Covariances in XML

David Brown

PAD Name - Directorate/Department Name

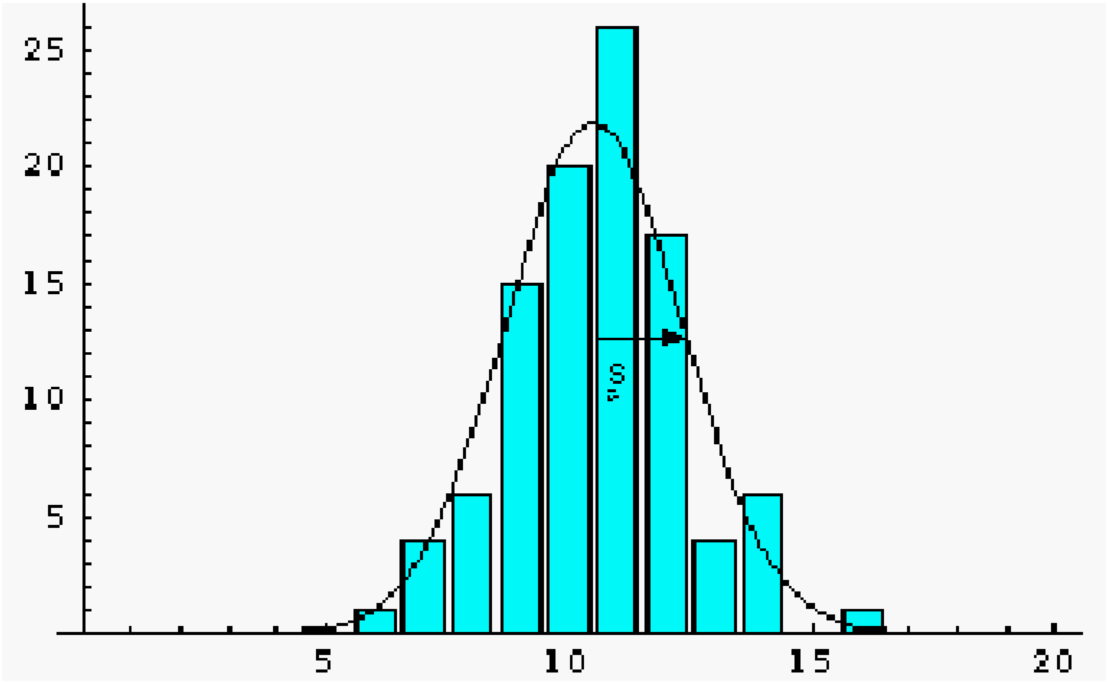
Option: Auspice statement or other directorate information

Lawrence Livermore National Laboratory

Outline

- What are covariances?
- Representing covariance data

Uncertainties, one measurement



- Width of histogram == uncertainty
- (insert verbiage about confidence intervals here)
- We always assume measurements have this Gaussian uncertainty shape

Users understand these

Uncertainties, N measurements

- N independent measurements, N uncertainties:
  - $X_1 \pm \delta X_1$
  - $X_2 \pm \delta X_2$
  - $X_3 \pm \delta X_3$
  - $X_4 \pm \delta X_4$
  - …
- Type A evaluation

Simple uncertainty propagation

- Suppose we have N measurements and we propagate that uncertainty into another parameter:
  - $y = f(x_1, x_2, x_3, \ldots)$
- Want δy, do Taylor series about x_1, etc.:
  - $y = f(x_1,\ldots)+\Sigma_{ij}\ \delta x_i\, \delta x_j\ df\ (x_1,\ldots)/dx_i\ df\ (x_1,\ldots)/dx_j$ + higher order
  - keeping leading order, get standard result:

$$\delta y = \mathrm{sqrt}(\Sigma_{ij}\ \delta x_i\, \delta x_j\ df\ (x_1,\ldots)/dx_i\ df\ (x_1,\ldots)/dx_j\ )$$

- Type B uncertainty

Users start to tune out here

Coupled data and covariance

- Suppose have M measurements, y_j, and they are really a function of N other measurements, x_i. Define the covariance of y_j as
- $\delta^2 y_{jj'} = \Sigma_{ii'} \delta x_i \delta x_{i'} df_j (x_1,\ldots)/dx_i df_{j'} (x_1,\ldots)/dx_i$

