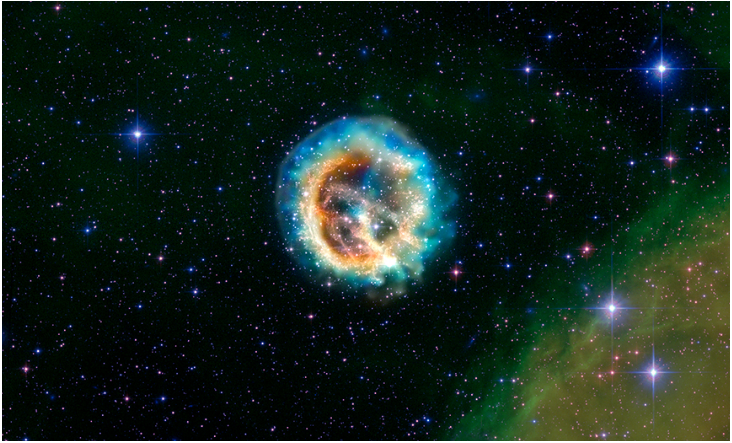
Calibration Concordance

Xiao-Li Meng, Harvard University

Joint work with Y. Chen, X. Wang, D. van Dyk, V. Kashyap, H. Marshall

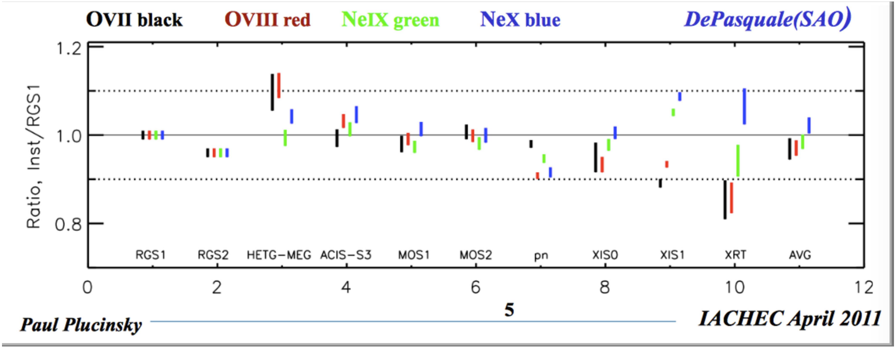
August 21, 2017

Calibration Concordance Problem (Example: E0102)



- Supernova remnant E0102
- Four sources correspond to four spectral lines in E0102

Calibration Concordance Problem (Example: E0102)



- Four spectral lines observed with 11 X-ray detectors
- Main challenge – the data/instruments do not agree

Outline

Notation

- N Instruments with true effective area A_i, $1 \le i \le N$.
  - For each instrument i, we know estimated a_i ($\approx A_i$) but not A_i.

Notation

- N Instruments with true effective area A_i, $1 \leq i \leq N$.
  - For each instrument i, we know estimated a_i ($\approx A_i$) but not A_i.
- M Sources with fluxes F_j, $1 \leq j \leq M$.
  - For each source j, F_j is unknown.

Notation

- N Instruments with true effective area A_i, $1 \leq i \leq N$.
  - For each instrument i, we know estimated a_i ($\approx A_i$) but not A_i.
- M Sources with fluxes F_j, $1 \leq j \leq M$.
  - For each source j, F_j is unknown.
- Photon counts c_{ij}: from measuring flux F_j with instrument i.

Notation

- N Instruments with true effective area A_i, $1 \le i \le N$.
  - For each instrument i, we know estimated a_i ($\approx A_i$) but not A_i.
- M Sources with fluxes F_j, $1 \le j \le M$.
  - For each source j, F_j is unknown.
- Photon counts c_{ij}: from measuring flux F_j with instrument i.
- Lower cases: data / estimators.
- Upper cases: parameter / estimand.

Calibration Concordance Problem

1. Astronomers' Dilemma:

$$\frac{c_{ij}}{a_i} \neq \frac{c_{i'j}}{a_{i'}} \text{ for } i \neq i'.$$

Different instruments give different estimated flux of the same object!

