Improved (iterative) Bayesian unfolding

Giulio D'Agostini

University and INFN Section of "Roma1"

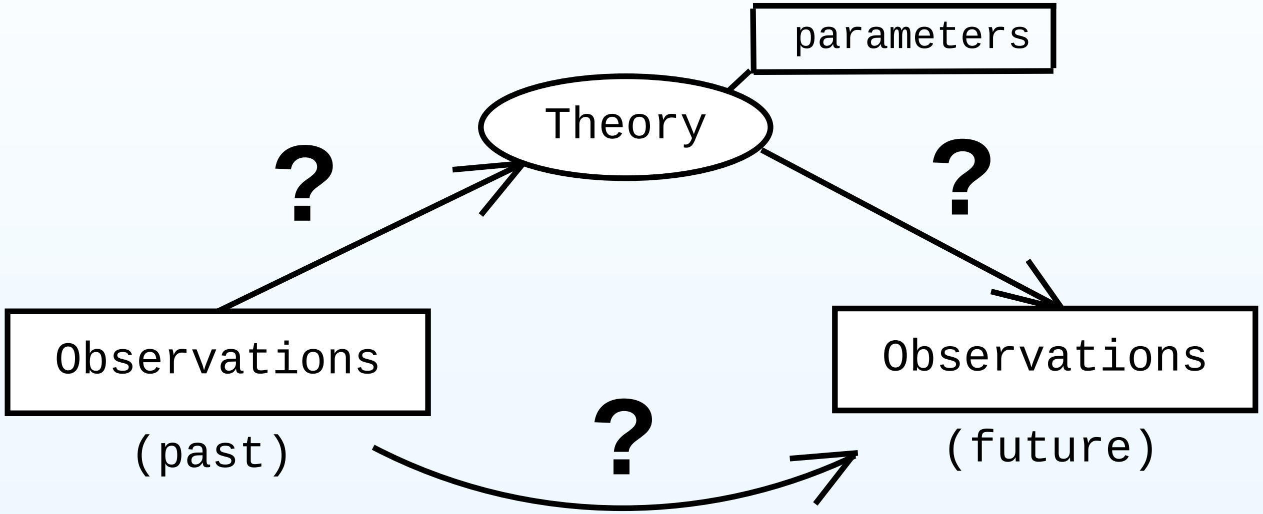
Outline

- Learning from data the probabilistic way
  - Causes $\longleftrightarrow$ Effects

    "The essential problem of the experimental method" *(Poincaré)*.
  - Graphical representation of probabilistic links
  - Learning about causes from their effects
- Parametric inference Vs unfolding
- From principles to real life... [the iteration 'dirty trick']
- The old code and its weak point
- Improvements:
  - use (conjugate) pdf's insteads of just 'estimates'
  - uncertainty evaluated by general rules of probability (instead of 'error propagation' formulae)

    $\Rightarrow$ integrals over the weighted possibilities $\rightarrow$ MC
- Some examples on toy models

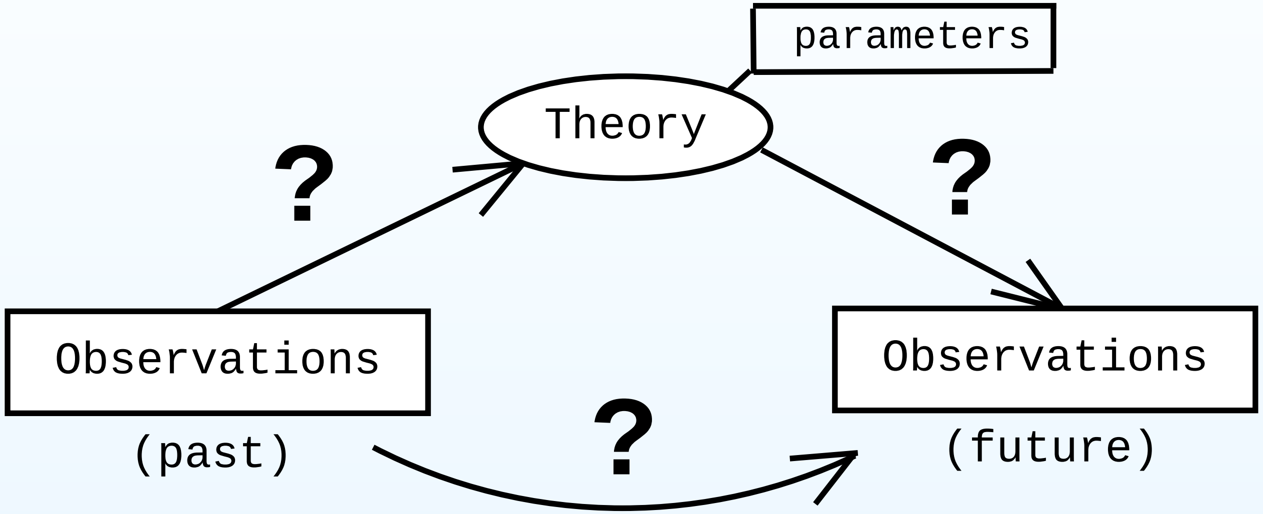
Learning from experience and source of uncertainty



