Pragmatic and Fully Bayesian Approaches

David A. van Dyk

Department of Mathematics, Statistics Section
Imperial College London

BIRS – Sytematics & Nuisance Parameters, April 2023

Imperial College
London

Disclaimer

My perspective is informed by:

- I am a Statistician
- I have worked with astrophysicists developing statistical methodology for over 25 years
- I'm a Bayesian Statistician *...but not overly so.*

Statisticians do not always agree on everything.

Some bits are rather philosophical.

I don't speak for ALL statisticians!

Systematics and Multi-Stage Analyses

My Interest in Systematics stems from Astrostatistics

- Massive new data streams allow explicit modelling of detailed physical processes.
- Often modularized into a chain of data analyses.
- Each conducted by different researchers with different data, assumptions, methods, expertise, etc.
- Output for one analysis is input for subsequent analyses.

Systematics in Physics

- Primary analysis involves nuisance parameters estimated with error in a Preliminary analyses.

Can we combine into principled omnibus analysis?
How do we properly quantify uncertainty?

Outline

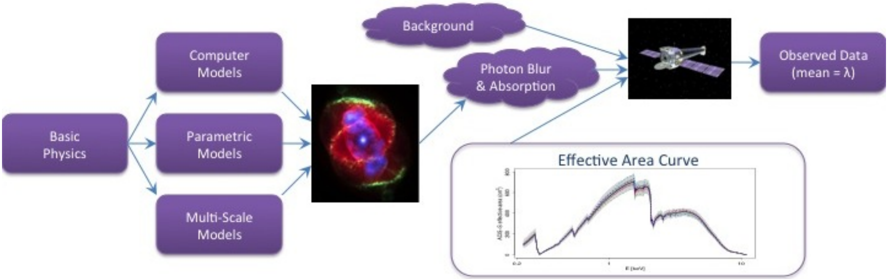
Bob Cousins at PhyStat-ν (2019)

"When you discover a new dimension/particle, can you convince the world you understand the systematics well enough to back up your claim?"

Three Topics

1. A Framework for Multi-Stage Statistical Analyses
2. Two Examples from Astrophysics
3. Why Do Many Physicists Avoid Bayesian Methods?

A Running Example – Calibration of X-ray Detectors



- Embed physics models into multi-level statistical models.
- X-ray and γ-ray detectors count a typically *small number of photons* in each of a *large number of pixels*.
- Must account for complexities of data generation.
- Effective area: instrument sensitivity as function of energy.

Accounting for Uncertainty in Effective Area

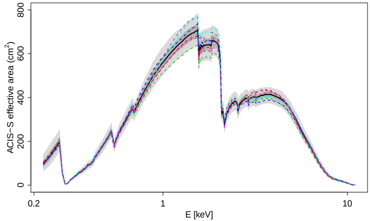
- Calibration scientists provide a sample representing uncertainty
- Introduce a Bayesian approach to ***reduce*** *prior assumptions.*
- Bayesian procedure: average

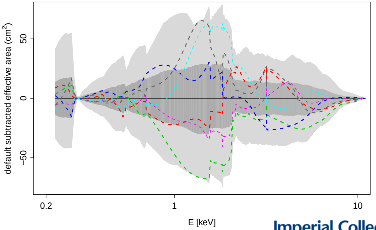
  standard model, $p(\theta|A, Y)$, over uncertainty in A, $p(A)$:

$$p(\theta|Y) = \int p(\theta|A, Y)p(A)dA.$$

Notation:

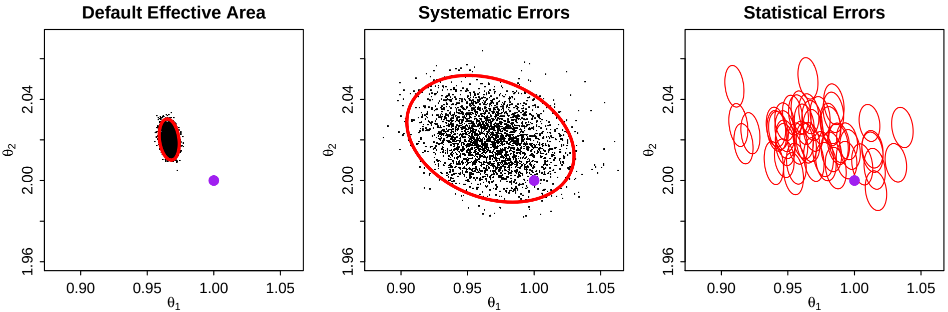
- Y = spectral data
- A = effective area – “nuisance parameter”
- θ = spectral parameters





Lee, DvD et al (2011) Astrophysical J **731**, 126.
Xu, DvD et al (2014) Astrophysical J **794**, 97.
Chen, DvD et al (2019) JASA, **114**, 1018.