Calibration Concordance Problem

1. Astronomers' Dilemma:

$$\frac{c_{ij}}{a_i} \neq \frac{c_{i'j}}{a_{i'}} \text{ for } i \neq i'.$$

Different instruments give different estimated flux of the same object!

2. Scientific Question:

- Are there systematic errors in 'known' effective areas?
- Can we derive properly adjusted effective areas?
- Can we unify estimates of the same flux with different instruments?

Scientific and Statistical Models

Scientific Model

Multiplicative in original scale and additive on the log scale.

$$\text{Counts} = \text{Exposure} \times \text{Effective Area} \times \text{Flux},$$
$$C_{ij} = T_{ij} A_i F_j, \quad \Leftrightarrow \quad \log C_{ij} = B_i + G_j,$$

where log area $= B_i = \log A_i$, log flux $= G_j = \log F_j$; let $T_{ij} = 1$.

Scientific and Statistical Models

Scientific Model

Multiplicative in original scale and additive on the log scale.

$$\text{Counts} = \text{Exposure} \times \text{Effective Area} \times \text{Flux},$$
$$C_{ij} = T_{ij} A_i F_j, \quad \Leftrightarrow \quad \log C_{ij} = B_i + G_j,$$

where log area $= B_i = \log A_i$, log flux $= G_j = \log F_j$; let $T_{ij} = 1$.

Statistical Model

$$\text{log counts } y_{ij} = \log c_{ij} - \alpha_{ij} = B_i + G_j + e_{ij}, \quad e_{ij} \overset{indep}{\sim} \mathcal{N}(0, \sigma_{ij}^2);$$

where $\alpha_{ij} = -0.5\sigma_{ij}^2$ to ensure $E(c_{ij}) = C_{ij} = A_i F_j$.

- **Known Variances**: σ_{ij} known.
- **Unknown Variances**: $\sigma_{ij} = \sigma_i$ unknown.

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

$$\text{log counts } | \textit{area} \,\&\, \textit{flux} \,\&\, \textit{variance} \overset{\text{indep}}{\sim} \text{Gaussian distribution},$$

$$y_{ij} \mid B_i,\ G_j,\ \sigma_i^2 \overset{\text{indep}}{\sim} \mathcal{N}\left(B_i + G_j,\ \sigma_i^2\right),$$

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

$$\begin{aligned}
\text{log counts } | \textit{area \& flux \& variance} &\overset{\text{indep}}{\sim} \text{Gaussian distribution},\\
y_{ij} \mid B_i,\ G_j,\ \sigma_i^2 &\overset{\text{indep}}{\sim} \mathcal{N}\left(B_i + G_j,\ \sigma_i^2\right),\\
B_i &\overset{\text{indep}}{\sim} N(b_i,\ \tau_i^2),\\
G_j &\overset{\text{indep}}{\sim} \text{flat prior},
\end{aligned}$$

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

$$\begin{aligned}
\text{log counts } | \textit{area \& flux \& variance} &\overset{\text{indep}}{\sim} \text{Gaussian distribution},\\
y_{ij} \mid B_i,\ G_j,\ \sigma_i^2 &\overset{\text{indep}}{\sim} \mathcal{N}\left(B_i + G_j,\ \sigma_i^2\right),\\
B_i &\overset{\text{indep}}{\sim} N(b_i,\ \tau_i^2),\\
G_j &\overset{\text{indep}}{\sim} \text{flat prior},\\
\text{If variance unknown: } \sigma_i^2 &\overset{\text{indep}}{\sim} \text{Inv-Gamma}(df_g,\ \beta_g).
\end{aligned}$$

Setting the prior parameters.

1. $b_i = \log a_i$, τ_i are given by astronomers.

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

$$\begin{aligned}
\text{log counts } | \textit{area} \,\&\, \textit{flux} \,\&\, \textit{variance} &\overset{\text{indep}}{\sim} \text{Gaussian distribution}, \\
y_{ij} \mid B_i,\ G_j,\ \sigma_i^2 &\overset{\text{indep}}{\sim} \mathcal{N}\left(B_i + G_j,\ \sigma_i^2\right), \\
B_i &\overset{\text{indep}}{\sim} N(b_i,\ \tau_i^2), \\
G_j &\overset{\text{indep}}{\sim} \text{flat prior}, \\
\text{If variance unknown: } \sigma_i^2 &\overset{\text{indep}}{\sim} \text{Inv-Gamma}(df_g,\ \beta_g).
\end{aligned}$$

Setting the prior parameters.