Uncertainty:

Theory	— **?** ⟶		
Past observations	— **?** ⟶		
Theory	— **?** ⟶	Future observations	

Learning from experience and source of uncertainty



Uncertainty:

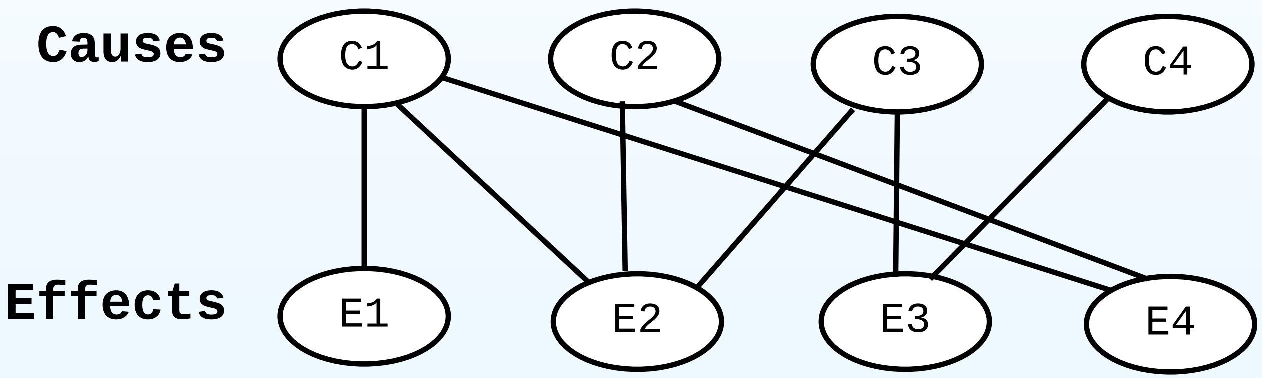
Theory	— ? ⟶		Future observations
Past observations	— ? ⟶		Theory
Theory	— ? ⟶		Future observations

$\Longrightarrow$ **Uncertainty about causal connections**

CAUSE $\Longleftrightarrow$ EFFECT

Causes → effects

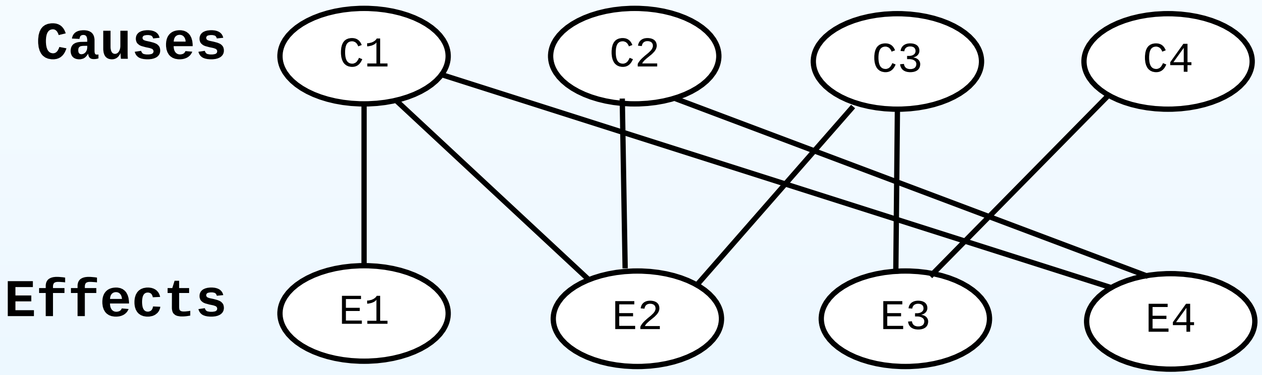
The same *apparent* cause might produce several,different effects



Given an observed effect, we are not sure about the exact cause that has produced it.