Systematic and Statistical Errors – Toy Example



Spectral Model (purple bullet = truth): $f(E_j) = \theta_1 E_j^{-\theta_2}$

Default: Use best fit effective area.
Systematic: Best fit given each of a sample of effective areas from $p(A)$.
Statistical: Statistical errors for a sample of effective areas from $p(A)$.

The systematic error in the effective area biases spectral analysis.

Methodology - Two Methods Used in Physics

.... I'm sure there are more!

Multiply the Likelihoods

$$L(\theta, A \mid Y, Y_0) \equiv L(\theta \mid A, Y) L(A \mid Y_0)$$

- Perhaps use profile likelihood: $L_p(\theta) = \max_A L(\theta, A \mid Y, Y_0)$.
- Note: Estimate of A depends **on both** Y and Y_0.

Bayesian Justification:

$$\begin{aligned} p(A, \theta \mid Y_0, Y) &\propto p(Y \mid A, \theta)\, p(Y_0 \mid A, \theta)\, p(A, \theta) \\ &\stackrel{?}{=} p(Y \mid A, \theta)\, p(Y_0 \mid A)\, p(A)\, p(\theta) \\ &\propto p(Y \mid A, \theta)\, p(A \mid Y_0)\, p(\theta) \end{aligned}$$

Information Accumulates: Posterior of A from preliminary analysis is prior for A for primary analysis.

Methodology - Second Method from Physics

OPAT Forward Propagation

- In preliminary analysis, compute:
  - $\hat{A}_L = \hat{A} - \sigma_A$ and $\hat{\theta}_L = g(\hat{A}_L, Y)$
  - $\hat{A}_U = \hat{A} + \sigma_A$ and $\hat{\theta}_U = g(\hat{A}_U, Y)$

  Use $\hat{\theta}_U$ - $\hat{\theta}_L$ to compute systematic error.
- Statistical error based on $L(\theta \mid \hat{A}, y)$
- Note: Estimate of A depends **only on** Y_0.

Questions:

- What if σ_A is asymmetric or maps non-monotonically to θ?
- What if A is high-dimensional with correlated components?

Possible Pragmatic Bayesian Solution:

- Sample $A \sim p(A \mid Y_0)$ and then $\theta \sim p(\theta \mid A, Y)$.

General Strategies for Two-Stage Analyses[1]

A PRAGMATIC BAYESIAN TARGET: $\pi_0(A, \theta) = p(A)p(\theta|A, Y)$.

THE FULLY BAYESIAN POSTERIOR: $\pi(A, \theta) = p(A|Y)p(\theta|A, Y)$.

[Suppressing conditioning on Y_0].

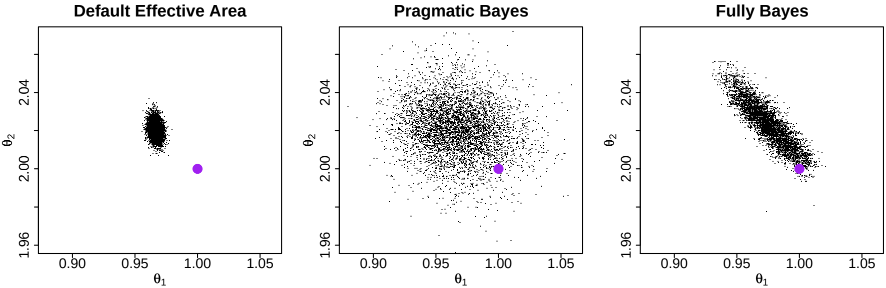
Concerns:

Statistical	Fully Bayes uses all data to reduce variance.
Cultural	Astronomers have concerns about letting the current data influence calibration products.
Future Bias	Misspecification of $p(Y \mid A, \theta)$ or $p(\theta)$, may bias estimate of A and future analyses.
Current Bias	Pragmatic Bayes – simpler model may reduce misspecification bias in current analysis. *[Event Selection]*
Computational	Pragmatic Bayesian target generally easier to sample.
Practical	How different are $p(A)$ and $p(A|Y)$?

Monte Carlo: resample nuisance parameters at each iteration.

