Calibrating SN Ia Standard Candles

suppose have already calibrated cepheids

function of distance: $\mu = m - M$ ← abs. mag

$s = 1, \dots, N$ measurements, $\hat{\mu}_s$, $\sigma_{\text{cepheid},s}$ from cepheids

$\hat{m}_s$, $\sigma_{m,s}$ from SN Ia

$\Rightarrow \hat{M}_s = \hat{m}_s - \hat{\mu}_s \Rightarrow \sigma_{\text{err},s}^2 = \sigma_{m,s}^2 + \sigma_{c,s}^2$ assume σ's known

(assume $\hat{\mu}_s$, $\hat{m}_s$ independent)

- assume $M_s \sim N(M_0, \sigma_{\text{int}}^2)$

each supernova drawn from normal distribution | σ_{int} unknown internal error

[could also say $M_s = M_0 + \varepsilon_s^M$, and $\varepsilon_s^M \sim N(0, \sigma_{\text{int}}^2)$]

so the errors in means normally dist. around true value

- Calibraters: SN $\hat{m}_s \sim N(m_s, \sigma_{m,s}^2)$ ← true values — LATENT VARIABLES

Cepheid $\hat{\mu}_s \sim N(\mu_s, \sigma_{\mu,s}^2)$

[could also write $\hat{m}_s = m_s + \varepsilon_{m,s}$ $\hat{\mu}_s = \mu_s + \varepsilon_{\mu,s}$] ← this is how you'd generate data by code

where $\varepsilon_{m,s} \sim N(0, \sigma_{m,s}^2)$ $\varepsilon_{\mu,s} \sim N(0, \sigma_{\mu,s}^2)$

- Latent variable: $M_s = m_s - \mu_s$, use estimater $\hat{M}_s = \hat{m}_s - \hat{\mu}_s$

$\Rightarrow \hat{M}_s \sim N(M_s, \sigma_{\text{err},s}^2 = \sigma_{\mu,s}^2 + \sigma_{m,s}^2)$

assume M_s, $\sigma_{\mu,s}$ known, measurement errors and $\varepsilon_{m,s} \perp \varepsilon_{\mu,s}$ indep

$$P(\{\hat{M}_s\} \mid M_0, \sigma_{\text{int}}^2) = \prod_{s=1}^{N} P(\hat{M}_s \mid M_0, \sigma_{\text{int}}^2)$$

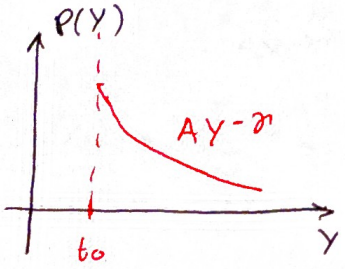
$$= \prod_{s=1}^{N} N(\hat{M}_s \mid M_0, \sigma_{\text{int}}^2 + \sigma_{\text{err},s}^2)$$

$\hat{M}_s \sim N(M_0, \sigma_{\text{int}}^2 + \sigma_{\text{err},s}^2)$

can't actually take analytic derivatives bc won't be able to solve for $\hat{\sigma}_{\text{int}}$ ⇒ need to solve numerically

Inference of Stellar Mass Function

Y_i = mass of star i $\quad Y_i \sim P(Y_i, \alpha) = \begin{cases} 0 & Y_i < t_0 \\ A Y_i^{-\alpha} & Y \geq t_0 \end{cases}$



$\Rightarrow L(\alpha) = \prod_{i=1}^{N} P(Y_i | \alpha)$

- suppose due to some effect couldn't see stars above some mass

SELECTION EFFECT in this case truncation at t_1

→ introduce indicator $I_i = \begin{cases} 0 & \text{otherwise} \\ 1 & \text{if } Y_i \text{ observable} \end{cases}$

$\Rightarrow P(I_i = 1 | Y_i, t_1) = \begin{cases} 1 & Y \leq t_1 \\ 0 & Y > t_1 \end{cases}$

- Modified likelihood function:

$$P(Y_i^{obs} | I_i = 1, \alpha) = \frac{P(I_i = 1, Y_i | \alpha)}{\int dy_i \, P(I_i = 1, y_i | \alpha)} = \frac{P(I_i = 1 | Y_i) P(Y_i | \alpha)}{\int dY_i \, P(I_i = 1 | Y_i) P(Y_i | \alpha)}$$

see example sheet for solution, this LKHD gives true value while naive LKHD overestimates α

QUANTIFYING UNCERTAINTY

- Suppose $\vec{X} = (X_1, \dots, X_N) \overset{iid}{\sim} Pe(x)$

$\vec{x}_{obs} = (x_1^{obs}, \dots, x_N^{obs})$, $\hat{G}(\vec{x}^{obs}) \pm \sqrt{Var(\hat{G})}$

- For case $x_i \overset{iid}{\sim} N(\mu, 1)$ $\quad \vec{x}^{obs} = (-0.639, -0.9313, 0.1577, -0.8813)$

$\bar{x}^{obs} = -0.5735$ $\quad \sigma_{\bar{x}} = 1/4 \Rightarrow \sigma_{\bar{x}} = 0.5$, $\quad \bar{X} \sim N(\mu, \sigma^2/N)$

$\Rightarrow \bar{X} \pm \bar{\sigma}_x = [-1.0735, -0.0735]$

This is a 68% confidence interval

- what does a confidence interval mean?

from a probability prospective (frequentist) only random variables have probabilities, so $P(\mu \in [\,,\,]) =$ ... doesn't make sense because freq. takes μ as given, not a rv.