Causes ⟶ effects

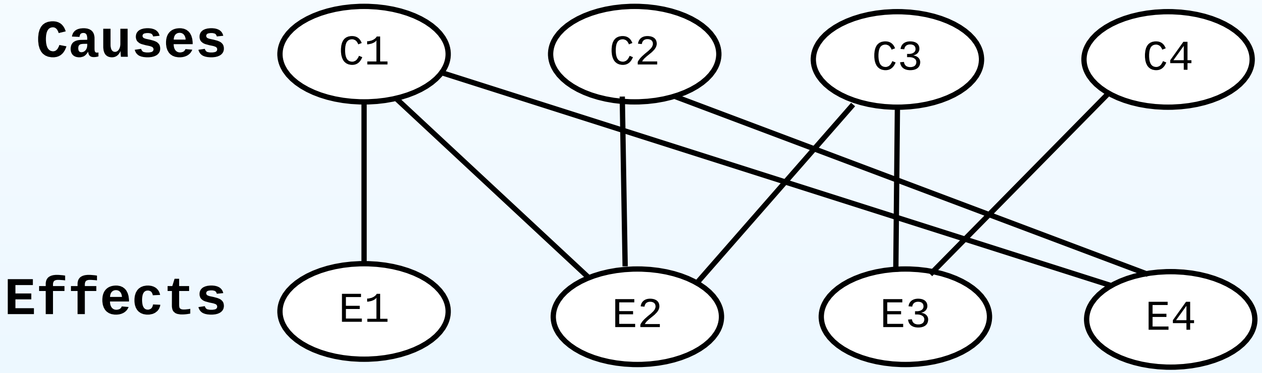
The same *apparent* cause might produce several,different effects



Given an observed effect, we are not sure about the exact cause that has produced it.

Causes $\longrightarrow$ effects

The same *apparent* cause might produce several,different effects



Given an observed effect, we are not sure about the exact cause that has produced it.

$$\mathbf{E_2} \Rightarrow \{C_1,\, C_2,\, C_3\}?$$

The essential problem of the experimental method

"Now, these problems are classified as *probability of causes*, and are most interesting of all their scientific applications. I play at *écarté* with a gentleman whom I know to be perfectly honest. What is the chance that he turns up the king? It is 1/8. This is a problem of the *probability of effects*.

The essential problem of the experimental method

"Now, these problems are classified as *probability of causes*, and are most interesting of all their scientific applications. I play at *écarté* with a gentleman whom I know to be perfectly honest. What is the chance that he turns up the king? It is 1/8. This is a problem of the *probability of effects*.

I play with a gentleman whom I do not know. He has dealt ten times, and he has turned the king up six times. What is the chance that he is a sharper? This is a problem in the probability of causes. It may be said that it is the essential problem of the experimental method."

(H. Poincaré – *Science and Hypothesis*)

Uncertainty and probability

We, as physicists, consider absolutely natural and meaningful statements of the following kind

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(170 \le m_{top}/\text{GeV} \le 180) \approx 70\%$
- $P(M_H < 200\,\text{GeV}) > P(M_H > 200\,\text{GeV})$

Uncertainty and probability

We, as physicists, consider absolutely natural and meaningful statements of the following kind

- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(170 \le m_{top}/\text{GeV} \le 180) \approx 70\%$
- $P(M_H < 200\,\text{GeV}) > P(M_H > 200\,\text{GeV})$

. . . although, such statements are considered
blaspheme to statistics gurus

Uncertainty and probability

We, as physicists, consider absolutely natural and meaningful statements of the following kind

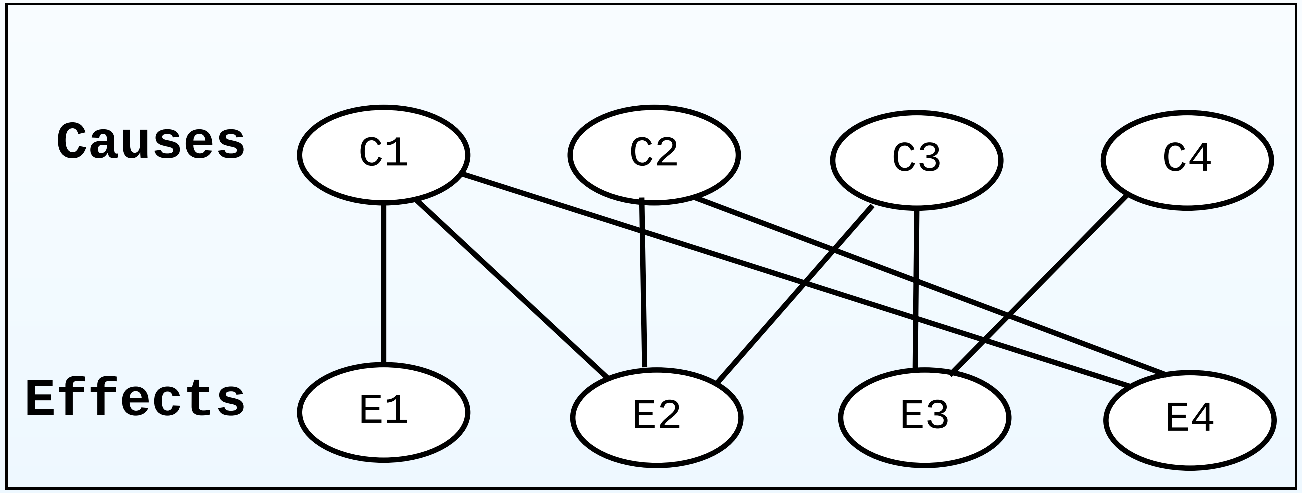
- $P(-10 < \epsilon'/\epsilon \times 10^4 < 50) >> P(\epsilon'/\epsilon \times 10^4 > 100)$
- $P(170 \le m_{top}/\text{GeV} \le 180) \approx 70\%$
- $P(M_H < 200\,\text{GeV}) > P(M_H > 200\,\text{GeV})$

