# Statistical and Systematic Uncertainties
## A Statistician's Perspective

David A. van Dyk

Statistics Section of Department of Mathematics
Imperial College London

PhyStat – Sytematics, November 2021

Imperial College London

# Disclaimer

**My perspective is informed by:**

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

Statisticians do not always agree on everything.

*Some bits are rather philosophical.*

*I don't speak for ALL statisticians!*

*Some overlap with Louis's talk, but perhaps from a different perspective....*

# 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.
- Output for one analysis is input for subsequent analyses.
- Can we combine into principled omnibus analysis?

**Systematics in Physics**

- Primary analysis is informed by Preliminary analyses.
- Louis's pendulum example: $g = 4\pi^2 L/\tau^2$
  - What if we use length or period $(L, \tau)$ estimated with error?
  - What if ideal conditions of theoretical pendulum don't hold?
- In either case the model is misspecified.

*Statistical theory generally assumes specified model is same as data-generation model.*

*How do we properly quantify uncertainty?*

# Outline

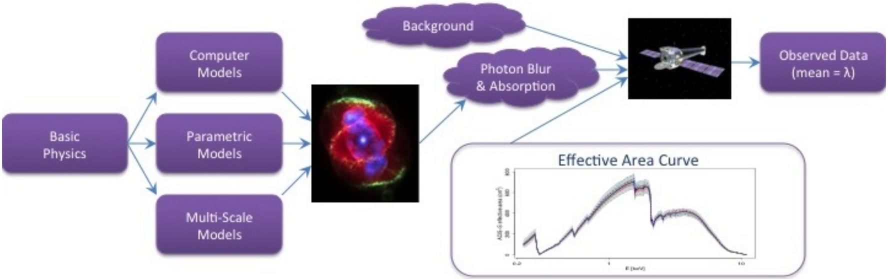
**Bob Cousins at PhyStat-$\nu$ (2019)**

*"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. Definitions and Preliminaries
2. A Framework for Multi-Stage Statistical Analyses
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 $\gamma$-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

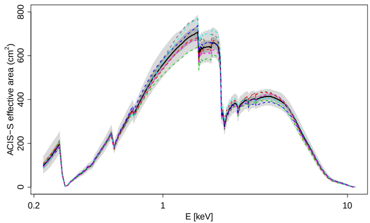
- Calibration scientists provide a sample representing uncertainty
- Introduce a Bayesian approach to ***reduce*** *prior assumptions.*
- Bayesian procedure: average

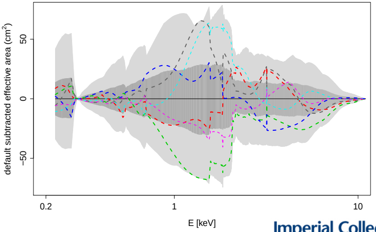
  standard model, $p(\theta|A, Y)$, over uncertainty in $A$, $p(A)$:

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

**Notation:**

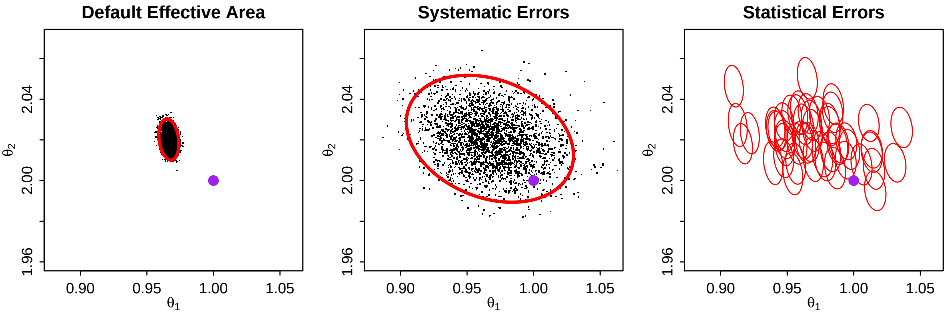
- $Y$ = spectral data
- $A$ = effective area – "nuisance parameter"
- $\theta$ = spectral parameters




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 $p(A)$.

*Systematic error in the effective area bias spectral analysis.*

Imperial College London

# Frequentist Analysis

**General Setup** (using notation of effective area example)

Preliminary Analysis: Use $Y_0$ to estimate $A$ ("nuisance parameter").

Primary Analysis: Use $Y$ to estimate $\theta$ (depends on $A$).