⇒ two estimaters! $[L(\vec{X}), U(\vec{X})]$ function of rv's → rv's

$= C(\vec{X})$

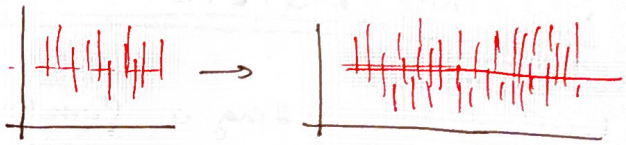
$P(C(\vec{X}) \text{ contains } \mu) \geq (1-\alpha)$ is an β percent confidence

$\geq \beta$ interval

the interval is random!

e.g.: run an experiment lots of times:

as $\# \to \infty$ the probability of μ being in confidence interval goes to above



- More complicated cases:

$\vec{X} \sim P_\theta(\vec{x})$, $\hat{q}(\vec{X})$ not a simple function $\mathrm{Var}(\hat{q}(\vec{X}))$ is intractable

- What could you do? above method worked bc gaussian, have not

Simulate: $\vec{X}_{sim,1}, \ldots, \vec{X}_{sim,M}$ ⇒ $\mathrm{Var}(\{\hat{q}_1, \ldots, \hat{q}_M\}) = \frac{1}{M-1}\sum_{m=1}^{M}(q_m - \bar{q})^2$

$\downarrow$ $\downarrow$

$\hat{q}_1$ $\hat{q}_M$

but a lot of times can't do / too hard / model incorrect

- Bootstrap

suppose have $\vec{X} = (X_1, \ldots, X_5)$

suppose real data that observe $\vec{x}^{obs} = (3, 8, 2, 4, 5)$

→ want to estimate skewness $\mathrm{skew}(X_i) = \frac{E[(X_i-\mu)^3]}{(\sigma^2)^{3/2}}$ third moment

- sample skewness $= \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)^3}$

⇒ create bootstrapped datasets — just drawing from $\vec{X}_{obs}$ — sampling with replacement

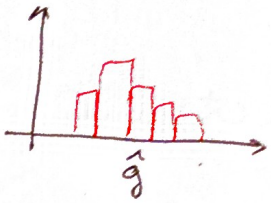
$\vec{X}_{boot,1} = (2, 5, 4, 4, 4)$

$\vec{X}_{boot,2} = (2, 4, 2, 8, 8)$

⋮

pretend these are equivalent to datasets that observed since you could have observed these numbers

⇒ $\hat{g}_1 = \hat{g}(\vec{X}_{boot,1})$, $\hat{g}_2 = \hat{g}(\vec{X}_{boot,2})$, etc



FITTING DISTRIBUTIONS

REGRESSION: fitting a function $E(y|x) = f(x;\theta)$ for mean relation between y and x

- Basic approaches: OLS, GLS, WLS, ML — real problems need more complicated modelling

Linear regression — fitting function is linear

- cepheids → lightcurve is time series

want to find relation of period-luminosity

$\underline{Y} = \underline{\underline{X}}\underline{\beta} + \underline{\varepsilon}$ $\quad E(\varepsilon_i) = 0$ $\quad Var(\varepsilon_i) = \sigma^2$ known — $\underline{\underline{X}}$ matrix

minimise wrt β: $RSS = (\underline{Y} - \underline{\underline{X}}\underline{\beta})^T(\underline{Y} - \underline{\underline{X}}\underline{\beta})$

↑ residual sum of squares

take derivative wrt β ⇒ $\hat{\beta}_{OLS} = (\underline{\underline{X}}^T\underline{\underline{X}})^{-1}\underline{\underline{X}}^T\underline{Y}$

if all assumptions are true can show $E(\hat{\beta}_{OLS}) = \beta$ so unbiased

$\hat{\sigma^2} = \frac{1}{N-k}(\underline{Y} - \underline{\underline{X}}\underline{\beta})^T(\underline{Y} - \underline{\underline{X}}\underline{\beta})$

Weighted Least squares — χ^2 minimisation

for heteroscedastic errors

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - \beta_0 - \sum \beta_j x_{ij})^2}{\sigma_i^2}$$

$$\underline{Y} = \underline{\underline{X}}\,\underline{\beta} + \underline{\varepsilon} \qquad \mathbb{E}(\varepsilon_i) = 0 \quad i = 1, \dots, N$$