. . . although, such statements are considered
blaspheme to statistics gurus

I stick to common sense (and physicists common sense) and assume that probabilities of causes, probabilities of of hypotheses, probabilities of the numerical values of physics quantities, etc. are sensible concepts that match the mind categories of human beings
(see D. Hume, C. Darwin + modern researches)

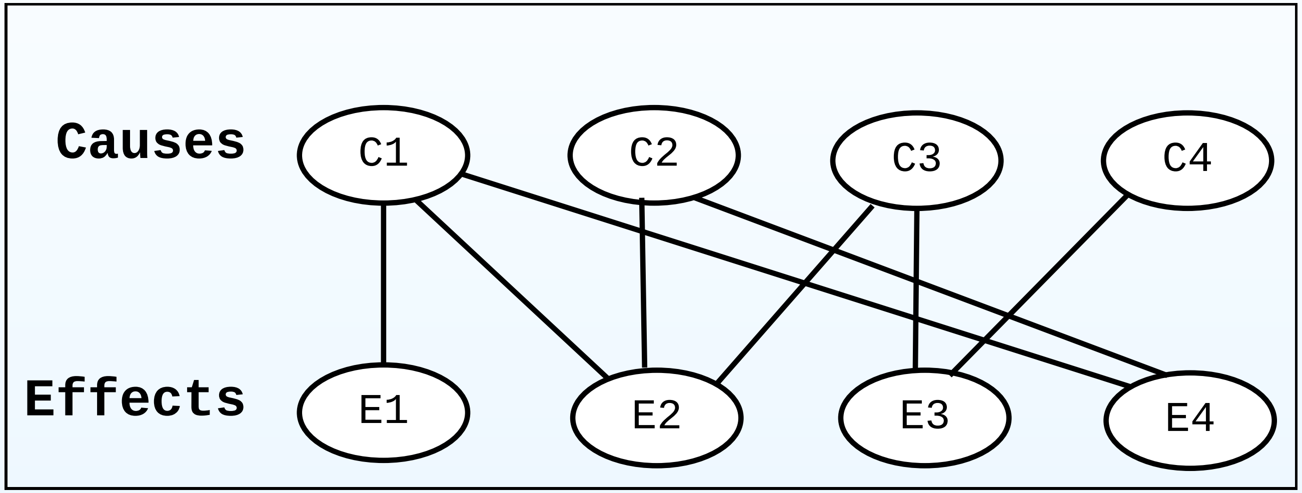
From causes to effects and back

Our original problem:



From causes to effects and back

Our original problem:

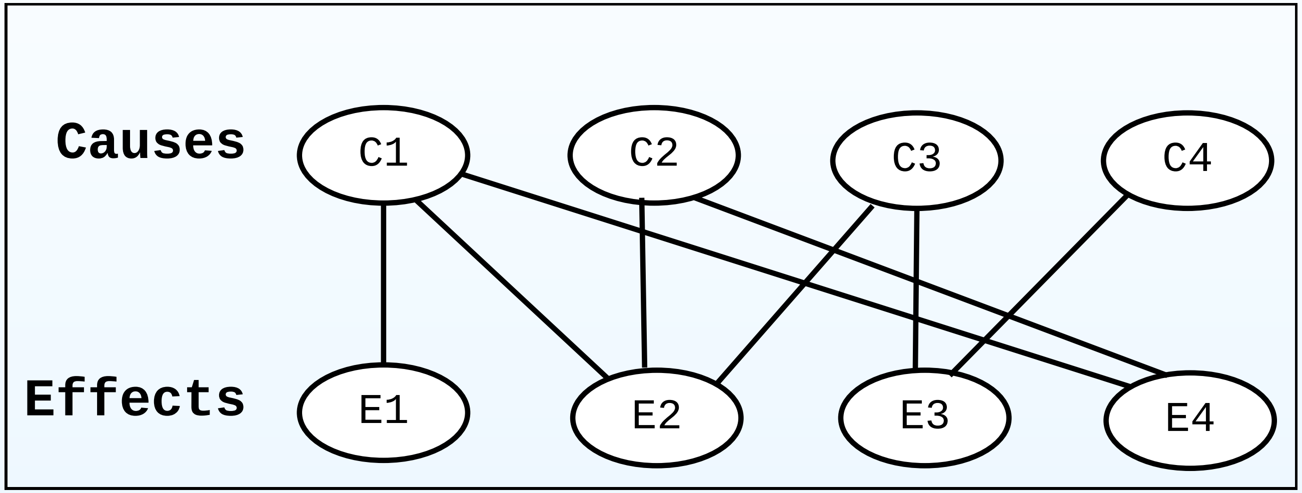


Our conditional view of probabilistic causation

$$\boxed{P(E_i \,|\, C_j)}$$

From causes to effects and back

Our original problem:



Our conditional view of probabilistic causation

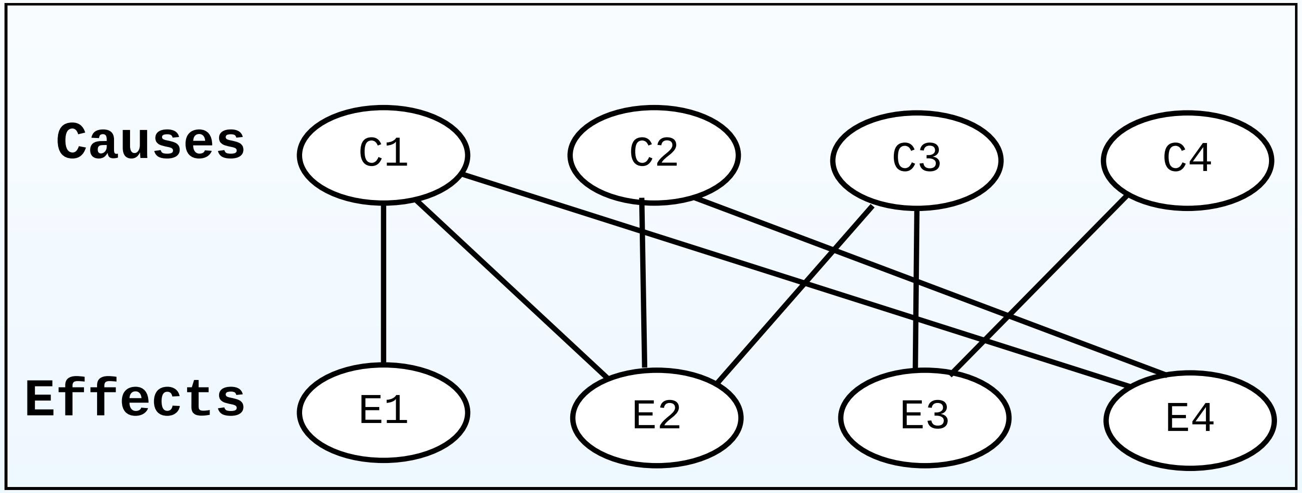
$$\boxed{P(E_i \,|\, C_j)}$$

Our conditional view of probabilistic inference

$$\boxed{P(C_j \,|\, E_i)}$$

From causes to effects and back

Our original problem:



Our conditional view of probabilistic causation

$$\boxed{P(E_i \,|\, C_j)}$$

Our conditional view of probabilistic inference

$$\boxed{P(C_j \,|\, E_i)}$$

The fourth basic rule of probability:

$$\boxed{P(C_j, E_i) = P(E_i \,|\, C_j)\, P(C_j) = P(C_j \,|\, E_i)\, P(E_i)}$$

Symmetric conditioning

Let us take basic rule 4, written in terms of hypotheses H_j and effects E_i, and rewrite it this way:

$$\frac{P(H_j \,|\, E_i)}{P(H_j)} = \frac{P(E_i \,|\, H_j)}{P(E_i)}$$

"The condition on E_i changes in percentage the probability of H_j as the probability of E_i is changed in percentage by the condition H_j."

