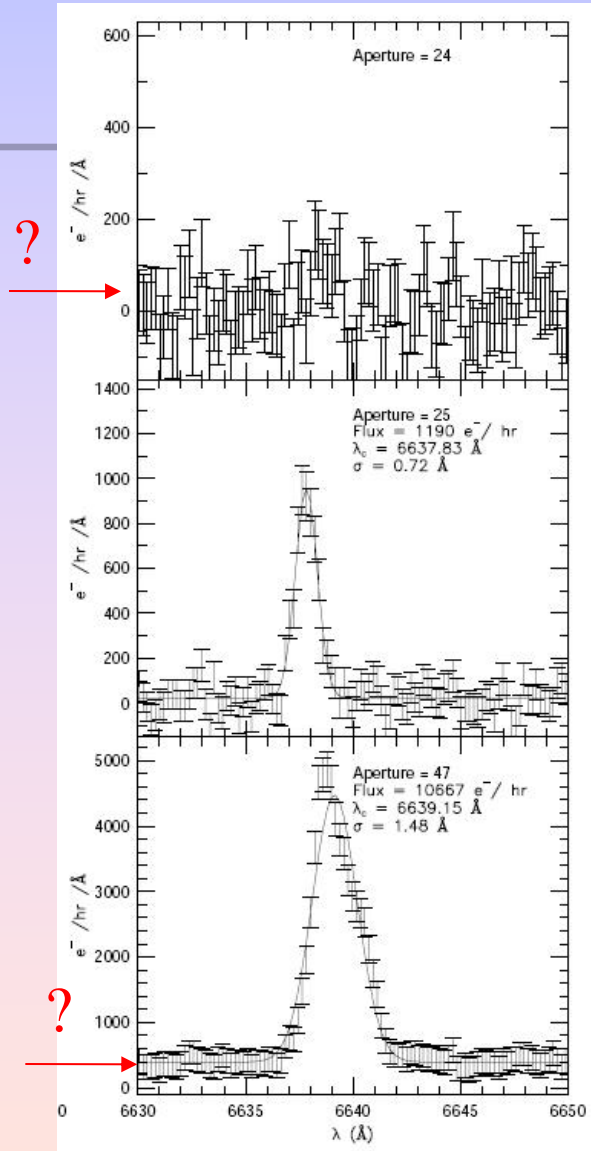
An aside: logarithmic derivatives

- The best way to present errors or to examine covariance:
- $d \log x = \frac{dx}{x}$ is scale free and gives the fractional error
- $\frac{d \ln x}{d \ln y} = \frac{y}{x} \frac{dx}{dy}$
- “neper-esque”

Are your errors any good?

- Look at your data and think about what it means.
- Think in terms of probability, specifically the probability distribution suitable for your adopted data model (typically Gaussian).
- For a Gaussian distribution of errors you expect
 - 68% of data lie within $\pm 1\sigma$
 - 95% of data lie within $\pm 2\sigma$
 - 99.7% of data lie within $\pm 3\sigma$

How do these look?



What assumptions do you make in your assessment?

Goodness of fit: χ^2

- Formal definition:
- Reduced- χ^2 : $= \chi_v^2 = \chi^2 / \nu$

$$\chi^2 = \sum_{i=1}^N \left\{ \left[y_i - y(x_1, \dots, x_n) \right]^2 / \sigma_i^2 \right\}$$

- y_i = data
- σ_i = errors on data
- $y(x_1, \dots, x_n)$ = model as function of parameters (x_1, \dots, x_n)
- $\nu = N - n - 1$, where n are the number of model parameters

