Seeking Effective Adjustments for Effective Areas

Xiao-Li Meng
Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul ....

Department of Statistics, Harvard University

October 5, 2015

A problem posed by Herman, Matteo and Vinay

Xiao-Li Meng Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul ....

A problem posed by Herman, Matteo and Vinay

Xiao-Li Meng
Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul ....

Systematic errors in comparing effective areas:
Speaking hypothetically, if we label the instruments by numbers $i = 1, \ldots, N$ and each has an attribute A that is used to measure the same $j = 1, \ldots, M$ astrophysical sources, with intrinsic attribute F_j where $C_{ij} = A_i F_j$ are the instrumental measurements, then the question is: "Is there a way to decide how (or whether) to change A_i when the values C_{ij}/A_i do not agree with F_j to within their statistical uncertainties s_i. In other words, each instrument provides an estimator f_j of F_j with statistical uncertainty s_j but $|f_j - F_j|/s_j$ is often large, not distributed as a Gaussian with unit variance (but can have zero mean if we define $F_j = \sum_j f_j s_j^{-2} / \sum_j s_j^{-2}$). How to estimate the systematic error on the A_i?

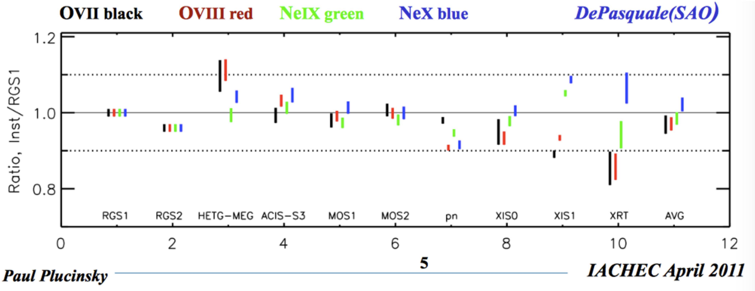
Instruments (i) and Sources (j)

- i are individual detectors (e.g., *Chandra*/ACIS-I, *Chandra*/HEG, *XMM*/EPIC-pn, *XMM*/EPIC-MOS1, *XMM*/RGS2, *Swift*/XRT, *Suzaku*/XIS, *NuSTAR*/FPMA, *Integral*/ISGRI, etc.), with counts obtained in specific passbands (e.g., soft=[0.5-2 keV], hard=[2-7 keV], ultra=[10-30 keV], etc.)

- j are individual sources (HZ 43, Capella, PKS 2155-304, Mkn 421, Crab, G21.5-09, etc.) with fluxes predicted in specific passbands

From Vinay ...

Xiao-Li Meng Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul ....



$i \equiv$ [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)

$j \equiv$ 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)

Avoiding mixing (deterministic) Estimands with (random) Estimators

Xiao-Li Meng
Working with
Herman
Marshall &
Matteo
Guainazzi,
Vinay, Aneta,
Jeremy, Paul
....

Avoiding mixing (deterministic) Estimands with (random) Estimators

Xiao-Li Meng
Working with
Herman
Marshall &
Matteo
Guainazzi,
Vinay, Aneta,
Jeremy, Paul
....

Use upper case for estimand/parameter; lower cases for estimator/data

Avoiding mixing (deterministic) Estimands with (random) Estimators

Xiao-Li Meng Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul ....

Use upper case for estimand/parameter; lower cases for estimator/data

- Let A_i be the *actual* effective area of instrument i; F_j be the *true* flux of source j; then the *expected* rate can be modelled as

$$C_{ij} = A_i F_j \quad \text{or} \quad \log C_{ij} = \log A_i + \log F_j$$

Avoiding mixing (deterministic) Estimands with (random) Estimators

Use upper case for estimand/parameter; lower cases for estimator/data

- Let A_i be the *actual* effective area of instrument i; F_j be the *true* flux of source j; then the *expected* rate can be modelled as

$$C_{ij} = A_i F_j \quad \text{or} \quad \log C_{ij} = \log A_i + \log F_j$$

- Let a_i be an estimator of A_i; f_j an estimator of F_j, and c_{ij} be the actual observation from source j detected by instrument i. Then it is *NOT* reasonable to expect $c_{ij} \approx a_i f_j$, in the sense of justifying the "regression" model

$$\log c_{ij} = \log a_i + \log f_j + e_{ij}, \quad E(e_{ij}) = 0.$$

Distributions cannot be manipulated as numbers

Xiao-Li Meng Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul
....