Symmetric conditioning

Let us take basic rule 4, written in terms of hypotheses H_j and effects E_i, and rewrite it this way:

$$\frac{P(H_j \,|\, E_i)}{P(H_j)} = \frac{P(E_i \,|\, H_j)}{P(E_i)}$$

"The condition on E_i changes in percentage the probability of H_j as the probability of E_i is changed in percentage by the condition H_j."

It follows

$$P(H_j \,|\, E_i) = \frac{P(E_i \,|\, H_j)}{P(E_i)}\, P(H_j)$$

Symmetric conditioning

Let us take basic rule 4, written in terms of hypotheses H_j and effects E_i, and rewrite it this way:

$$\frac{P(H_j\,|\,E_i)}{P(H_j)} = \frac{P(E_i\,|\,H_j)}{P(E_i)}$$

"The condition on E_i changes in percentage the probability of H_j as the probability of E_i is changed in percentage by the condition H_j."

It follows

$$P(H_j\,|\,E_i) = \frac{P(E_i\,|\,H_j)}{P(E_i)}\,P(H_j)$$

Got 'after' Calculated 'before'

(where 'before' and 'after' refer to the knowledge that E_i is true.)

Symmetric conditioning

Let us take basic rule 4, written in terms of hypotheses H_j and effects E_i, and rewrite it this way:

$$\frac{P(H_j \,|\, E_i)}{P(H_j)} = \frac{P(E_i \,|\, H_j)}{P(E_i)}$$

"The condition on E_i changes in percentage the probability of H_j as the probability of E_i is changed in percentage by the condition H_j."

It follows

$$P(H_j \,|\, E_i) = \frac{P(E_i \,|\, H_j)}{P(E_i)}\, P(H_j)$$

"post illa observationes" *"ante illa observationes"*

(Gauss)

Symmetric conditioning

Let us take basic rule 4, written in terms of hypotheses H_j and effects E_i, and rewrite it this way:

$$\frac{P(H_j \,|\, E_i)}{P(H_j)} = \frac{P(E_i \,|\, H_j)}{P(E_i)}$$

"The condition on E_i changes in percentage the probability of H_j as the probability of E_i is changed in percentage by the condition H_j."

It follows

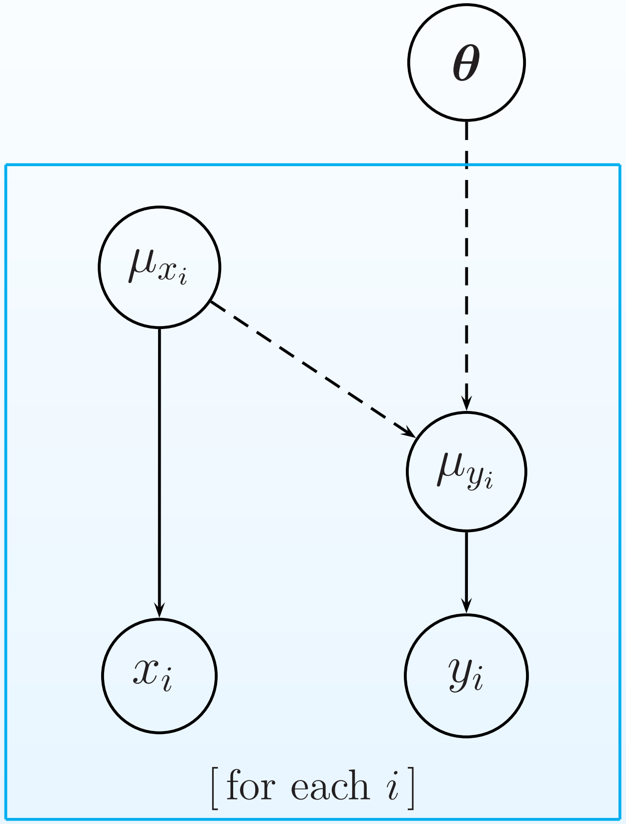
$$P(H_j \,|\, E_i) = \frac{P(E_i \,|\, H_j)}{P(E_i)}\, P(H_j)$$

"post illa observationes" *"ante illa observationes"*

(Gauss)

$\Rightarrow$ Bayes theorem

A different way to view fit issues



- Determistic link μ_x's to μ_y's
- Probabilistic links $\mu_x \longrightarrow x$, $\mu_y \longrightarrow y$

$\Rightarrow$ aim of fit: $\{\boldsymbol{x}, \boldsymbol{y}\} \ \longrightarrow \boldsymbol{\theta} \Rightarrow f(\boldsymbol{\theta} \,|\, \{\boldsymbol{x}, \boldsymbol{y}\})$