- If you see a covariance matrix, think underlying measurement, even if you don't know what it was

Users are probably lost here

Storing it in XEndl

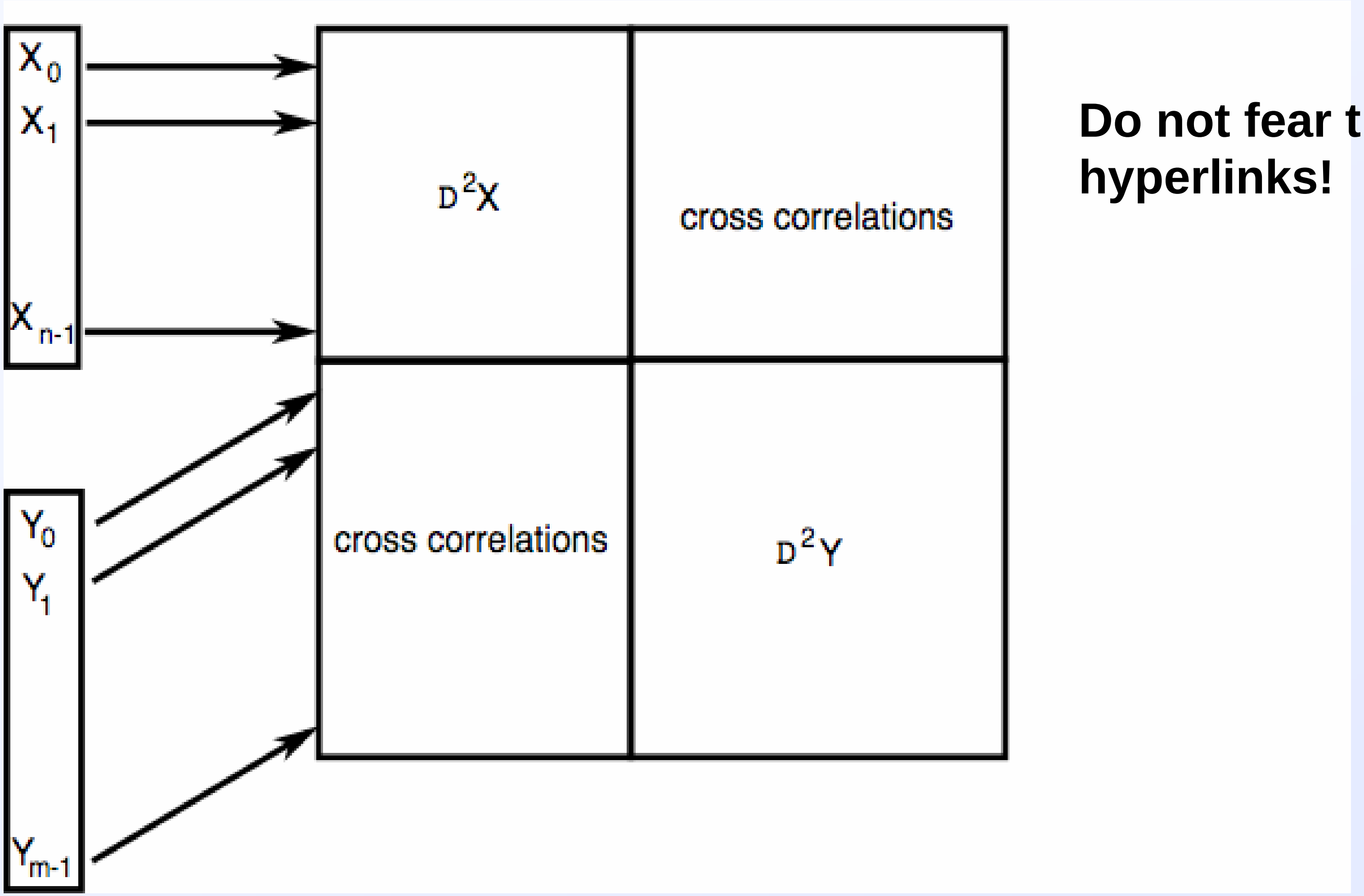
- Matrices can be big, how do we write them?
- How do we pack our data into them?
- XML representation