a 2nd moment,
error-weighted

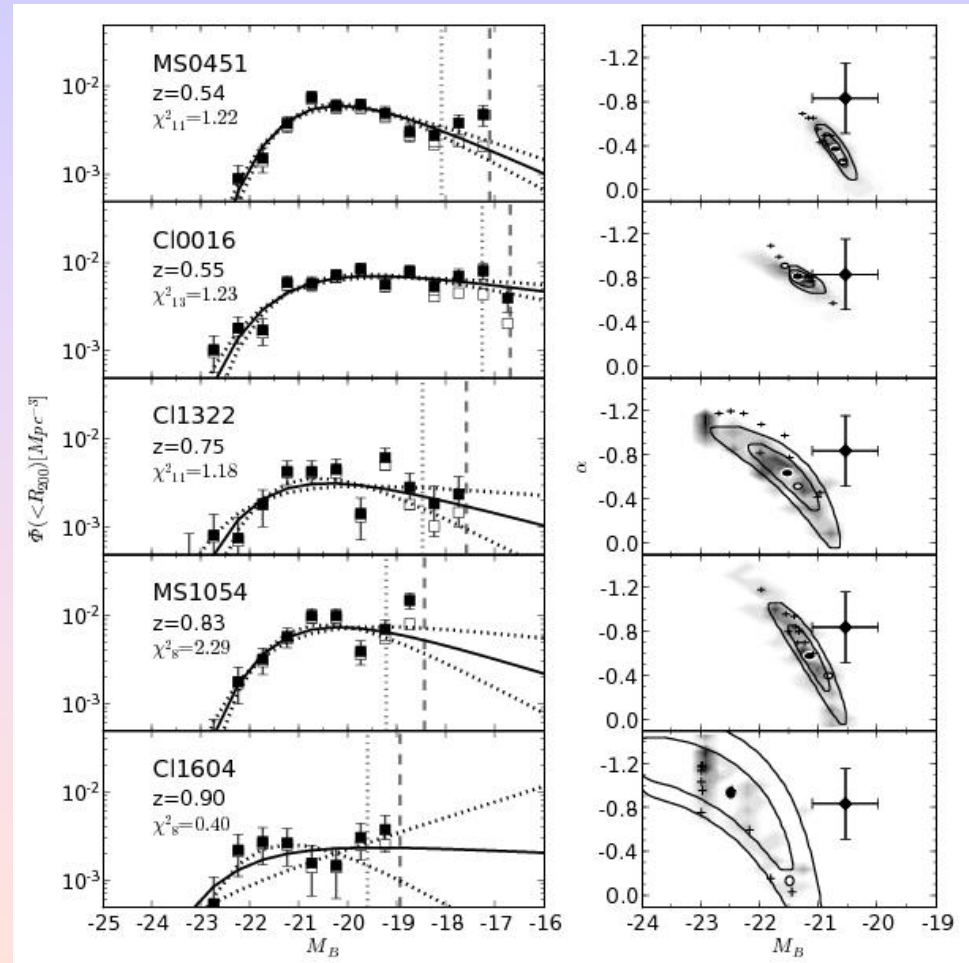
- What does it mean? For a good model and well defined errors you expect: $\chi_v^2 = 1$.
- χ^2, χ_v^2 useful to estimate goodness of fit of model as well as for defining *confidence intervals* on model parameters by using probability distribution.
 - Some assumptions and limitations apply. See Bevington and Press et al. (Num.Rec.) for discussion.
- Must have good error estimates.

Goodness of fit: example

- Schechter luminosity function fitting

$$\phi(M) dM = 0.4 \ln 10 \times \phi^* 10^{-0.4(M-M^*)(\alpha+1)} \times \exp \left[-10^{-0.4(M-M^*)} \right] dM.$$

- *lhs*:
 - best fit (solid curve)
 - 68% CL fits (dotted curves)
- *rhs*
 - 65, 95% joint confidence intervals on α - M^* (contours)
 - Jack-knife errors (greyscale)