[1] Xu, J., van Dyk, D., Kashyap, V., Siemiginowska, A., et al. (2014). A Fully Bayesian Method for Jointly Fitting Instrumental Calibration and X-ray Spectral Models. *The Astrophysical Journal*, **794**, 97.

Effective Area Results - Toy Example



Spectral Model (purple bullet = truth): $f(E_j) = \theta_1 E_j^{-\theta_2}$.

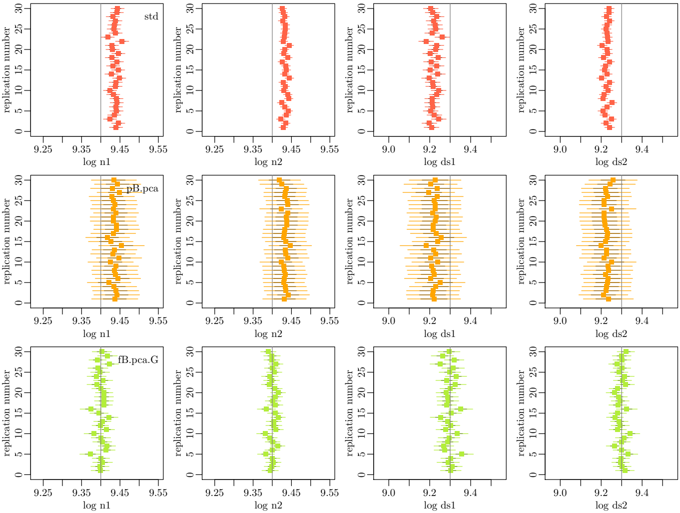
Questions for Physicists:

- Should primary analysis update nuisance parameters?
    - Forward propagation approximates Pragmatic Bayes.
    - Multiplying Likelihoods approximates Fully Bayes.

Frequentist Bias and Variance

Bias and variance of default, pragmatic, fully Bayes methods.

Replicate: Resampling data from primary experiment.

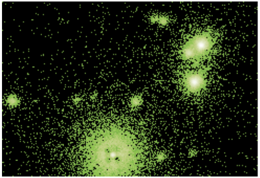


Example 1: Event/Object Selection

Parameter estimation or detection can proceed under either Fully or Pragmatic Bayes.

Event Selection

- Event selection in preliminary analysis.
- Analyse selected events in primary analysis.



Three Approaches:

1. Default Analysis: Takes classification and fixed and known.
2. Pragmatic: Account for uncertainty in classification.
3. Fully-Bayes: Use additional data in stage two to update classification probabilities .
   [Example 2: Requires models for all sources.... more models = more bias!].

Easier with probabilistic event-selection model.

Disentangling Overlapping Sources

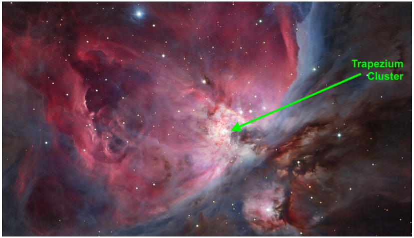
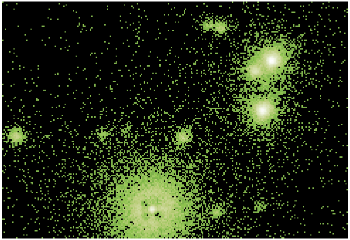


Image Credit & Copyright: László Francsics



Chandra Image of (part of) Trapezium Cluster.

We would like to separate and analyse the sources:

Stage 1: Clustering – compute $\Pr(Z_i = j \mid X, Y)$.

Stage 2: Fit source-specific spectral models.

Account for clustering uncertainty in Stage 2.

Might photon energies and arrival times improve classification?

Stage 1: A Finite Mixture Model [2] [3] [4]

Sky-coordinate and spectral model for source j:

$$(X_i, Y_i) \mid (\mu, Z_i = j) \sim \text{PSF centered at } \mu_j$$

$$E_i \mid (\alpha_j, \gamma_j, Z_i = j) \sim gamma(\alpha_j, \alpha_j/\gamma_j)$$

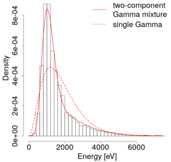


Figure: Fitting single Gamma and mixture of Gammas model to HBC 515 data. Source: Meyer et al. [2021]

Full mixture model:

- Spectral model simple and data-driven – no science parameters.
  - Can use for general catalogue – no model assumptions.
- Add flexibility – mixtures of two gamma dist'ns for spectra.
- Mix over k sources, where $k \sim \mathrm{Poisson}(\lambda)$.
- 2021 paper adds photon arrival time

[2] Jones, Kashyap, van Dyk, (2015). Disentangling Sources using Spatial-Spectral Data. *ApJ*, **808**, 137

[3] Meyer et al. (2021). Disentangling Sources Part II: Spatial-Spectral-Temporal Data. *MNRAS*, **506**, 6160

[4] Sottosanti et al. (2023+). Identification of High-Energy Astrophysical Point Sources. *arXiv:2104.11492*

Stage 1: Gamma Spectral Model

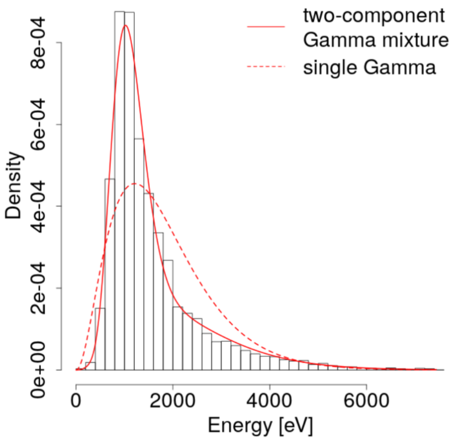
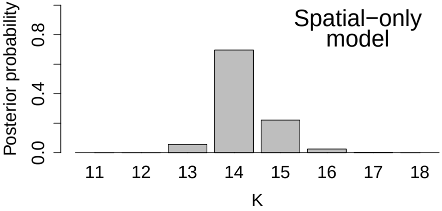
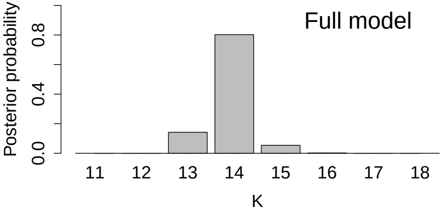


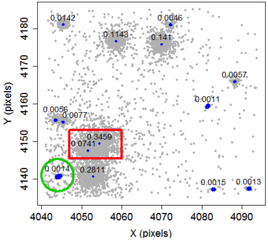
Figure: Fitting single Gamma and mixture of Gammas model to HBC 515 data. Source: Meyer et al. [2021]

Stage 1: Results for Trapezium

Posterior Distribution of k =number of sources.







***Sources in red box analyzed further.*

Spectral data yield more precise estimate.

A Two-Stage Analysis

We aim to fit Science-based spectral models:

Stage 1: Clustering – *Mixture Model gives* $\Pr(Z_i = j)$.

Stage 2: Fit source-specific spectral models. Parameters = θ_j.

Default Analysis: Use photons within fixed radii of src location

Fully Bayes: Sample from

$$p(\theta, \phi, Z \mid E, X, Y) = p(\theta \mid \phi, Z, E, X, Y) \; p(Z, \phi \mid E, X, Y)$$

$$= \left[\prod_{j=1}^{k} p(\theta_j \mid Z = j, E) \right] \; p(Z, \phi \mid E, X, Y)$$

Source Indicator (Z)

Other Parameters (ϕ)

Spectral Parameters (θ)

Stage 2: Source-by-Source Spectral Analyses (first factor)

Stage 1: Clustering via Mixture Model (second factor)

But the spectral models are not congenial....

A Two-Stage Analysis - Con't

We aim to fit Science-based spectral models:

Fully Bayes:

$$p(\theta, \phi, Z \mid E, X, Y) = \left[\prod_{j=1}^{k} p(\theta_j \mid Z = j, E)\right] p(Z, \phi \mid E, X, Y)$$

Pragmatic Bayes:

Stage 2 / Stage 1

$$p(\theta, \phi, Z \mid E, X, Y) = \boxed{\left[\prod_{j=1}^{k} p(\theta_j \mid Z = j, E)\right]} \; \boxed{p(Z, \phi \mid X, Y)}$$

Note Stage 1: Energies (E) not used in clustering – no spectral model.

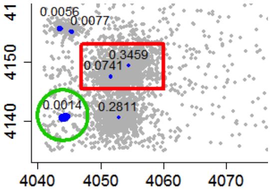
Including energy in Stage-1 classification approximates fully Bayes, but with non-congenial spectral models.

