The Bayesian Statistics behind Calibration Concordance

Yang Chen

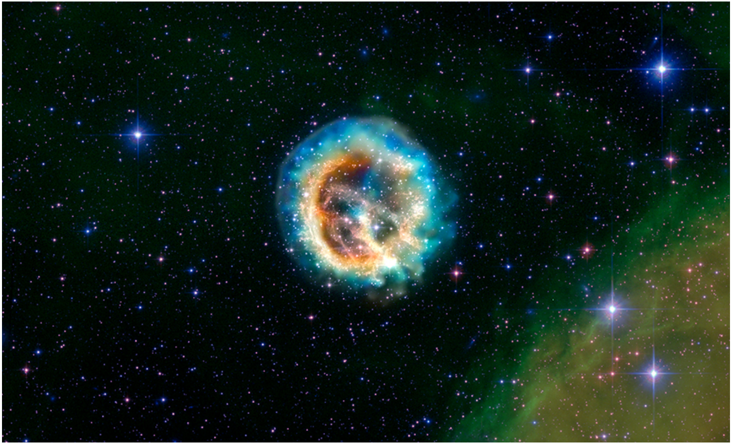
Harvard University

June 5, 2017

Outline

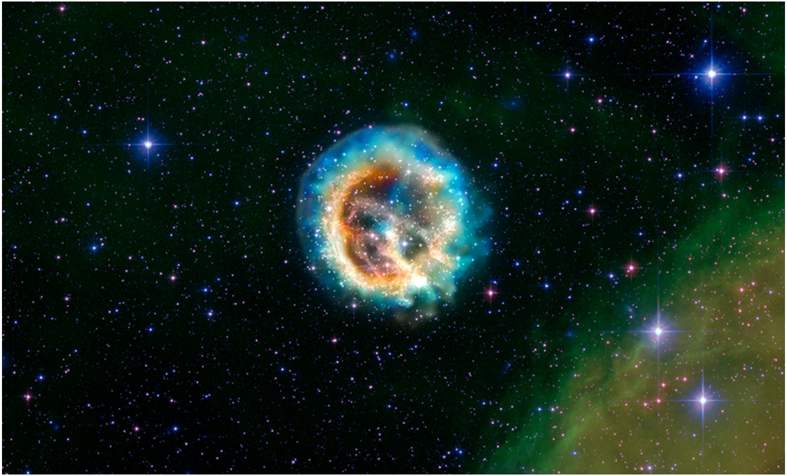
Calibration Concordance Problem (Example: E0102)

E0102 – the remnant of a supernova that exploded in a neighboring galaxy known as the Small Magellanic Cloud.



Calibration Concordance Problem (Example: E0102)

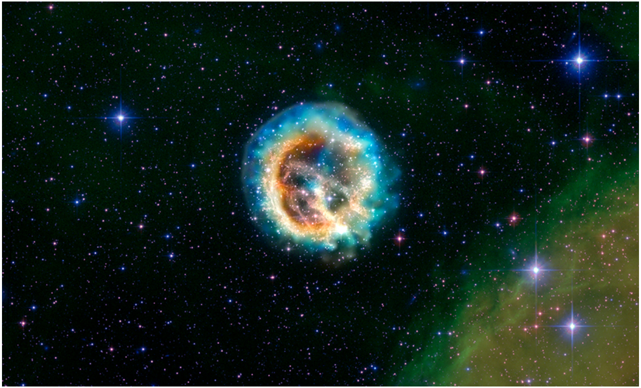
Four “sources” – spectral lines that appear in the E0102 spectrum.



Calibration Concordance Problem (Example: E0102)

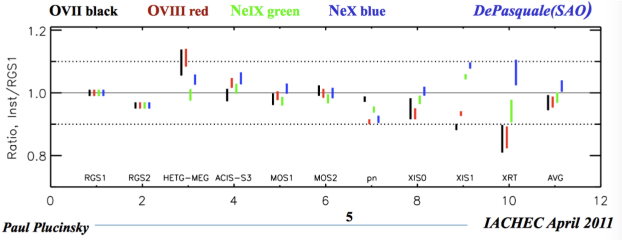
2 lines — Hydrogen like O VIII at 18.969Å & the resonance line of O VII from the Helium like triplet at 21.805Å.

2 lines – Hydrogen like Ne X at 12.135Å & the resonance line of Ne IX from the Helium like triplet at 13.447Å.