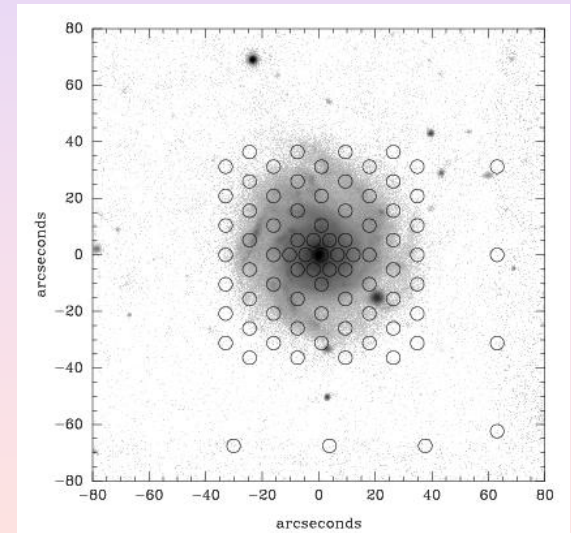
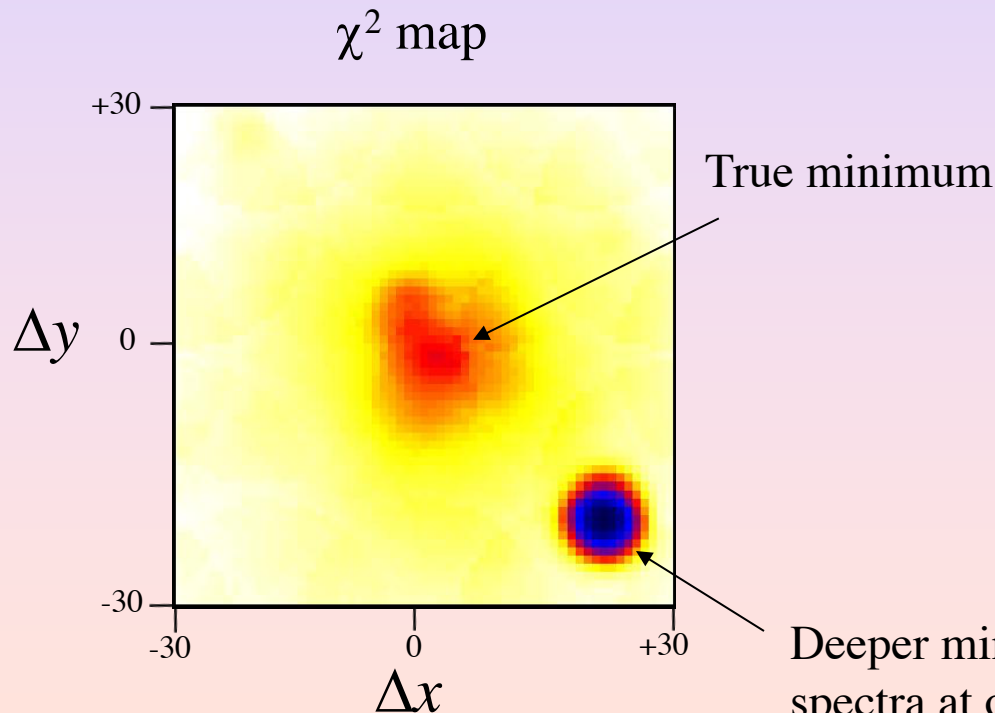
Crawford et al. '09 (ApJ, 690, 1158)

Minimization: χ^2

- Several methods for finding the model parameters yielding the minimum χ^2 for your data (the “best fit”).
 - See Press et al. (Num.Rec.)
- Brute force – just map it out over N-dimensional parameter space
- Downhill simplex: N+1 dimensional “amoeba” steps toward minimum
 - Requires no function derivatives
 - Easy to implement
 - slow
- Levenberg-Marquardt: smooth variation from inverse-Hessian to steepest-descent methods
 - Standard non-linear least-squares routine
 - Computationally fast
 - Yields covariance matrix for parameter error estimation
 - Requires function derivatives
- All methods suffer from trapping in false (local) minima.
- Must looking at χ^2 maps to assess parameter covariance and uniqueness

χ^2 : False minima

- Example: Spatial registration of SparsePak (sparsely-sampled IFU) spectral-continuum data to 2D broad-band images
- Which minima correctly identifies the centering?



Deeper minimum due to star in image not in IFU spectra at offset co-incident with galaxy center.

Outside the box

- What does it mean if $\chi_v^2 > 1$ or $\chi_v^2 < 1$?

- Under or over estimated your errors
- Bad model

➤ Can still use χ^2 for confidence-interval estimation on model parameter uncertainties, but if the model or errors are bad then so too are the uncertainties.

- What *else* can you do?

- Force $\chi_v^2 = 1$ by adjusting errors in an “unbiased” way

$$\chi^2 = \sum_{i=1}^N \left\{ \left[y_i - y(x_j, \dots, x_n) \right]^2 / \left(\sigma_i^2 + \sigma_m^2 \right) \right\}$$