... tradeoff between using full information and a properly specified model

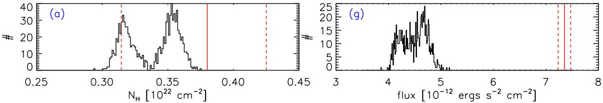
Results of Two Stage Analysis

Conduct Stage-2 analysis for overlapping sources in red box.

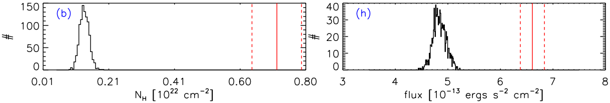
- MCMC - reassign photons at each iteration.
- Science-based spectral model
  (absorbed single temp thermal model)
- Top source is ∼five times brighter.
- Vertical lines: default fits ($\pm\sigma$, statistical)
- Histogram: uncertainty due to photon allocation



Bright source:



Faint source:



Classification uncertainty in non-negligible.

Imperial College London

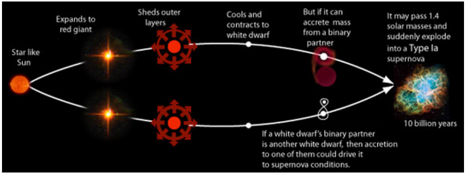
Example 2: Studying Expansion History of Universe

Could there be an advantage of the Pragmatic approach?

Type Ia Supernovae had a common "flashpoint"

Absolute magnitudes:

$$M_j^{\mathrm{Ia}} \sim \mathrm{N}(M_0^{\mathrm{Ia}}, \sigma_{\mathrm{int}}^{\mathrm{Ia}}).$$



Non-linear Regression: $m_{Bj} = g(z_j, \Omega_\Lambda, \Omega_M, H_0) + M_j^{\mathrm{Ia}}$

[e.g., Λ-CDM: function of density of dark energy and of total matter]

[part of a (second-stage) fully-Bayesian Hierarchical model ⋆]

First Stage Analysis: Classify Supernova into Type Ia, non Type Ia.

[New general method for handling a non-representative training set⋆⋆]

For Non Type Ia: $M_j^{\mathrm{Ia}'} \sim \mathrm{Distribution}(M_0^{\mathrm{Ia}'}, \sigma_{\mathrm{int}}^{\mathrm{Ia}'})$ with $\sigma_{\mathrm{int}}^{\mathrm{Ia}'} \gg \sigma_{\mathrm{int}}^{\mathrm{Ia}}$

⋆ Shariff, Jiao, Trotta, and van Dyk (2016). BAHAMAS: New SNIa Analysis Reveals Inconsistencies with Standard Cosmology. *The Astrophysical Journal*, **827**, 1

⋆⋆ Autenrieth, van Dyk, Trotta, Stenning (2023+). Stratified Learning. . . , *arXiv:2106.11211*

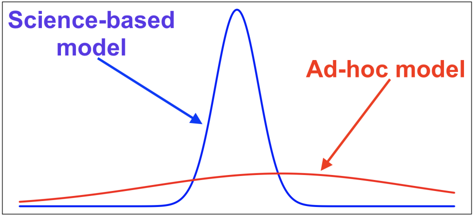
Bias-Variance Trade Off

In Fully Bayesian analysis, given θ, the relative densities:

[θ = cosmological parameters; Y = apparent magnitudes]

Type Ia: $p(Y \mid \theta, \text{Type Ia})$ and

Non-Type Ia: $p(Y \mid \theta, \text{Not Ia})$ ...*will inform* $p(\textit{Type Ia} \mid Y, \theta)$.



Insofar as model for Non-Type Ia selected for convenience and may suffer misspecification, pragmatic Bayes may reduce bias.

Work in Progress... but my bias is toward a pragmatic approach!!

Summary

Default / Naïve Methods

- Underestimate uncertainty and can introduce bias.
- Avoid unless nuisance parameters are very well estimated.

Pragmatic Bayesian Method

- Simple way to avoid problems of default / naïve approach.
- Can overstate uncertainty and exhibit bias.

... but it is better to overstate than to understate uncertainty

Pragmatic Bayesian Method

- Best use of data – If model is perfectly specified
- Requires coordination of preliminary and primary analyses
- May require additional model assumptions

Bayesian Methods

Bayes Theorem

$$\Pr(\theta \mid Y) = \frac{\Pr(Y \mid \theta)\Pr(\theta)}{\int \Pr(Y \mid \theta)\Pr(\theta)\mathrm{d}\theta}$$

Bayesian methods

- have cleaner mathematical foundations
- signpost principled methodology *[e.g., multiplying likelihoods]*
- can help identify assumptions *[e.g., of OPAT]*
- more directly answer scientific questions

But they depend on ***prior distributions***

- $\Pr(\theta)$ quantifies likely values of θ before having seen data.