For (deterministic) numbers Y and X

$$\textit{If} \quad Y = \rho X, \quad \text{then} \quad X = \rho^{-1} Y.$$

Distributions cannot be manipulated as numbers

For (deterministic) numbers Y and X

$$\textit{If} \quad Y = \rho X, \quad \text{then} \quad X = \rho^{-1} Y.$$

For distributional (random variables) Y and X

If regressing Y on X yields (both have zero mean and unit var):

$$Y = \rho X,$$

Then regressing X on Y is NOT $X = \rho^{-1} Y$, but rather

$$X = \rho Y.$$

Here ρ is the *correlation* between X and Y .

Distributions cannot be manipulated as numbers

Xiao-Li Meng Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul
....

For (deterministic) numbers Y and X

$$\textit{If} \quad Y = \rho X, \quad \text{then} \quad X = \rho^{-1} Y.$$

For distributional (random variables) Y and X

If regressing Y on X yields (both have zero mean and unit var):

$$Y = \rho X,$$

Then regressing X on Y is NOT $X = \rho^{-1} Y$, but rather

$$X = \rho Y.$$

Here ρ is the *correlation* between X and Y .

- Do not follow "The Rule of Three" (Stephen Stigler, *Seven Pillars of Statistics*; ASA President Address, 2014).

Setting up an Appropriate Model for Calibration

Setting up an Appropriate Model for Calibration

Notation is important

$$b_i = \log a_i, \quad B_i = \log A_i, \quad i \in I = \{1, \ldots, N\}.$$

Setting up an Appropriate Model for Calibration

Notation is important

$$b_i = \log a_i, \quad B_i = \log A_i, \quad i \in I = \{1, \ldots, N\}.$$

$$g_j = \log f_j, \quad G_j = \log F_j, \quad j \in J = \{1, \ldots, M\}.$$

Setting up an Appropriate Model for Calibration

Notation is important

$$b_i = \log a_i, \quad B_i = \log A_i, \quad i \in I = \{1, \ldots, N\}.$$

$$g_j = \log f_j, \quad G_j = \log F_j, \quad j \in J = \{1, \ldots, M\}.$$

$$y_{ij} = \log c_{ij}, \quad i \in I, \quad j \in J_i \subset J.$$

Setting up an Appropriate Model for Calibration

Notation is important

$$b_i = \log a_i, \quad B_i = \log A_i, \quad i \in I = \{1, \ldots, N\}.$$

$$g_j = \log f_j, \quad G_j = \log F_j, \quad j \in J = \{1, \ldots, M\}.$$

$$y_{ij} = \log c_{ij}, \quad i \in I, \quad j \in J_i \subset J.$$

log-normal model for c

$$y_{ij} = \alpha_{ij} + B_i + G_j + e_{ij}$$

Setting up an Appropriate Model for Calibration

Notation is important

$$b_i = \log a_i, \quad B_i = \log A_i, \quad i \in I = \{1, \ldots, N\}.$$
$$g_j = \log f_j, \quad G_j = \log F_j, \quad j \in J = \{1, \ldots, M\}.$$
$$y_{ij} = \log c_{ij}, \quad i \in I, \quad j \in J_i \subset J.$$

log-normal model for c

$$y_{ij} = \alpha_{ij} + B_i + G_j + e_{ij}$$

$$e_{ij} \sim indep\ N(0, \sigma_{ij}^2)$$

Setting up an Appropriate Model for Calibration

Xiao-Li Meng Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul ....