Storing really really big matrices

Approach	Relation to Covariance Matrix	Pros	Cons
Covariance matrix C	n/a	Simple	Possibly very large; Must synchronize with uncertainties
Correlation matrix R	$C_{ij} = R_{ij}\delta x_i \delta x_j$	Simple; Don't need to syncronize with uncertainty	Possibly very large
Sensitivity matrix O, S	$C_{ij} = \sum_{kl} O^T_{ik} S_{kl} O_{lj}$	May be very compact	Complicated; Must synchronize with uncertainties
Normalized Sensitivity matrix $\hat{O}, S$	$C_{ij} = \sum_{kl} \delta x_i \delta x_j \hat{O}^T_{ik} S_{kl} \hat{O}_{lj}$	May be very compact; Don't need to syncronize with uncertainty	Very complicated

Table 1: Possible approaches to the implementation of covariance matrices. The sensitivity matrix based approaches all require a notion of matrix multiplication which must somehow be denoted in the format and defined in any application code.

Packing the covariance matrix



Do not fear the hyperlinks!

An XML implementation

- Covariance, Correlation, Sensitivity and Normed Sensitivity matrices all can be stored as vanilla array’s (array is a simple array implementation we have written)
- Surround array’s with appropriate tags (covarianceMatrix, etc.)
- Hyperlink uncertainty fields in data to corresponding covariance data
- Coupling data in uncertainty field to specify range in covariance matrix that a certain data set points too (covarianceDatum, covarianceRange)

An XML implementation, cont.

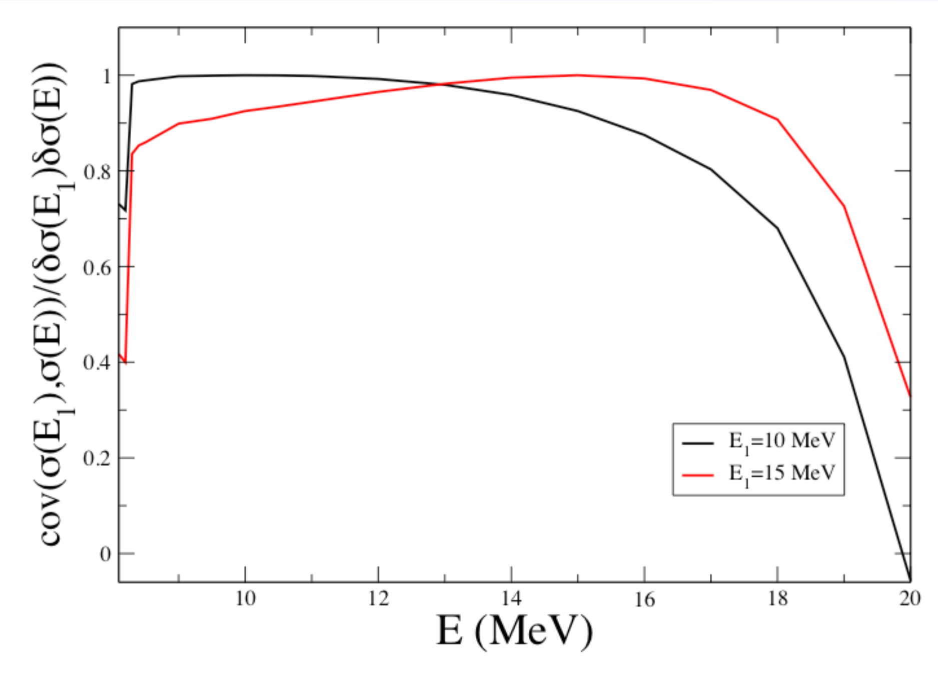
- Can decouple uncertainty from covariance so users don't have to eat it all
- Can "discover" if two sets co-vary by comparing hyperlinks
- Evaluator is charged with doing the actual packing, and the user is charged with doing the unpacking
- Hyperlinks provide elegant solution to cross-material correlations

Backup slides

Generating covariance data

In general this is pretty hard.

Some very simple things have been done so far:



Cov. estimate for ^{74}As(n,2n)

How do you sample several random variables?

i) Independent variables: sample $P_i(x)$ independently

ii) Correlated data:

write $C = A^T A$ (C is the covariance matrix)

and sample using

$x = \langle x \rangle + A^T z$, with z_i a vector of independent unit deviation random variables