Calibration Concordance Problem (Example: E0102)

13 detectors over 4 telescopes, *Chandra* (ACIS-S with and without HETG, and ACIS-I), XMM-*Newton* (RGS, EPIC-MOS, EPIC-pn), *Suzaku* (XIS), and *Swift* (XRT). (Plucinsky et al. 2017).



i ≡ [RGS1, RGS2, HETG-MEG, ACIS-S3, MOS1, MOS2, pn, XIS0, XIS1, XRT] **x** [560-574 eV, 654 eV, 905-922 eV, 1022 eV] (i=1..10,11..20,21..30,31..40)

j ≡ E0102 fluxes in [OVII, OVIII, NeIX, NeX] (j=1..4)

- $c_{1,1}$ = observed counts in RGS2/[560-574 eV], $c_{12,2}$ = in HETG-MEG/[654 eV], $c_{23,3}$ = in ACIS-S3/[905-922 eV], etc.
- a_i = effective area, f_j = expected flux, α_{ij} = exposure time of instrument i for source j (in this case, $\alpha_{k(\cdot)}$ are identical for $k=\{l, l+10, l+20, l+30\}$, l=1..10)

Calibration Concordance Problem (Example: E0102)

Notations

- 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.

Calibration Concordance Problem (Example: E0102)

Notations

- 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.

Calibration Concordance Problem (Example: E0102)

Notations

- 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.

Calibration Concordance Problem (Example: E0102)

Notations

- 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.
- Lower cases: data / estimators. Upper cases: parameter / estimand.

Calibration Concordance Problem (Example: E0102)

Notations

- 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.

Original Questions

Systematic errors in comparing effective areas $\Rightarrow$ absolute measurements.

Calibration Concordance Problem (Example: E0102)

Notations

- 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.
- Lower cases: data / estimators. Upper cases: parameter / estimand.

Original Questions

Systematic errors in comparing effective areas $\Rightarrow$ absolute measurements.

1. How to adjust A_i s.t. $c_{ij}/A_i \approx F_j$ within statistical uncertainty?
2. How to estimate the systematic error on the A_i?

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.

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

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

, ,

.

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

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

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

, ,

.

Setting up priors for unknowns.

1.
2.
3.

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

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

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

, ,

.

Setting up priors for unknowns.

1. Prior for log-flux G_j: flat (improper, non-informative).
2. Prior for log-area B_i: $\mathcal{N}(b_i, \tau_i^2)$ (conjugate, proper).
3.

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

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

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

$$B_i \overset{\text{indep}}{\sim} N(b_i,\ \tau_i^2),\ G_j \overset{\text{indep}}{\sim} \text{flat prior},$$

.

Setting up priors for unknowns.

1. Prior for log-flux G_j: flat (improper, non-informative).
2. Prior for log-area B_i: $\mathcal{N}(b_i, \tau_i^2)$ (conjugate, proper).
3.

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

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

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

$$B_i \overset{\text{indep}}{\sim} N(b_i,\ \tau_i^2),\ G_j \overset{\text{indep}}{\sim} \text{flat prior},$$

.

Setting up priors for unknowns.

1. Prior for log-flux G_j: flat (improper, non-informative).
2. Prior for log-area B_i: $\mathcal{N}(b_i, \tau_i^2)$ (conjugate, proper).
3. Unknown variance: Prior for σ_i^2: inverse Gamma (conjugate, proper).

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

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

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

$$B_i \overset{\text{indep}}{\sim} N(b_i,\ \tau_i^2),\ G_j \overset{\text{indep}}{\sim} \text{flat prior},$$

$$\text{Unknown variance: } \sigma_i^2 \overset{\text{indep}}{\sim} \text{Inv-Gamma}(df_g,\ \beta_g).$$

Setting up priors for unknowns.

1. Prior for log-flux G_j: flat (improper, non-informative).
2. Prior for log-area B_i: $\mathcal{N}(b_i, \tau_i^2)$ (conjugate, proper).
3. Unknown variance: Prior for σ_i^2: inverse Gamma (conjugate, proper).

Shrinkage Estimators (Known Variances)

Hierarchical model $\Rightarrow$ Shrinkage estimators [Example: temperature.]
(weighted averages of evidence from 'Prior' and evidence from 'Data').

Shrinkage Estimators (Known Variances)

Hierarchical model $\Rightarrow$ Shrinkage estimators [Example: temperature.] (weighted averages of evidence from 'Prior' and evidence from 'Data').