Parametric inference Vs unfolding

$f(\boldsymbol{\theta}\,|\,\{\boldsymbol{x},\boldsymbol{y}\})$:

Parametric inference Vs unfolding

$f(\boldsymbol{\theta}\,|\,\{\boldsymbol{x},\boldsymbol{y}\})$:

probabilistic parametric inference
$\Rightarrow$ it relies on the kind of functions parametrized by $\boldsymbol{\theta}$

$$\boldsymbol{\mu}_y = \boldsymbol{\mu}_y(\boldsymbol{\mu}_x;\boldsymbol{\theta})$$

Parametric inference Vs unfolding

$f(\boldsymbol{\theta}\,|\,\{\boldsymbol{x},\boldsymbol{y}\})$:

probabilistic parametric inference
$\Rightarrow$ it relies on the kind of functions parametrized by $\boldsymbol{\theta}$

$$\boldsymbol{\mu}_y = \boldsymbol{\mu}_y(\boldsymbol{\mu}_x;\boldsymbol{\theta})$$

$\Rightarrow$ data distilled into $\boldsymbol{\theta}$;

BUT sometimes we wish to interpret the data as little as possible

$\Rightarrow$ just public 'something equivalent' to an experimental distribution, with the bin contents fluctuating according to an underlying multinomial distribution, but having possibly got rid of physical and instrumental distortions, as well as of background.

Parametric inference Vs unfolding

$f(\boldsymbol{\theta}\,|\,\{\boldsymbol{x},\boldsymbol{y}\})$:

probabilistic parametric inference
$\Rightarrow$ it relies on the kind of functions parametrized by $\boldsymbol{\theta}$

$$\boldsymbol{\mu}_y = \boldsymbol{\mu}_y(\boldsymbol{\mu}_x;\boldsymbol{\theta})$$

$\Rightarrow$ data distilled into $\boldsymbol{\theta}$;

BUT sometimes we wish to interpret the data as little as possible

$\Rightarrow$ just public ‘something equivalent’ to an experimental distribution, with the bin contents fluctuating according to an underlying multinomial distribution, but having possibly got rid of physical and instrumental distortions, as well as of background.

$\Rightarrow$ Unfolding (deconvolution)

Why unfolding?

Te idea is to provide somethimg similar to an experimental spectrum, with a minimal interpretation by the experimentalist, a part from correcting from distortions due to physics and detector effects (including background).

(The alternative would be to give a parametrized description of the true spectrum – a fit)

Smearing matrix $\longrightarrow$ unfolding matrix

Invert smearing matrix?

Smearing matrix $\longrightarrow$ unfolding matrix

Invert smearing matrix?

In general is a bad idea:

not a rotational problem

but an inferential problem!

Smearing matrix $\longrightarrow$ unfolding matrix

Imagine $S = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$: $\longrightarrow U = S^{-1} = \begin{pmatrix} 1.33 & -0.33 \\ -0.33 & 1.33 \end{pmatrix}$

Let the true be $s_t = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$: $\longrightarrow s_m = S \cdot s_t = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$;

If we measure $s_m = \begin{pmatrix} 8 \\ 2 \end{pmatrix} \longrightarrow S^{-1} \cdot s_m = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$ $\surd$

Smearing matrix $\longrightarrow$ unfolding matrix

Imagine $S = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$: $\rightarrow U = S^{-1} = \begin{pmatrix} 1.33 & -0.33 \\ -0.33 & 1.33 \end{pmatrix}$

Let the true be $s_t = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$: $\rightarrow s_m = S \cdot s_t = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$;

If we measure $s_m = \begin{pmatrix} 8 \\ 2 \end{pmatrix} \rightarrow S^{-1} \cdot s_m = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$ $\surd$

BUT

if we had measured $\begin{pmatrix} 9 \\ 1 \end{pmatrix} \rightarrow S^{-1} \cdot s_m = \begin{pmatrix} 11.7 \\ -1.7 \end{pmatrix}$

if we had measured $\begin{pmatrix} 10 \\ 0 \end{pmatrix} \rightarrow S^{-1} \cdot s_m = \begin{pmatrix} 13.3 \\ -3.3 \end{pmatrix}$