$\underline{\underline{X}}$: $N \times k$ matrix

$\chi^2 \sim \chi^2_{N-k}$ χ^2 random variable with $N-k$ degrees of freedom

$\mathbb{E}(q\chi^2_{N-k}) = N-k$ so $\frac{\chi^2}{N-k} \approx 1$ for large $N-k$

ie if have enough data

- e.g.: even for 0 measurement error there could be intrinsic dispersion

- Special cases: correlated error $\mathrm{Var}(\underline{\varepsilon}) = \mathrm{Cov}[\underline{\varepsilon}, \underline{\varepsilon}^T] = \underline{\underline{W}}$ known

$$\Rightarrow RSS = (\underline{Y} - \underline{\underline{X}}\,\underline{\beta})^T \underline{\underline{W}}^{-1} (\underline{Y} - \underline{\underline{X}}\,\underline{\beta})$$

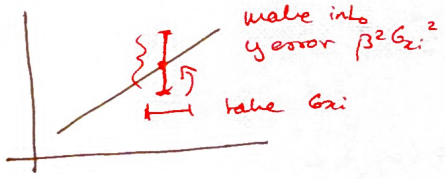
can think in terms of ML! assuming Gaussian with covariance matrix $\underline{\underline{W}}$

$$L(\underline{\beta}) = P(\underline{Y} \mid \underline{\beta}, \underline{\underline{X}}) = N(\underline{Y} \mid \underline{\underline{X}}\,\underline{\beta}, \underline{\underline{W}})$$

if $\underline{Y} \sim N(\underline{\underline{X}}\,\underline{\beta}, \underline{\underline{W}})$

maximise ⇒ get previous results

- Example: FITEXY estimator idea



physicist $\chi^2 = \sum \frac{(y_i - \alpha - \beta x_i)^2}{\sigma_{y,i}^2}$

Tremaine et al. 2002 add measurement error σ^2 to denominator as well $\frac{1}{\sigma^2 + \sigma_{y,i}^2 + \beta^2 \sigma_{x,i}^2}$

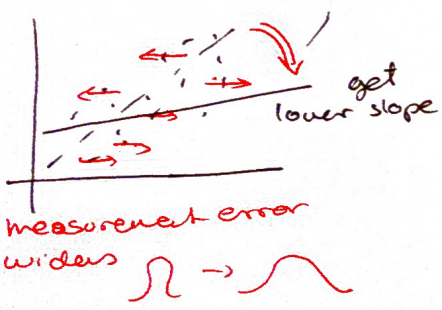
Bad bc could make σ^2 huge to min.

Kelly et al. 2017

⇒ OLS always underestimated

FITEXY always overestimated

But MLE is unbiased



Probabilistic Generative Modelling

Start with conceptual framework: FORWARD MODEL

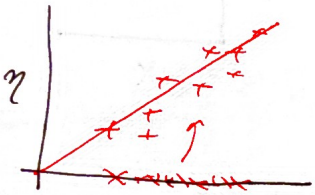
can introduce latent variables $\underline{\alpha}$ along the way intermediate

⇒ get sampling distribution $P(D|\underline{\theta}) = \int P(D|\underline{\alpha})\, P(\underline{\alpha}|\underline{\theta})\, d\underline{\alpha}$

could integrate out "things that didn't see" — measurement model

⇒ using observed D, draw inferences about $\underline{\theta}$ from $\mathcal{L}(\underline{\theta}) = P(D|\underline{\theta})$

- quasars:



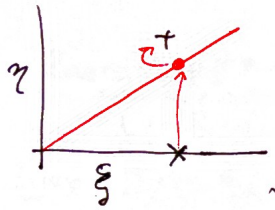
② generate slope and map ξ to them

③ add errors and get η

with model

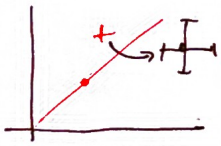
① generate 100 draws of $\xi \sim N(\mu, \tau^2)$

For one:



$\xi \sim N(\mu, \tau^2)$

$\varepsilon \sim N(0, \sigma^2)$



$P(\eta_i, \xi_i | \alpha, \beta, \sigma, \mu, \tau)$

$= P(\eta_i | \xi_i, \alpha, \beta, \sigma) \cdot P(\xi_i | \mu, \tau)$

get latent variables from pop distribution

Now get observed data: $P([x_i, y_i] | \eta_i, \xi_i) = N([x_i, y_i] | [\eta_i, \xi_i], \underline{\underline{\Sigma}})$:

$$\underline{\underline{\Sigma}} = \begin{pmatrix} \sigma_y^2 & 0 \\ 0 & \sigma_x^2 \end{pmatrix}$$

$\underline{\theta} = (\alpha, \beta, \sigma^2)$ reg. param.

$\underline{\Psi} = (\mu, \tau)$ indep variable "hyperparam"

(ξ_i, η_i) latent (true) variables

(x_i, y_i) data

① $\xi \sim N(\mu, \tau^2)$

② $\eta_i | \xi_i$

⇒ formulate likelihood function by integrating out latent variables