added term

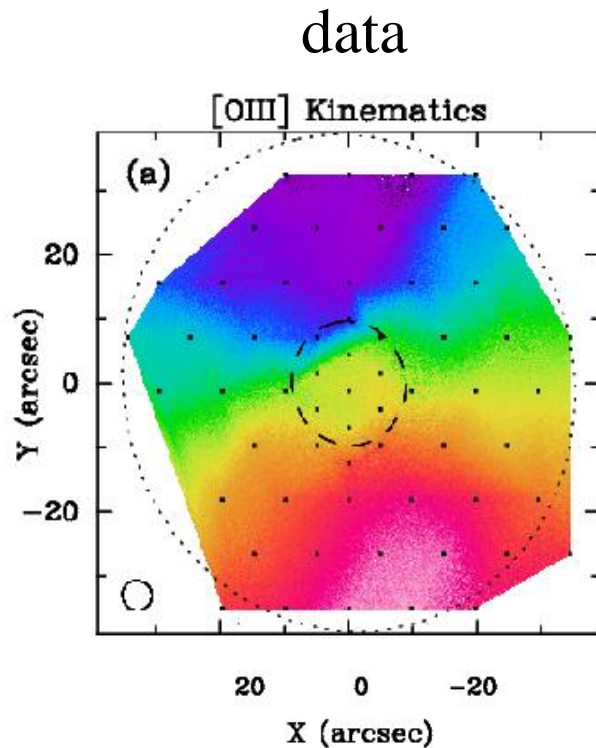
- σ_m is a constant “model error” for all data y_i , chosen to make $\chi_v^2 = 1$.

- χ^2 can be used for confidence-interval estimation on model parameter uncertainties with better fidelity.
- When can you do this, i.e., when is it ok to “fudge it?”
- Is this voodoo statistics or are we ahead of the statisticians?

WARNING:
no commonly
accepted norm

Outside the box: example

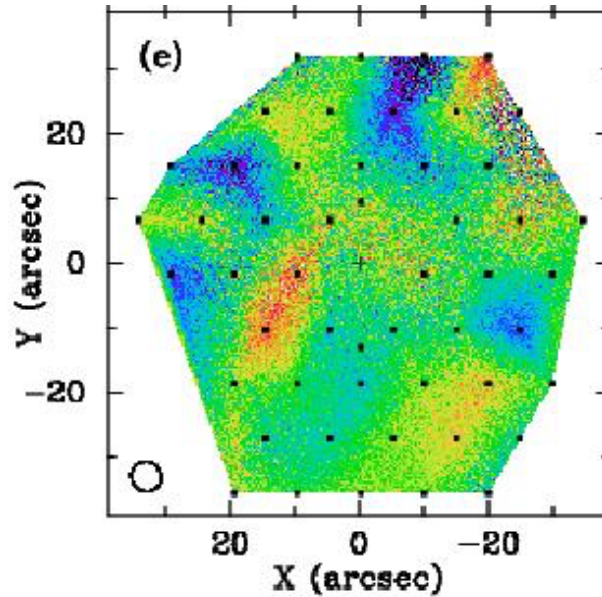
- Ionized-gas velocity fields of galaxy disks
 - Andersen (2001, Ph.D.), Westfall (2009, Ph.D.)



PLAW

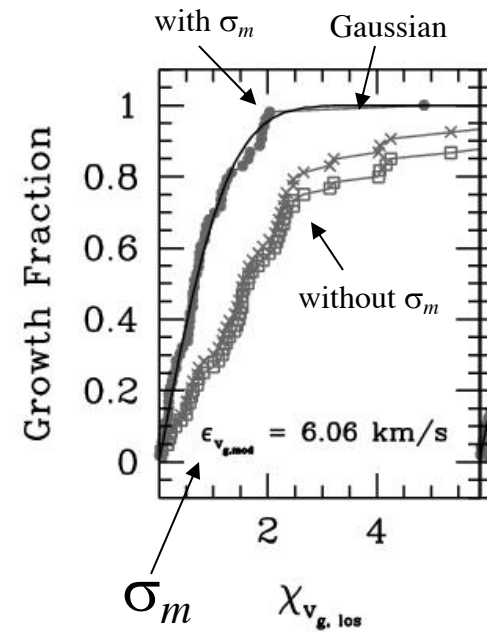
1050 1100 1150 1200
 v_{los} (km s⁻¹)

data - model



-10 0 10
 $v_{\text{obs}} - v_{\text{mod}}$ (km s⁻¹)