**First: Suppose $A$ is known in Primary Analysis**

- Let $\hat{\theta} = g(A, Y)$ be unbiased estimator of $\theta$:

$$\mathrm{E}(\hat{\theta}) = \mathrm{E}[g(A, Y)] = \theta,$$

  with expectation over sampling distribution $p(Y \mid A, \theta)$.

- Statistical Error = $\hat{\theta} - \theta$.

  *[If* $\mathrm{VAR}(\hat{\theta}) \to 0$ *as* $n \to \infty$*, then* $\mathrm{MSE} = \mathrm{E}[(\hat{\theta} - \theta)^2]$ *also goes to zero.]*

*But what if we use an estimate, $\hat{A} \neq A$?*

# A Misspecified Primary Model

**Replacing $A$ with an estimate from Preliminary analysis**

- $\hat{\theta}$ is typically biased

$$\mathrm{E}[g(\hat{A}, Y)] \neq \theta$$

with $\hat{A} \neq A$ and expectation over $p(Y \mid A, \theta)$.

Statistical Error $= \hat{\theta} - \mathrm{E}[g(\hat{A}, Y)]$ *(Goes to zero as $n \to \infty$.)*

Systematic Error $= \mathrm{E}[g(\hat{A}, Y)] - \theta$ *(Does not depend on $Y$ or $n$.)*

*[a.k.a., the bias due to model misspecification]*

## A Possible Definition

Statistical Error: Errors that dissipate as $n \to \infty$.

Systematic Error: Errors that do not dissipate as $n \to \infty$.

With $n =$ sample size of the primary analysis.

# Reference Frame Matters

**A Possible Definition**

Statistical Error: Errors that dissipate as $n \to \infty$.

Systematic Error: Errors that do not dissipate as $n \to \infty$.

With $n =$ sample size of the primary analysis.

**Reference Frame Matters**

- Frequentist evaluation of combined analysis:
  - Sample of size $n$ from $p(Y_0, Y \mid A, \theta)$.
  - All errors are statistical and typically dissipate as $n \to \infty$
- Actually we have fixed samples of each!
  - Why imagine $n_0$ is fixed and $n \to \infty$?

*Consider a Bayesian perspective?*

# 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 $\sigma_A$ is asymmetric or maps nonlinearly onto $\theta$?
- 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 $Y_0$].*

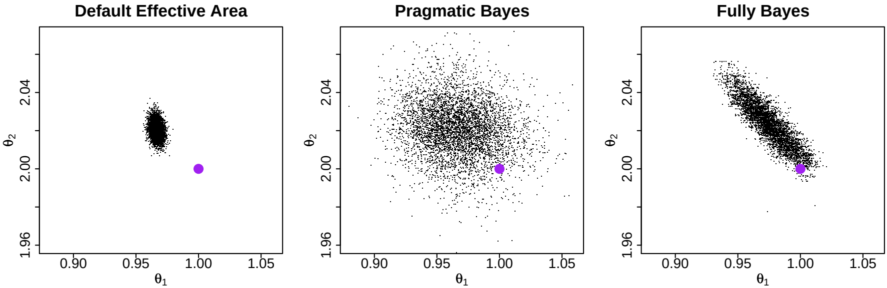
**Concerns:**

| | |
|---|---|
| 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 $p(A|Y)$? |

*With MCMC: sample 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.

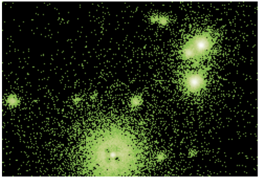
Imperial College London

# Event 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:

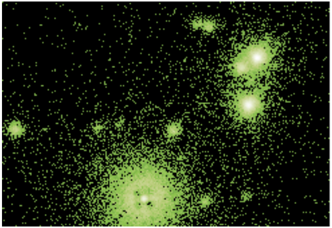
1. Naive Analysis: Takes classification and fixed and known.
2. Pragmatic: Account for uncertainty in classification.
3. Fully-Bayes: Use additional data in stage two to update classification probabilities .

*[Requires models for all sources.... more models = more bias!].*

*Easier with probabilistic event-selection model.*

# Event Selection - Citations

*An unrepresentative sample of fully-Bayesian methods for high-energy astrophysics:*



- Jones etal (2015): Improve event selection witf flexible spectral models.
- Meyer et al. (2021): Further improve by also using light curves.
- Sottosanti et al. (2021): Improve with flexible (B spline) bkgd models.

*[Full citations given at the end.]*

Jones et al (2015) illustrates two-stage analysis.