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

where

$$W_i = \frac{\tau_i^{-2}}{\tau_i^{-2} + |J_i|\sigma_i^{-2}}$$

are the precisions of the direct information in the b_i relative to the indirect information for estimating the B_i with

$$\bar{G}_i = \frac{\sum_{j \in J_i} \widehat{G}_j \sigma_i^{-2}}{\sum_{j \in J_i} \sigma_i^{-2}}, \quad \bar{B}_j = \frac{\sum_{i \in I_j} \widehat{B}_i \sigma_i^{-2}}{\sum_{i \in I_j} \sigma_i^{-2}}, \ \bar{y}'_{i\cdot} = \frac{\sum_{j \in J_i} y'_{ij} \sigma_i^{-2}}{\sum_{j \in J_i} \sigma_i^{-2}}, \quad \bar{y}'_{\cdot j} = \frac{\sum_{i \in I_j} y'_{ij} \sigma_i^{-2}}{\sum_{i \in I_j} \sigma_i^{-2}}.$$

Shrinkage Estimators (A special case)

Assume that $G_j = g_j$ is known, i.e. fluxes known apriori. Then

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

where $\tilde{c}_{i \cdot}$ and $\tilde{f}_i$ are the geometric means,

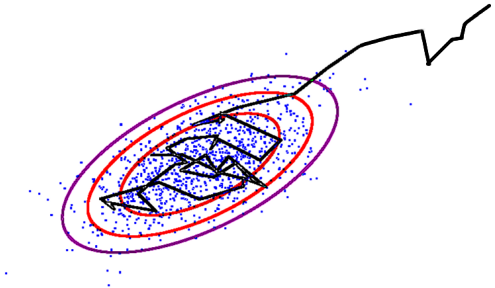
$$\tilde{c}_{i \cdot} = \left[\prod_{j \in J_i} c_{ij} \right]^{1/M_i} \quad \text{and} \quad \tilde{f}_i = \left[\prod_{j \in J_i} f_j \right]^{1/M_i}.$$

Bayesian Computation: MCMC

Markov chain Monte Carlo

Construct a biased random walk that explores target dist $P^\star(x)$

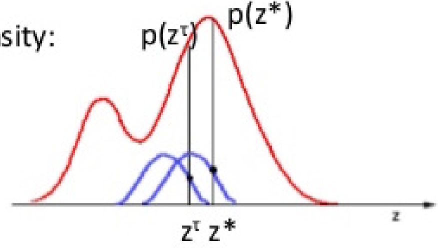
Markov steps, $x_t \sim T(x_t \leftarrow x_{t-1})$



MCMC gives approximate, correlated samples from $P^\star(x)$

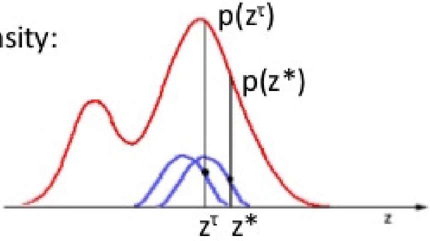
Bayesian Computation: MCMC

Increase in density:



$\alpha = 1$

Decrease in density:



$$\alpha = \frac{\tilde{p}(z^*)}{\tilde{p}(z^\tau)}$$

M. Dümcke

Bayesian Computation (Unknown Variances)

Markov Chain Monte Carlo (MCMC) algorithms.

Bayesian Computation (Unknown Variances)

Markov Chain Monte Carlo (MCMC) algorithms.

- Gibbs Sampling: update parameters one-at-a-time.

Bayesian Computation (Unknown Variances)

Markov Chain Monte Carlo (MCMC) algorithms.

- Gibbs Sampling: update parameters one-at-a-time.
- Block Gibbs Sampling: update vectors of parameters.

Bayesian Computation (Unknown Variances)

Markov Chain Monte Carlo (MCMC) algorithms.

- Gibbs Sampling: update parameters one-at-a-time.
- Block Gibbs Sampling: update vectors of parameters.
  - The joint distribution of the B_i and G_j is Gaussian.

Bayesian Computation (Unknown Variances)

Markov Chain Monte Carlo (MCMC) algorithms.

- Gibbs Sampling: update parameters one-at-a-time.
- Block Gibbs Sampling: update vectors of parameters.
  - The joint distribution of the B_i and G_j is Gaussian.
- Hamiltonian Monte Carlo (HMC) – STAN package.

Bayesian Computation (Unknown Variances)

Markov Chain Monte Carlo (MCMC) algorithms.