1. $b_i = \log a_i$, τ_i are given by astronomers.
2. df_g, β_g are given based on the variability in data.

Advantages of Our Approach

1. Intuitive Interpretation: Shrinkage Estimators
2. Adjusted Estimates of Effective Area
3. Calibration Concordance
4. Avoiding Pitfalls of Wrong ‘Known Variances’

Shrinkage Estimators: Known Fluxes and Errors

Hierarchical model ⇒ Shrinkage estimators

- weighted averages of evidence from 'Prior' and evidence from 'Data').

Shrinkage Estimators: Known Fluxes and Errors

Hierarchical model $\Rightarrow$ Shrinkage estimators

- weighted averages of evidence from 'Prior' and evidence from 'Data').

When fluxes and variances are known,

Original Scale

$$\widehat{A}_i = a_i^{W_i} \left[(\tilde{c}_{i\cdot} \tilde{f}^{-1}) e^{\sigma_i^2/2} \right]^{1-W_i},$$

where

$$\tilde{c}_{i\cdot} = \prod_j c_{ij}^{1/M}, \ \ \tilde{f} = \prod_j f_j^{1/M}$$

are geometric means.

Log-Scale

$$\widehat{B}_i = W_i b_i + (1 - W_i)(\bar{y}_{i\cdot} - \bar{G}),$$

where

$$\bar{G} = \frac{\sum_j g_j}{M}, \bar{y}_{i\cdot} = \frac{\sum_j y_{ij}}{M}$$

are arithmatic means.

The 'weights', $W_i = \frac{\tau_i^{-2}}{\tau_i^{-2} + M\sigma_i^{-2}}$, represents the direct information in b_i relative to indirect information in fluxes.

Shrinkage Estimators: Known Errors

When fluxes are unknown and variances are known,

$$\widehat{B}_i = W_i b_i + (1 - W_i)(\bar{y}_{i\cdot} - \bar{G}_i), \quad \widehat{G}_j = \bar{y}_{\cdot j} - \bar{B},$$

where $\bar{G}_i = \frac{\sum_j \widehat{G}_j}{M}, \quad \bar{B} = \frac{\sum_i \widehat{B}_i \sigma_i^{-2}}{\sum_i \sigma_i^{-2}}, \; \bar{y}_{i\cdot} = \frac{\sum_j y_{ij}}{M}, \quad \bar{y}_{\cdot j} = \frac{\sum_i y_{ij} \sigma_i^{-2}}{\sum_i \sigma_i^{-2}}.$

In practice, we use MCMC to fit the full model.

Shrinkage Estimators: Known Errors

When fluxes are unknown and variances are known,

$$\widehat{B}_i = W_i b_i + (1 - W_i)(\bar{y}_{i\cdot} - \bar{G}_i), \quad \widehat{G}_j = \bar{y}_{\cdot j} - \bar{B},$$

where $\bar{G}_i = \frac{\sum_j \widehat{G}_j}{M}, \quad \bar{B} = \frac{\sum_i \widehat{B}_i \sigma_i^{-2}}{\sum_i \sigma_i^{-2}}, \; \bar{y}_{i\cdot} = \frac{\sum_j y_{ij}}{M}, \quad \bar{y}_{\cdot j} = \frac{\sum_i y_{ij} \sigma_i^{-2}}{\sum_i \sigma_i^{-2}}.$

In practice, we use MCMC to fit the full model.