χ^2 distribution

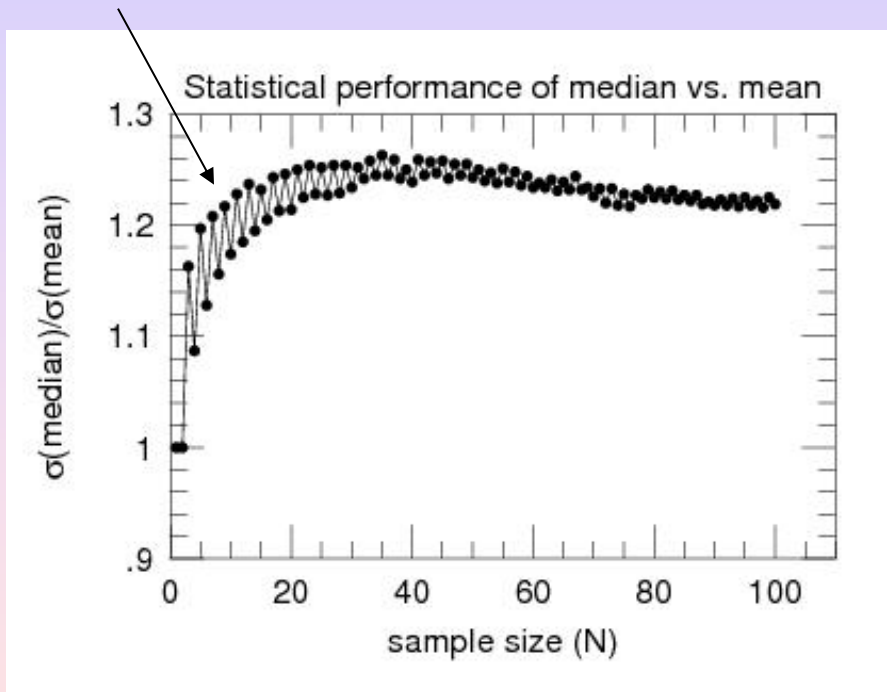


Robust estimation

- Usually trade-offs between precision and accuracy
- Robust estimators typically improve accuracy at the sacrifice of precision.
- When to make this trade is a judgment call based on the quality and “issues” associated with your data.
- You should understand (i.e., model) the amplitude of the trade-off.
- Two examples:
 - (1) median vs. mean
 - (2) σ -clipping

Robust estimation: median

Why fluctuations?



Experiment: compute variance (σ) in mean and median estimated for N data points sampling the same intrinsic Gaussian distribution (same parent mean and σ).

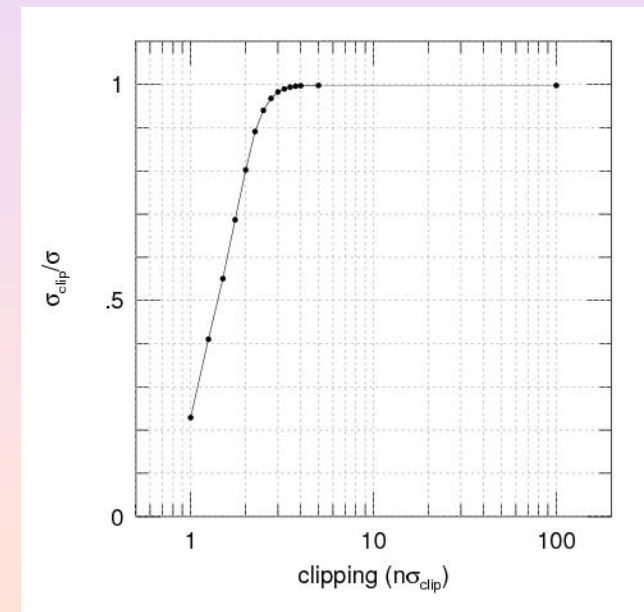
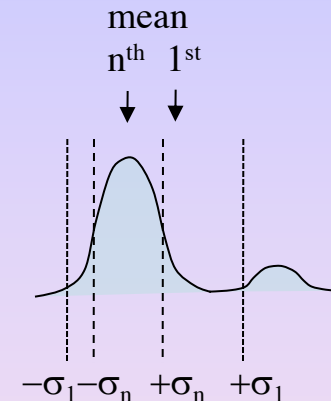
- The median is the 50th-percentile of a sorted list.
- The median is a robust estimator of the mean because it is relatively impervious to outlier data, e.g., a non-Gaussian tail.
- As long as the majority of data cluster around the true mean, then the median is more reliable (i.e., *more accurate*) than the mean.
- However, the median is 20-25% noisier than the mean when the distribution is close to Gaussian, i.e., it is *less precise*.