Notation is important

$$b_i = \log a_i, \quad B_i = \log A_i, \quad i \in I = \{1, \ldots, N\}.$$
$$g_j = \log f_j, \quad G_j = \log F_j, \quad j \in J = \{1, \ldots, M\}.$$
$$y_{ij} = \log c_{ij}, \quad i \in I, \quad j \in J_i \subset J.$$

log-normal model for c

$$y_{ij} = \alpha_{ij} + B_i + G_j + e_{ij}$$
$$e_{ij} \sim indep\ N(0, \sigma^2_{ij})$$
$$\alpha_{ij} = -0.5\sigma^2_{ij}.$$

Setting up an Appropriate Model for Calibration

Notation is important

$$b_i = \log a_i, \quad B_i = \log A_i, \quad i \in I = \{1, \ldots, N\}.$$
$$g_j = \log f_j, \quad G_j = \log F_j, \quad j \in J = \{1, \ldots, M\}.$$
$$y_{ij} = \log c_{ij}, \quad i \in I, \quad j \in J_i \subset J.$$

log-normal model for c

$$y_{ij} = \alpha_{ij} + B_i + G_j + e_{ij}$$

$$e_{ij} \sim indep\ N(0, \sigma^2_{ij})$$
$$\alpha_{ij} = -0.5\sigma^2_{ij}.$$

- Model I: $\sigma^2_{ij} = \sigma^2_i$

Setting up an Appropriate Model for Calibration

Notation is important

$$b_i = \log a_i, \quad B_i = \log A_i, \quad i \in I = \{1, \ldots, N\}.$$

$$g_j = \log f_j, \quad G_j = \log F_j, \quad j \in J = \{1, \ldots, M\}.$$

$$y_{ij} = \log c_{ij}, \quad i \in I, \quad j \in J_i \subset J.$$

log-normal model for c

$$y_{ij} = \alpha_{ij} + B_i + G_j + e_{ij}$$

$$e_{ij} \sim indep\ N(0, \sigma_{ij}^2)$$

$$\alpha_{ij} = -0.5\sigma_{ij}^2.$$

- Model I: $\sigma_{ij}^2 = \sigma_i^2$
- Model II: $\sigma_{ij}^2 = \sigma_i^2 + \sigma_j^2$

Setting up an Appropriate Model for Calibration

Xiao-Li Meng Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul
....

Notation is important

$$b_i = \log a_i, \quad B_i = \log A_i, \quad i \in I = \{1, \ldots, N\}.$$

$$g_j = \log f_j, \quad G_j = \log F_j, \quad j \in J = \{1, \ldots, M\}.$$

$$y_{ij} = \log c_{ij}, \quad i \in I, \quad j \in J_i \subset J.$$

log-normal model for c

$$y_{ij} = \alpha_{ij} + B_i + G_j + e_{ij}$$

$$e_{ij} \sim indep\ N(0, \sigma_{ij}^2)$$

$$\alpha_{ij} = -0.5\sigma_{ij}^2.$$

- Model I: $\sigma_{ij}^2 = \sigma_i^2$
- Model II: $\sigma_{ij}^2 = \sigma_i^2 + \sigma_j^2$

log-normal models for a and f

Setting up an Appropriate Model for Calibration

Notation is important

$$b_i = \log a_i, \quad B_i = \log A_i, \quad i \in I = \{1, \ldots, N\}.$$

$$g_j = \log f_j, \quad G_j = \log F_j, \quad j \in J = \{1, \ldots, M\}.$$

$$y_{ij} = \log c_{ij}, \quad i \in I, \quad j \in J_i \subset J.$$

log-normal model for c

$$y_{ij} = \alpha_{ij} + B_i + G_j + e_{ij}$$

$$e_{ij} \sim indep\ N(0, \sigma^2_{ij})$$

$$\alpha_{ij} = -0.5\sigma^2_{ij}.$$

- Model I: $\sigma^2_{ij} = \sigma^2_i$
- Model II: $\sigma^2_{ij} = \sigma^2_i + \sigma^2_j$

log-normal models for a and f

$$b_i = \beta_i + B_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \tau^2_i), \ \beta_i = -0.5\tau^2_i$$