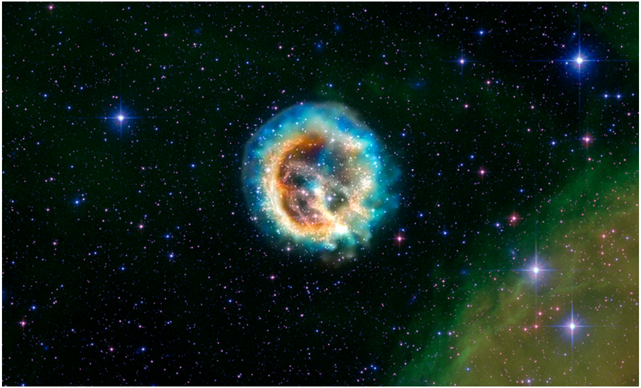
When fluxes and variances are unknown,

again, we use MCMC to fit the full model.

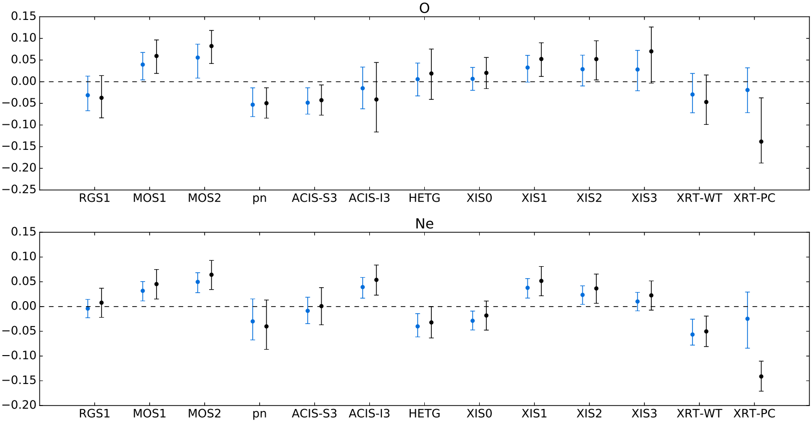
Numerical Results (E0102)

Recap: Supernova remnant E0102.

Four sources are four spectral lines in E0102.



Estimates of $B_i = \log A_i$ ($M = 2$ each panel)

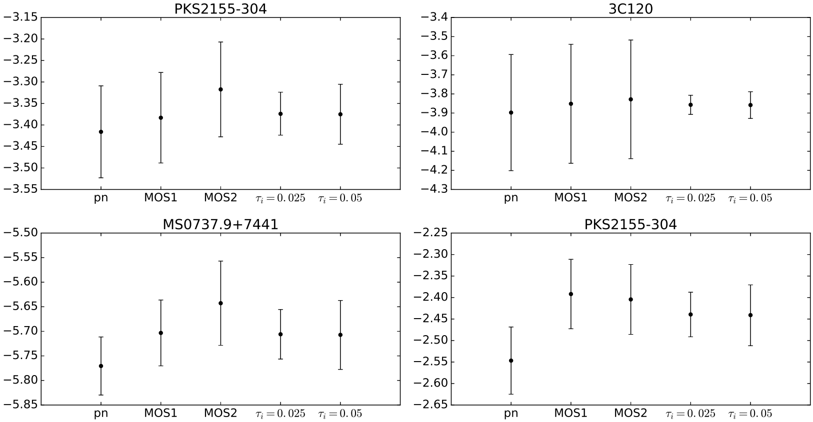


- Adjusted so that default effective area, $b_i = \log a_i = 0$.
- 95% posterior intervals (black:$\tau = 0.05$; blue: $\tau = 0.025$).
- Some instruments systematically high, others low.

Numerical Results (XCAL)

- **XCAL data**: Bright active galactic nuclei from the XMM-Newton cross-calibration sample.
- **Pileup**: Image data are clipped to eliminate the regions affected by pileup, determined using `epatplot`.
- **Three detectors**: MOS1, MOS2 and pn.
- **Sources**: M=103 (in medium band).

Numerical Results (XCAL): Calibration Concordance



- y-axis: G (log flux)
- vertical bars (left 3): mean $\pm$ 2 s.d. based on observed fluxes
  (right 2): 95% our posterior intervals.
- Calibration Concordance: *A single estimate of each flux!*

Benefits of Fitting σ_i^2

- Tolerance to model/error model misspecification.