Frequentist Properties Are Also Compelling

Frequentist justification of likelihood based methods:

under certain conditions...

1. $\hat{\theta}_{\text{MLE}}$ is an *asymptotically* unbiased estimator of θ
2. The sampling variance of $\hat{\theta}_{\text{MLE}}$ goes to zero *as n* $\rightarrow \infty$.
3. (standardized) $\hat{\theta}_{\text{MLE}}$ *converges* in distribution to normal.

Bayesian estimates enjoy the same asymptotic properties!

if prior assigns positive probability to a neighborhood of θ

- Large sample asymptotics are primary justification for likelihood-based methods.
- Bayesian methods enjoy an alternative (small sample) justification.

Profile or Marginalize?

Profile Likelihood

$$L_p(\theta) = \max_A L(\theta, A \mid Y, Y_0)$$

Marginal Likelihood

$$L(\theta \mid Y, Y_0) = \int p(Y \mid \theta, A)\, p(A \mid Y_0)\mathrm{d}A$$

What is the justification for the profile likelihood?

In the large sample asymptotic case.... *again under certain conditions...*

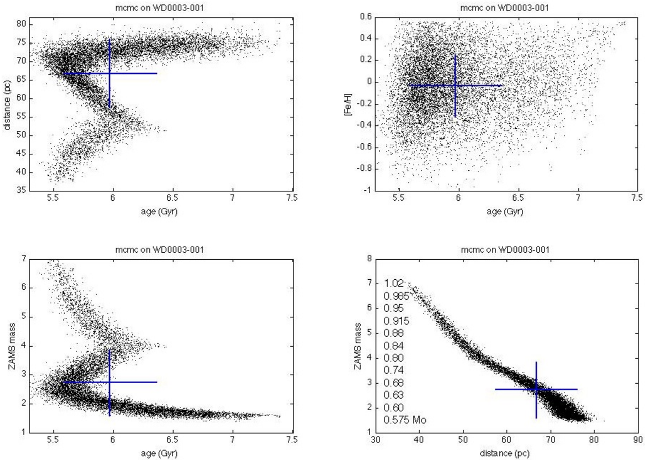
... the log-likelihood is quadratic in the parameter (i.e., Gaussian) and

... the profile and marginal likelihoods are equivalent.

But this is the easy case!

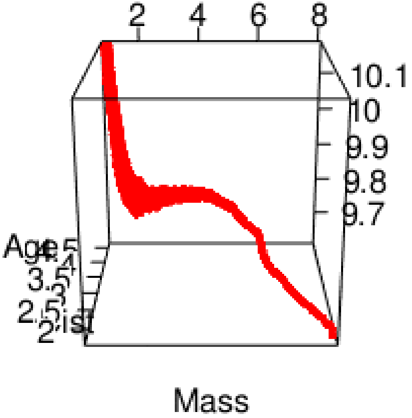
Want to Bet on Asymptotic Gaussians?

A few examples from my work:



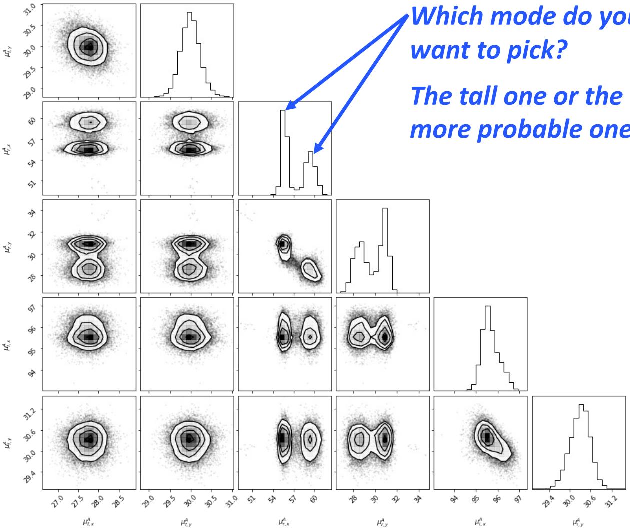
Highly non-linear relationship among stellar parameters.

Want to Bet on Asymptotic Gaussians?



Highly non-linear relationships among stellar parameters.