Setting up an Appropriate Model for Calibration

Notation is important

$$b_i = \log a_i, \quad B_i = \log A_i, \quad i \in I = \{1, \dots, N\}.$$

$$g_j = \log f_j, \quad G_j = \log F_j, \quad j \in J = \{1, \dots, M\}.$$

$$y_{ij} = \log c_{ij}, \quad i \in I, \quad j \in J_i \subset J.$$

log-normal model for c

$$y_{ij} = \alpha_{ij} + B_i + G_j + e_{ij}$$

$$e_{ij} \sim indep\ N(0, \sigma^2_{ij})$$

$$\alpha_{ij} = -0.5\sigma^2_{ij}.$$

- Model I: $\sigma^2_{ij} = \sigma^2_i$
- Model II: $\sigma^2_{ij} = \sigma^2_i + \sigma^2_j$

log-normal models for a and f

$$b_i = \beta_i + B_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \tau^2_i), \ \beta_i = -0.5\tau^2_i$$

$$g_j = \gamma_j + G_j + \delta_j$$

$$\delta_j \sim N(0, \eta^2_j), \ \gamma_j = -0.5\eta^2_j$$

Setting up an Appropriate Model for Calibration

Xiao-Li Meng Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul ....

Notation is important

$$b_i = \log a_i, \quad B_i = \log A_i, \quad i \in I = \{1, \ldots, N\}.$$

$$g_j = \log f_j, \quad G_j = \log F_j, \quad j \in J = \{1, \ldots, M\}.$$

$$y_{ij} = \log c_{ij}, \quad i \in I, \quad j \in J_i \subset J.$$

log-normal model for c

$$y_{ij} = \alpha_{ij} + B_i + G_j + e_{ij}$$

$$e_{ij} \sim indep\ N(0, \sigma^2_{ij})$$

$$\alpha_{ij} = -0.5\sigma^2_{ij}.$$

- Model I: $\sigma^2_{ij} = \sigma^2_i$
- Model II: $\sigma^2_{ij} = \sigma^2_i + \sigma^2_j$

log-normal models for a and f

$$b_i = \beta_i + B_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \tau_i^2), \ \beta_i = -0.5\tau_i^2$$

$$g_j = \gamma_j + G_j + \delta_j$$

$$\delta_j \sim N(0, \eta_j^2), \ \gamma_j = -0.5\eta_j^2$$

- set $\eta_j = \infty$ when f_j's are estimated by $\{y_{ij}\}$

Why do we need the half-variance correction?

Why do we need the half-variance correction?

Xiao-Li Meng
Working with
Herman
Marshall &
Matteo
Guainazzi,
Vinay, Aneta,
Jeremy, Paul
....

This is because, by Jensen's inequality,

$$E(c_{ij}) = E(e^{y_{ij}}) > e^{E(y_{ij})}$$

Why do we need the half-variance correction?

Xiao-Li Meng
Working with
Herman
Marshall &
Matteo
Guainazzi,
Vinay, Aneta,
Jeremy, Paul
....

This is because, by Jensen's inequality,

$$E(c_{ij}) = E(e^{y_{ij}}) > e^{E(y_{ij})}$$

It is well known that

$$\textit{if} \quad \log X \sim N(\mu, \sigma^2), \quad \textit{then} \quad E(X) = e^{u + \frac{\sigma^2}{2}}$$

Why do we need the half-variance correction?

Xiao-Li Meng
Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul
....

This is because, by Jensen's inequality,

$$E(c_{ij}) = E(e^{y_{ij}}) > e^{E(y_{ij})}$$

It is well known that

$$\textit{if} \quad \log X \sim N(\mu, \sigma^2), \quad \textit{then} \quad E(X) = e^{u+\frac{\sigma^2}{2}}$$

Therefore, when we set

$$E(y_{ij}) = -0.5\sigma_{ij}^2 + B_i + G_j,$$

$$\textit{We obtain} \;\; E(c_{ij}) = E(e^{y_{ij}}) = e^{E(y_{ij})+\sigma_{ij}^2/2} = e^{B_i+G_j} = A_i F_j.$$

Xiao-Li Meng Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul ....