Robust estimation: MAD

- Median absolute deviation: a robust estimator of the 2nd moment as the median is for the 1st moment
- $\text{MAD} = \text{median}(|x_i - \bar{x}|)$
- $\sigma = k * \text{MAD}$
- $k = 1.4826$ for a normal distribution

Robust estimation: σ -clipping

- Iterative-clipping of moments: good for mean and σ
 - Also applies to χ^2
- Choose a clipping level
- Define a convergence criterion
- Skew is good choice for assumption of normal distribution ($s=0$)
 - Can check assumption by calculating kurtosis
- Note correction factor needed to complete integral for both σ and k (even moments)

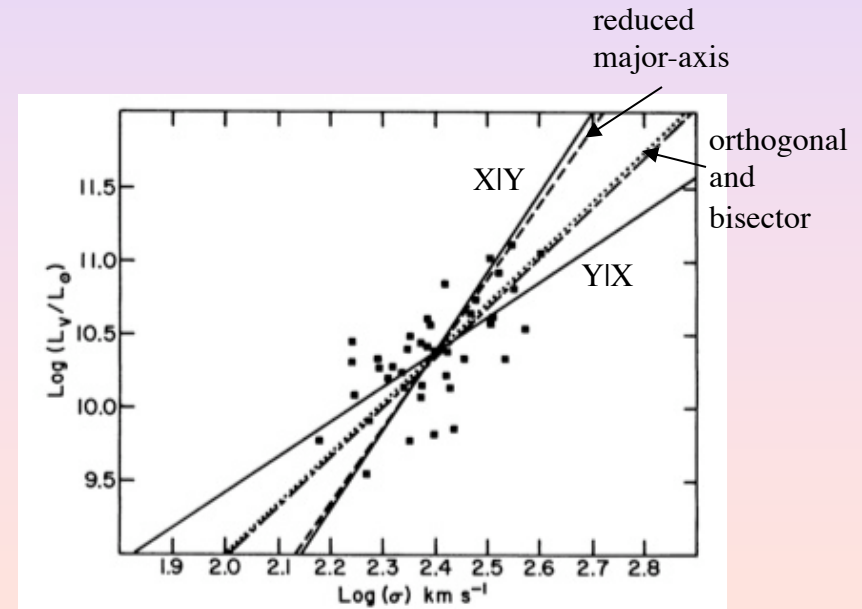
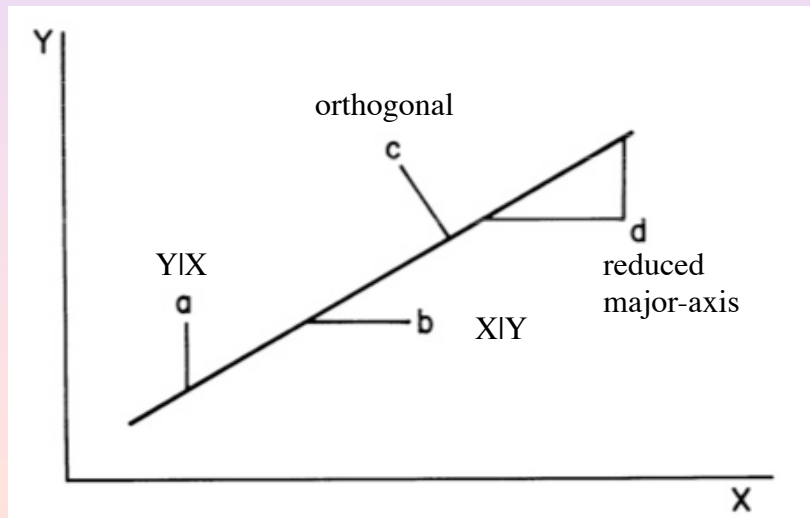


Linear Regressions

- Regressions are based on solving a set of linear equations based on different moments of the data and weighted by errors or priors.
- There are different kinds of regression models (moments)
 - $X|Y$, $Y|X$, bisector, orthogonal
 - There are different assumptions to be made about errors that also lead to different moments and regressions
 - Is there an independent variable? (one variable with no errors)
 - Are the errors heteroscedastic or homoscedastic? (different or the same for all data)
 - Is there intrinsic scatter (usually other dimensions not known)?
- *There is no right regression model (it depends what you want to learn), but there are correct and incorrect errors models and assumptions.*
 - Social science analysis is plagued by systematic errors due to inaccurate models, but we're not free of such pitfalls because the universe is complicated.