Benefits of Fitting σ_i^2

- Tolerance to model/error model misspecification.
- Pitfalls of assuming ‘known’ variances:
  - Overly optimistic ‘known variances’

    $\Rightarrow$ overly narrow confidence intervals

    $\Rightarrow$ possible false discoveries

Benefits of Fitting σ_i^2

- Tolerance to model/error model misspecification.
- Pitfalls of assuming 'known' variances:
  - Overly optimistic 'known variances'

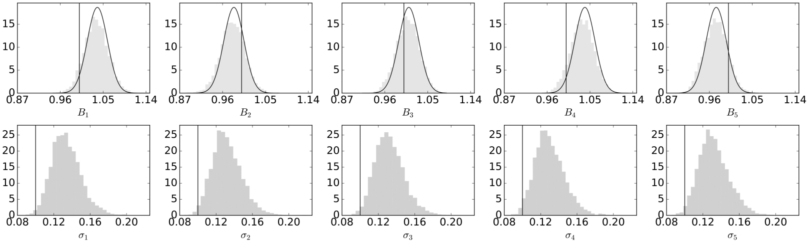
    $\Rightarrow$ overly narrow confidence intervals

    $\Rightarrow$ possible false discoveries
  - 'known variances' $\geq$ true variability

    $\Rightarrow$ noninformative results

Benefits of Fitting σ_i^2: Example

Simulated Data: Poisson data with $N = 10, M = 40,\ B_i = 1,\ G_j = 3$.



Histograms: posterior distributions.

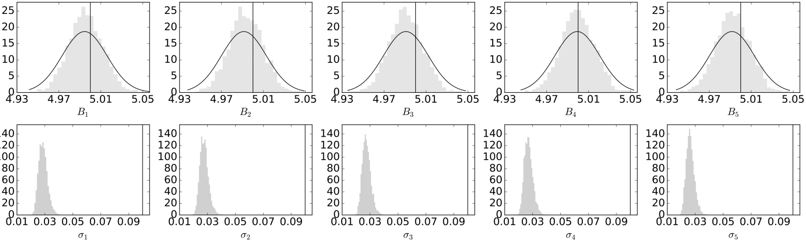
Vertical line: true values

Black Curve: Results with 'known variances' $\sigma_i^2 = 0.1^2$, ($\approx$ fit)

No cost to fitting σ, even when values are known correctly.

Benefits of Fitting σ_i^2: Example

Simulated Data: Poisson data with $N = 10, M = 40, B_i = 5, G_j = 3.$



Histograms: posterior distributions.

Vertical line: true values

Black Curve: Results with 'known variances' $\sigma_i^2 = 0.1^2$, ($>$ fit)

When 'known' σ is off, under/over estimate errors in fit.

Summary

Statistics

1. *Multiplicative* mean modeling:

   log-Normal hierarchical model.

Summary

Statistics

1. *Multiplicative* mean modeling:

   log-Normal hierarchical model.

2. Shrinkage estimators.

Summary

Statistics

1. *Multiplicative* mean modeling:

   log-Normal hierarchical model.

2. Shrinkage estimators.
3. Bayesian computation: MCMC & STAN.

Summary

Statistics

1. *Multiplicative* mean modeling:

   log-Normal hierarchical model.

2. Shrinkage estimators.
3. Bayesian computation: MCMC & STAN.
4. The potential pitfalls of assuming 'known' variances.

Summary

Statistics

1. *Multiplicative* mean modeling:

   log-Normal hierarchical model.

2. Shrinkage estimators.
3. Bayesian computation: MCMC & STAN.
4. The potential pitfalls of assuming 'known' variances.

Astronomy

1. Adjustments of effective areas of each instrument.

Summary

Statistics

1. *Multiplicative* mean modeling:

   log-Normal hierarchical model.

2. Shrinkage estimators.
3. Bayesian computation: MCMC & STAN.
4. The potential pitfalls of assuming 'known' variances.

Astronomy

1. Adjustments of effective areas of each instrument.
2. Calibration concordance.

Acknowledgement

Yang Chen (UMich), Xufei Wang (Two Sigma), Xiao-Li Meng (Harvard), David van Dyk (ICL), Herman Marshall (MIT) & Vinay Kashyap (cfA)