Why do we need the half-variance correction?

This is because, by Jensen's inequality,

$$E(c_{ij}) = E(e^{y_{ij}}) > e^{E(y_{ij})}$$

It is well known that

$$\textit{if} \quad \log X \sim N(\mu, \sigma^2), \quad \textit{then} \quad E(X) = e^{u+\frac{\sigma^2}{2}}$$

Therefore, when we set

$$E(y_{ij}) = -0.5\sigma_{ij}^2 + B_i + G_j,$$

$$\textit{We obtain} \;\; E(c_{ij}) = E(e^{y_{ij}}) = e^{E(y_{ij})+\sigma_{ij}^2/2} = e^{B_i+G_j} = A_i F_j.$$

When variances are known, simply "correct" the data

$$y'_{ij} = y_{ij} + 0.5\sigma_{ij}^2$$

Shrinkage estimators: combining information

Shrinkage estimators: combining information

Two sources of information

- *Prior/other-data estimator* for B_i:

$$\hat{B}_i^{prior} = b_i', \quad \text{with} \;\; \mathrm{Var} = \tau_i^2$$

Shrinkage estimators: combining information

Xiao-Li Meng Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul ....

Two sources of information

- *Prior/other-data estimator* for B_i:

$$\hat{B}_i^{prior} = b_i', \quad \text{with} \;\; \mathrm{Var} = \tau_i^2$$

- *Calibration-data estimator* for B_i:

Shrinkage estimators: combining information

Xiao-Li Meng
Working with
Herman
Marshall &
Matteo
Guainazzi,
Vinay, Aneta,
Jeremy, Paul
....

Two sources of information

- *Prior/other-data estimator* for B_i:

$$\hat{B}_i^{prior} = b_i', \quad \text{with Var} = \tau_i^2$$

- *Calibration-data estimator* for B_i:

$$\hat{B}_i^{data} = \bar{y}_{i.}' - \bar{G}, \quad \text{with Var} = \frac{\sigma_i^2}{M}$$

- **Relative precision:** $w_i = \tau_i^{-2} / (\tau_i^{-2} + M\sigma_i^{-2})$

Shrinkage estimators: combining information

Two sources of information

- *Prior/other-data estimator* for B_i:

$$\hat{B}_i^{prior} = b_i', \quad \text{with } \text{Var} = \tau_i^2$$

- *Calibration-data estimator* for B_i:

$$\hat{B}_i^{data} = \bar{y}_{i.}' - \bar{G}, \quad \text{with } \text{Var} = \frac{\sigma_i^2}{M}$$

- **Relative precision:** $w_i = \tau_i^{-2}/(\tau_i^{-2} + M\sigma_i^{-2})$

Maximum Likelihood Estimation: Linear shrinkage on log-scale

$$\hat{B}_i - \bar{B} = w_i(b_i' - \bar{B}) + (1 - w_i)(\bar{y}_{i.}' - \bar{y}')$$

Shrinkage estimators: combining information

Two sources of information

- *Prior/other-data estimator* for B_i:

$$\hat{B}_i^{prior} = b'_i, \quad \text{with } \text{Var} = \tau_i^2$$

- *Calibration-data estimator* for B_i:

$$\hat{B}_i^{data} = \bar{y}'_{i.} - \bar{G}, \quad \text{with } \text{Var} = \frac{\sigma_i^2}{M}$$

- **Relative precision:** $w_i = \tau_i^{-2}/(\tau_i^{-2} + M\sigma_i^{-2})$

Maximum Likelihood Estimation: Linear shrinkage on log-scale

$$\hat{B}_i - \bar{B} = w_i(b'_i - \bar{B}) + (1 - w_i)(\bar{y}'_{i.} - \bar{y}')$$

$$\bar{y}' = \frac{\sum_i \bar{y}'_{i.}\sigma_i^{-2}}{\sum_i \sigma_i^{-2}}, \quad \text{and} \quad \bar{B} = \frac{\sum_i \hat{B}_i\sigma_i^{-2}}{\sum_i \sigma_i^{-2}}.$$

When variance is unknown ...