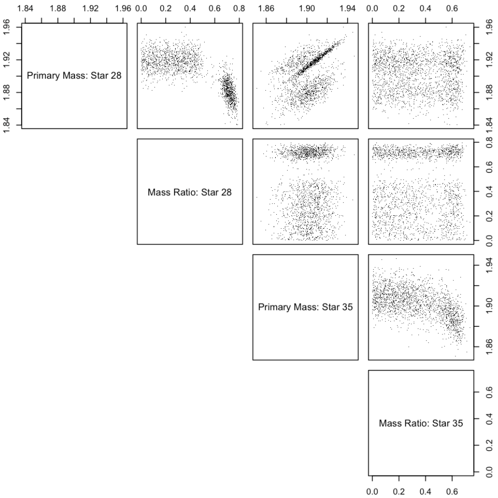
Want to Bet on Asymptotic Gaussians?



Which mode do you want to pick?

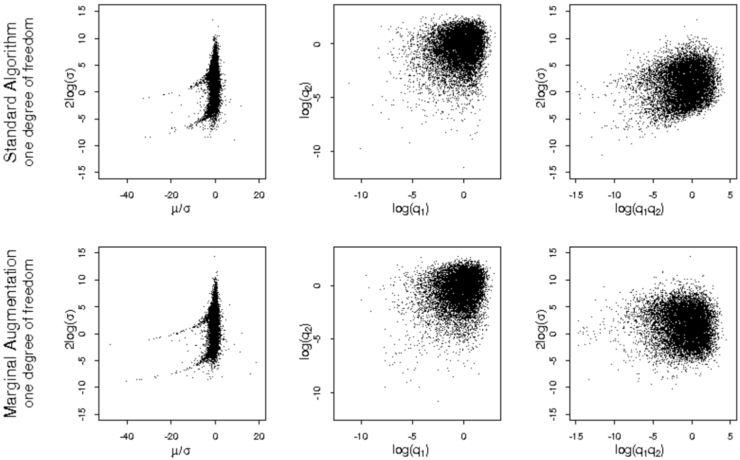
The tall one or the more probable one?

Want to Bet on Asymptotic Gaussians?

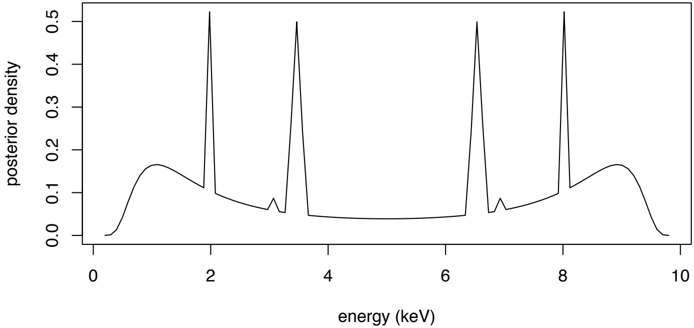


The classification of certain stars as field or cluster stars can cause multiple modes in the distributions of other parameters.

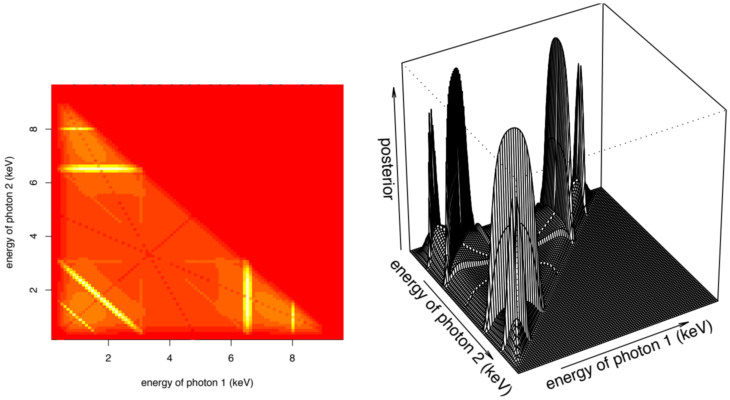
Want to Bet on Asymptotic Gaussians?



Want to Bet on Asymptotic Gaussians?



Want to Bet on Asymptotic Gaussians?



When to worry

Confession: I use profile likelihood, but I worry when I do.

If your analyses are based on asymptotic properties,

- your data being Gaussian is not enough.