# 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., forward propagation]*
- more directly answer scientific questions

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

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

# Frequentist Properties Are Also Compelling

Frequentist justification of likelihood based methods:

*under certain conditions...*

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

**Bayesian estimates enjoy the same asymptotic properties!**

*if prior assigns positive probability to a neighborhood of* $\theta$

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

# Examples

## Profile Likelihood

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

*Is profiling ever better than marginalizing?* [2]

- LRT is fine, as long as you are in the asymptotic regime.
- Confession: I use profile likelihoods, but I worry when I do.
- The marginal posterior is always justified.

## High-Dimensional Nuisance Parameters

- Neyman-Scott Paradox: nuisance dimension grows with $n$.
- Violates likelihood asymptotics.
- Bayesian: put prior on nuisance parameters and marginalize.



[2] "Information for Statisticians" provided for conference attendees.

# When to worry

If your analyses are based on asymptotic frequency properties,

- your data being Gaussian is not enough.

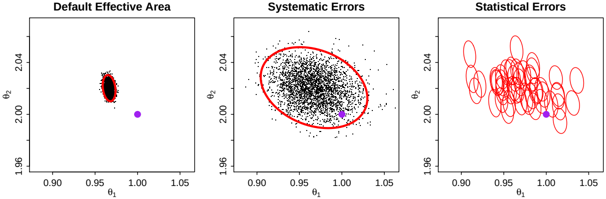
You need to watch for warning signs....

- strange (non-convex?) contours
- MLE/MAP on boundary of parameter space
- confidence intervals are assymetric 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:**

estimate $\pm$ statistical error $\pm$ systematic error

- How is the systematic error computed?

  *Likelihood doesn't distinguish; dealing with correlations is complicated.*

**Bayesians might use Law of Total Variance:**

$$\text{VAR}(\theta) = \text{VAR}\big[\text{E}(\theta \mid A)\big] + \text{E}\big[\text{VAR}(\theta \mid A)\big]$$

$$= \text{systematic var} + \text{expected statistical var}$$

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

## 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.

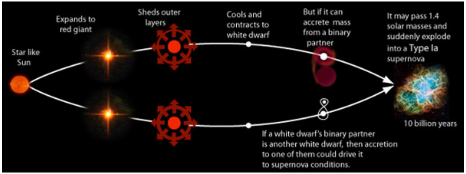
# Studying the Expansion History of Universe

*Could there be an advantage of the Pragmatic approach?*

Type Ia Supernovae had a common "flashpoint"

**Absolute magnitudes:**

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




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

*[function of density of dark energy and of total matter]*

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

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

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

*[New general method for non-representative training set★★, Cosmology Session, Max Autenrieth Wed @ 3:30pm]*

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

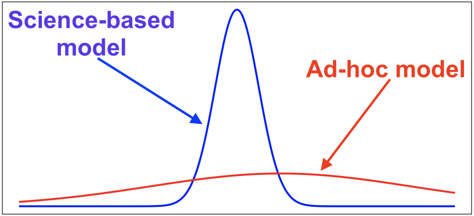
★★ Autenrieth, van Dyk, Trotta, Stenning (2021). Stratified Learning..., *arXiv:2106.11211*

# Bias-Variance Trade Off

In Fully Bayesian analysis, given $\theta$, the relative densities:

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

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

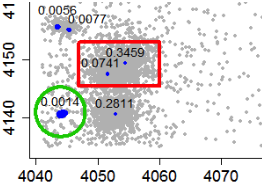



*Insofar as model for Non-Type Ia is selected for convenience and suffers misspecification, pragmatic Bayes may reduce bias.*

# Background – Pragmatic Bayesian Results

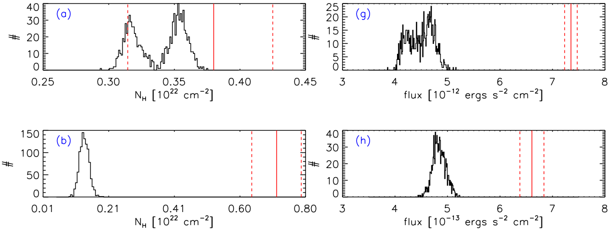
## Conduct Stage-2 analysis for overlapping sources in red box.

- MCMC - reassign photon to sources/background at each iteration.
- Partially Collapsed Gibbs Sampler.*
- Science-based spectral model (4 param)
  - Vertical lines are naive fits ($\pm\sigma$)
- Top source is $\sim$ five times brighter.




Bright source:

Faint source:




*Uncertainty due to overlap carried forward to spectral fit!*

Imperial College London

* van Dyk and Jiao (2015). MH within Partially Collapsed Gibbs Samplers. *JCGS*, **24**, 301–327.