When variance is unknown ...

The MLE is also a (non-linear) shrinkage estimator

$$\hat{\sigma}_i^2 = 2\left[\sqrt{1 + S_{y,i}^2} - 1\right] \equiv R_i S_{y,i}^2,$$

$$R_i = \frac{2}{1 + \sqrt{1 + S_{y,i}^2}} \leq 1$$

$$\begin{aligned} S_{y,i}^2 &= \frac{1}{M}\sum_{j=1}^{M}(y_{ij} - \hat{B}_i - \hat{G}_j)^2 \\ &= \frac{1}{M}\sum_{j=1}^{M}[y_{ij} - \bar{y}'_{\cdot j} - (\hat{B}_i - \bar{B})]^2 \end{aligned}$$

When variance is unknown ...

The MLE is also a (non-linear) shrinkage estimator

$$\hat{\sigma}_i^2 = 2\left[\sqrt{1 + S_{y,i}^2} - 1\right] \equiv R_i S_{y,i}^2,$$

$$R_i = \frac{2}{1 + \sqrt{1 + S_{y,i}^2}} \leq 1$$

$$\begin{aligned} S_{y,i}^2 &= \frac{1}{M}\sum_{j=1}^{M}(y_{ij} - \hat{B}_i - \hat{G}_j)^2 \\ &= \frac{1}{M}\sum_{j=1}^{M}[y_{ij} - \bar{y}'_{\cdot j} - (\hat{B}_i - \bar{B})]^2 \end{aligned}$$

$$\bar{y}'_{\cdot j} = \frac{\sum_i y'_{ij}\hat{\sigma}_i^{-2}}{\sum_i \hat{\sigma}_i^{-2}} = \frac{\sum_i y_{ij}\hat{\sigma}_i^{-2} + 0.5}{\sum_i \hat{\sigma}_i^{-2}}$$

How about Poisson Model?

How about Poisson Model?

Directly model the counts

$$c_{ij} \sim Poisson(C_{ij})$$

How about Poisson Model?

Directly model the counts

$$c_{ij} \sim Poisson(C_{ij})$$

- Model $C_{ij} = A_i F_j$?

How about Poisson Model?

Directly model the counts

$$c_{ij} \sim Poisson(C_{ij})$$

- Model $C_{ij} = A_i F_j$?
- But need to take care of the model error/imperfection:

$$C_{ij} = \alpha_{ij} A_i F_j + \beta_{ij}, \qquad \beta_{ij} - background\ rate$$

How about Poisson Model?

Xiao-Li Meng Working with Herman Marshall & Matteo Guainazzi, Vinay, Aneta, Jeremy, Paul ....

Directly model the counts

$$c_{ij} \sim Poisson(C_{ij})$$

- Model $C_{ij} = A_i F_j$?
- But need to take care of the model error/imperfection:

$$C_{ij} = \alpha_{ij} A_i F_j + \beta_{ij}, \qquad \beta_{ij} - background \ \ rate$$

- Do we give a distribution to α_{ij}, and impose $E(\alpha_{ij}) = 1$?

How about Poisson Model?

Directly model the counts

$$c_{ij} \sim Poisson(C_{ij})$$

- Model $C_{ij} = A_i F_j$?
- But need to take care of the model error/imperfection:

$$C_{ij} = \alpha_{ij} A_i F_j + \beta_{ij}, \qquad \beta_{ij} - background\ rate$$

- Do we give a distribution to α_{ij}, and impose $E(\alpha_{ij}) = 1$?
- This will be an over-dispersion model because

$$\mathrm{Var}(c_{ij}) = E(c_{ij}) + A_i^2 F_j^2 \,\mathrm{Var}(\alpha_{ij})$$