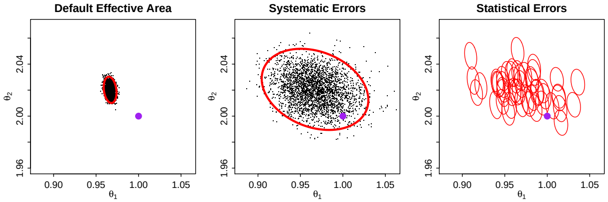
Watch for warning signs....

- strange (non-convex?) contours
- MLE/MAP on boundary of parameter space
- confidence intervals are assymmetric or contain non-physical values

If asymptotics don't apply investigate frequency properties via Monte Carlo!

... or base inference on small sample justification of Bayesian analyses.

Quantifying Total Uncertainty



Physicists often decompose the errors:

$$\text{estimate} \pm \text{statistical error} \pm \text{systematic error}$$

- How is the systematic error computed?
  Likelihood doesn't distinguish; dealing with correlations is complicated.

Bayesians might use Law of Total Variance:

$$\mathrm{VAR}(\theta) = \mathrm{VAR}\big[\mathrm{E}(\theta \mid A)\big] + \mathrm{E}\big[\mathrm{VAR}(\theta \mid A)\big]$$
$$= \text{systematic var} + \text{expected statistical var}$$

...where all moments are conditional on Y_0 and Y.

Imperial College London

Collaborators

- Maximilian Autenrieth (Imperial College)
- Alessandra Brazzale (Padova)
- Alanna Connors (deceased)
- Alex Geringer-Sameth (Lawrence Livermore)
- Xiyun Jiao (SUSTech)
- David Jones (Texas A& M)
- Vinay Kashyap (Harvard Smithsonian CfA)
- Hyunsook Lee (Lam Research)
- Antoine Meyer (Imperial College)
- Xiao-Li Meng (Harvard)
- Esben Revsbech (Jyske Bank)
- Hikmatali Shariff (Imperial College)
- Andrea Sottosanti (Padova)
- Aneta Siemiginowska (Harvard Smithsonian CfA)
- David Stenning (Simon Fraser)
- Roberto Trotta (Imperial College & SISSA - Trieste)
- Jin Xu (Adobe)
- Andreas Zezas (Crete)

Sponsored by:

Calibration and Multi-Stage Analyses

- Lee, Kashyap, van Dyk, Connors, Drake, Izem, Min, Park, Ratzlaff, et al.
  Accounting for Calibration Uncertainties in X-ray [Spectral] Analysis
  The Astrophysical Journal, **731**, 126–144, 2011.
- Xu, van Dyk, Kashyap, Siemiginowska, Connors, Drake, Meng, Ratzlaff, and Yu.
  A Fully Bayesian Method for Jointly Fitting Calibration and X-ray Spectral Models
  The Astrophysical Journal, **794**, 97 (21pp), 2015.
- Yu, Del Zanna, Stenning, Cisewski-Kehe, Kashyap, Stein, van Dyk, Warren, et al.
  Incorporating Uncertainties in Atomic Data Into [Solar Analyses]
  The Astrophysical Journal, **866**, 146 (20 pages), 2018.

Separating Source From Background

- Jones, Kashyap, and van Dyk.
  Disentangling Overlapping Sources using Spatial and Spectral Information.
  The Astrophysical Journal, **808**, 137 (24 pp), 2015.
- Meyer, van Dyk, Kashyap, Campos, Jones, Siemiginowska, and Zezas.
  eBASCS: Disentangling Overlapping Sources II, using Temporal Information.
  Monthly Notices of the Royal Astronomical Society, **506**, 5160 (21pp), 2021.
- Sottosanti, Bernardi, Brazzale, Geringer-Sameth, Stenning, Trotta, and van Dyk.
  Identification of High-Energy Astrophysical Point Sources.
  arXiv:2104.11492, 2021.

Type Ia Supernova Cosmology

Autenrieth, M., van Dyk, D. A., Trotta, R., and Stenning, D. C.
Stratified Learning: A General-Purpose Statistical Method for Improved Learning under Covariate Shift
arXiv:2106.11211, 2021.

Revsbech, E., Trotta, R., and van Dyk, D. A.
STACCATO: A Novel Solution to Supernova Photometric Classification...
Monthly Notices of the Royal Astronomical Society, **473**, 3969–3986, 2018.

Shariff, H., Jiao, X., Trotta, R., and van Dyk, D. A.
BAHAMAS: SNIa Analysis Reveals Inconsistencies with Standard Cosmology.
The Astrophysical Journal, **827**, 1 (25 pp), 2016.