- Gibbs Sampling: update parameters one-at-a-time.
- Block Gibbs Sampling: update vectors of parameters.
  - The joint distribution of the B_i and G_j is Gaussian.
- Hamiltonian Monte Carlo (HMC) – STAN package.
  - Highly correlated parameters, high-dim parameter space.

Bayesian Computation (STAN)

From STAN homepage —

Users specify log density functions in Stan's probabilistic programming language and get:

- full Bayesian statistical inference with MCMC sampling (NUTS, HMC)
- approximate Bayesian inference with variational inference (ADVI)
- penalized maximum likelihood estimation with optimization (L-BFGS)

Bayesian Computation (STAN Example)

Start by writing a Stan program for the model.

```
// saved as 8schools.stan
data {
  int<lower=0> J; // number of schools
  real y[J]; // estimated treatment effects
  real<lower=0> sigma[J]; // s.e. of effect estimates
}
parameters {
  real mu;
  real<lower=0> tau;
  real eta[J];
}
transformed parameters {
  real theta[J];
  for (j in 1:J)
    theta[j] = mu + tau * eta[j];
}
model {
  target += normal_lpdf(eta | 0, 1);
  target += normal_lpdf(y | theta, sigma);
}
```

Bayesian Computation (STAN Example)

Assuming we have the 8schools.stan file in our working directory, we can prepare the data and fit the model as the following R code shows.

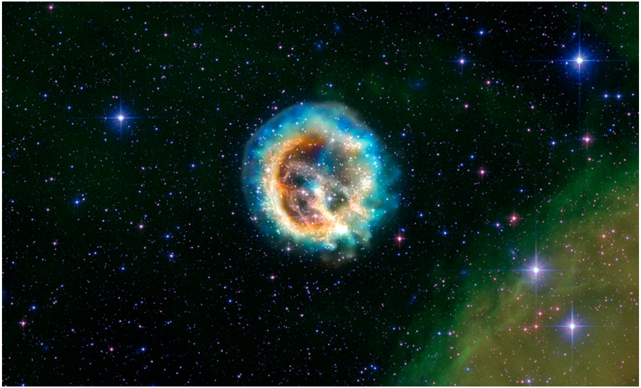
```
schools_dat <- list(J = 8,
                    y = c(28,  8, -3,  7, -1,  1, 18, 12),
                    sigma = c(15, 10, 16, 11,  9, 11, 10, 18))

fit <- stan(file = '8schools.stan', data = schools_dat,
            iter = 1000, chains = 4)
```

Bayesian Computation (STAN Example)