Different Regressions

- $X|Y$, $Y|X$, bisector, orthogonal regressions
- Isobe et al. (1990, ApJ, 364, 104):
 - ordinary least squares (OLS) – no errors
- Akritas & Bershady (1996, ApJ, 470, 706)
 - bivariate correlated errors (heteroscedastic) and intrinsic scatter (BCES)



OLS Regression formulae

TABLE 1
LINEAR REGRESSION FORMULAE FOR SLOPES

| Method | Expression for Slope | Estimate of the Variance of the Slope $\widehat{\text{Var}}(\beta_i)$ |
|-----------------------------|--|---|
| OLS($X Y$) | $\beta_1 = \frac{S_{xy}}{S_{xx}}$ | $\frac{1}{S_{xx}^2} \left[\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \beta_1 x_i - \bar{y} + \beta_1 \bar{x})^2 \right]$ |
| OLS($Y X$) | $\beta_2 = \frac{S_{yx}}{S_{xy}}$ | $\frac{1}{S_{xy}^2} \left[\sum_{i=1}^n (y_i - \bar{y})^2 (y_i - \beta_2 x_i - \bar{y} + \beta_2 \bar{x})^2 \right]$ |
| OLS bisector | $\beta_3 = (\beta_1 + \beta_2)^{-1} [\beta_1 \beta_2 - 1 + \sqrt{(1 + \beta_1^2)(1 + \beta_2^2)}]$ | $\frac{\beta_3^2}{(\beta_1 + \beta_2)^2 (1 + \beta_1^2)(1 + \beta_2^2)} [(1 + \beta_2^2)^2 \widehat{\text{Var}}(\beta_1) + 2(1 + \beta_1^2)(1 + \beta_2^2) \widehat{\text{Cov}}(\beta_1, \beta_2) + (1 + \beta_1^2)^2 \widehat{\text{Var}}(\beta_2)]$ |
| Orthogonal regression | $\beta_4 = \frac{1}{2} [(\beta_2 - \beta_1^{-1}) + \text{Sign}(S_{xy}) \sqrt{4 + (\beta_2 - \beta_1^{-1})^2}]$ | $\frac{\beta_4^2}{4\beta_1^2 + (\beta_1 \beta_2 - 1)^2} [\beta_1^{-2} \widehat{\text{Var}}(\beta_1) + 2 \widehat{\text{Cov}}(\beta_1, \beta_2) + \beta_1^2 \widehat{\text{Var}}(\beta_2)]$ |
| Reduced major-axis | $\beta_5 = \text{Sign}(S_{xy}) (\beta_1 \beta_2)^{1/2}$ | $\frac{1}{4} \left[\frac{\beta_2}{\beta_1} \widehat{\text{Var}}(\beta_1) + 2 \widehat{\text{Cov}}(\beta_1, \beta_2) + \frac{\beta_1}{\beta_2} \widehat{\text{Var}}(\beta_2) \right]$ |

Errors on Regressions

- How do you estimate errors on slope and intercept?
 - Resample your data:
 - Boot-strap – pick N data points out of sample of N , m times. Each pick is a random selection from N data points with equal probability of selecting i^{th} element.
 - Jack-knife – recalculate leaving out one datum, N times (N data)
 - Monte Carlo simulation – artificial data
-
- ✧ When in doubt, “Monte Carlo” your data
 - ✧ This applies not just to linear regressions but any modeling.

When in doubt....

- “Monte Carlo (MC) your data”
- Monte Carlo: a town in Monaco (country in SE France) famous for gambling casinos
- What you need:
 - Model of data
 - Model of errors
 - Model of data sampling (range, censorship, incompleteness, spurious source (when applicable)).
 - A good random-number generator
 - A modicum of computing skill and cpu time.
- How good is it?
- Only as good as your assumptions (i.e., model)
- Test your assumptions by comparing distributions (and their characterization) generated by MC against those from the data.

Miscellaneous topics in statistics

- Items you need to explore on your own as an astronomer:
 - Markov-chain Monte Carlo (MCMC)
 - How do distribution- and selection-functions bias regressions?
 - Famous biases in astronomy
 - Malmquist
 - Eddington
 - What linear regression should you use? ...or:
 - Should you even use a linear regression?