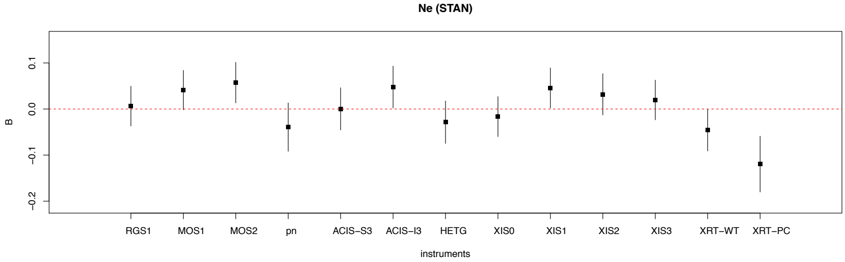
```
> print(fit, digits = 1)
Inference for Stan model: 8schools.
4 chains, each with iter=1000; warmup=500; thin=1;
post-warmup draws per chain=500, total post-warmup draws=2000.

          mean se_mean  sd  2.5%   25%   50%   75% 97.5% n_eff Rhat
mu         8.2     0.2 5.4  -1.9   4.8   8.1  11.3  19.3   480    1
tau        6.8     0.3 6.2   0.3   2.5   5.2   9.2  21.7   425    1
eta[1]     0.4     0.0 1.0  -1.5  -0.3   0.4   1.0   2.2  2000    1
eta[2]     0.0     0.0 0.8  -1.7  -0.6   0.0   0.5   1.7  2000    1
eta[3]    -0.2     0.0 1.0  -2.1  -0.9  -0.2   0.4   1.7  2000    1
eta[4]    -0.1     0.0 0.9  -1.8  -0.7  -0.1   0.5   1.7  2000    1
eta[5]    -0.4     0.0 0.9  -2.1  -1.0  -0.4   0.2   1.4  2000    1
eta[6]    -0.2     0.0 0.9  -1.9  -0.8  -0.2   0.4   1.5  1731    1
eta[7]     0.3     0.0 0.9  -1.4  -0.2   0.4   0.9   2.0  1507    1
eta[8]     0.0     0.0 0.9  -1.9  -0.6   0.0   0.7   1.8  1988    1
theta[1]  11.5     0.3 8.8  -2.4   5.9  10.1  15.6  32.9   977    1
theta[2]   7.8     0.1 6.2  -4.7   4.1   7.9  11.6  20.3  2000    1
theta[3]   6.1     0.2 7.7 -11.2   2.1   6.4  10.5  20.2  2000    1
theta[4]   7.6     0.1 6.5  -4.9   3.8   7.8  11.4  21.3  2000    1
theta[5]   5.0     0.1 6.6  -9.3   1.2   5.6   9.3  16.7  2000    1
theta[6]   6.2     0.2 6.7  -8.2   2.2   6.5  10.5  18.5  2000    1
theta[7]  10.8     0.2 7.0  -1.3   6.1  10.1  15.1  26.8  2000    1
theta[8]   8.7     0.2 8.2  -7.3   3.9   8.4  12.8  27.2  1446    1
lp__     -39.5     0.1 2.6 -45.1 -41.2 -39.4 -37.7 -35.1   590    1

Samples were drawn using NUTS(diag_e) at Fri May  5 10:41:43 2017.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

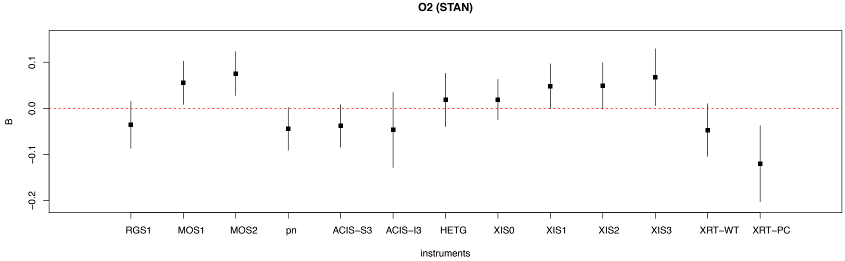
Numerical Results (E0102)

Recap: Highly ionized Oxygen (2 lines). Neon (2 lines). 13 detectors over 4 telescopes, *Chandra* (ACIS-S with & without HETG, ACIS-I), XMM - *Newton* (RGS, EPIC-MOS, EPIC-pn), *Suzaku* (XIS), & *Swift* (XRT).



Numerical Results (E0102)





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.

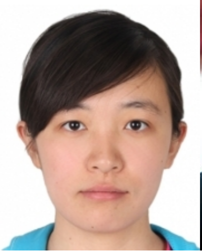
2. Shrinkage estimators.
3. Bayesian computation: MCMC & STAN.
4. The potential pitfalls of assuming 'known' variances.

Astronomy

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

Acknowledgement

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





Numerical Results (XCAL)

- **XCAL data**: Bright active galactic nuclei from the XMM-Newton cross-calibration sample.

Numerical Results (XCAL)

- **XCAL data**: Bright active galactic nuclei from the XMM-Newton cross-calibration sample.
- **The "pileup"**: Image data are clipped to eliminate the regions affected by pileup, determined using `epatplot`.

Numerical Results (XCAL)

- **XCAL data**: Bright active galactic nuclei from the XMM-Newton cross-calibration sample.
- **The "pileup"**: Image data are clipped to eliminate the regions affected by pileup, determined using `epatplot`.
- **Three data sets**: the hard, medium, and soft energy bands.

Numerical Results (XCAL)

- **XCAL data**: Bright active galactic nuclei from the XMM-Newton cross-calibration sample.
- **The "pileup"**: Image data are clipped to eliminate the regions affected by pileup, determined using `epatplot`.
- **Three data sets**: the hard, medium, and soft energy bands.
- **Three detectors**: MOS1, MOS2 and pn.

Numerical Results (XCAL)

- **XCAL data**: Bright active galactic nuclei from the XMM-Newton cross-calibration sample.
- **The "pileup"**: Image data are clipped to eliminate the regions affected by pileup, determined using `epatplot`.
- **Three data sets**: the hard, medium, and soft energy bands.
- **Three detectors**: MOS1, MOS2 and pn.
- **Sources**: 94 (hard band), 103 (medium band), and 108 (soft band).

Numerical Results (XCAL)