Smearing matrix $\longrightarrow$ unfolding matrix

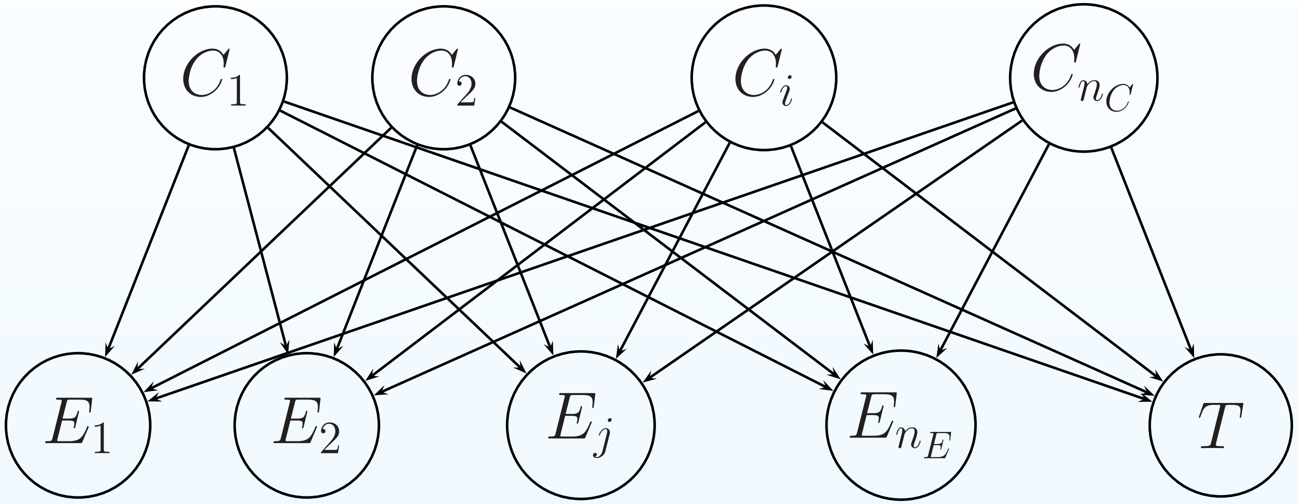
Imagine $S = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$: $\rightarrow U = S^{-1} = \begin{pmatrix} 1.33 & -0.33 \\ -0.33 & 1.33 \end{pmatrix}$

Let the true be $s_t = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$: $\rightarrow s_m = S \cdot s_t = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$;

If we measure $s_m = \begin{pmatrix} 8 \\ 2 \end{pmatrix} \rightarrow S^{-1} \cdot s_m = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$ $\surd$

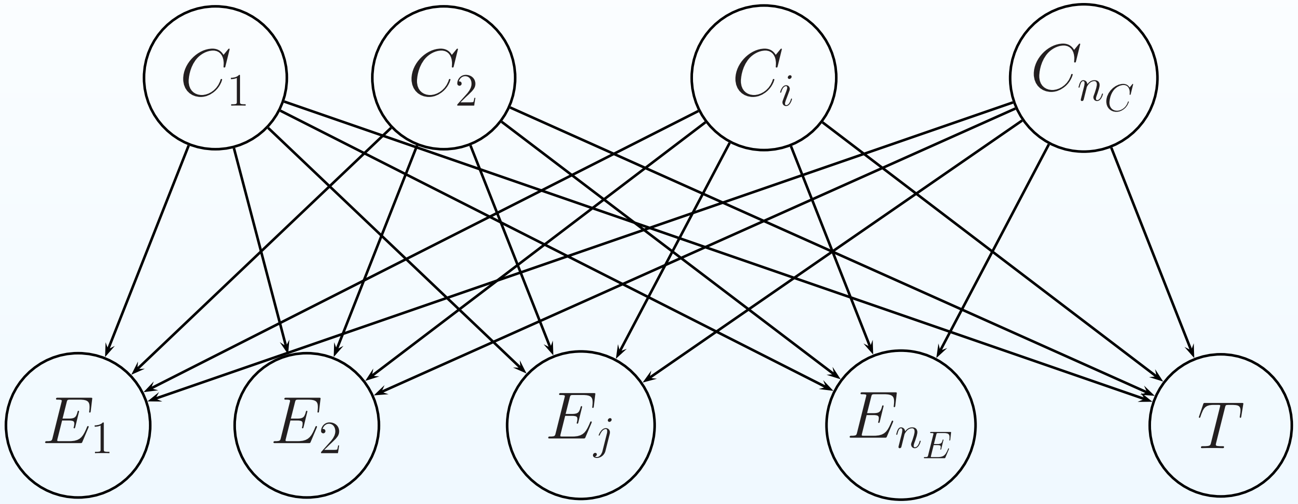
Indeed, matrix inversion is recognized to producing ‘crazy spectra’ and even negative values (unless such large numbers in bins such fluctuations around expectations are negligeable)

Discretized unfolding



(T: ‘trash’)

Discretized unfolding

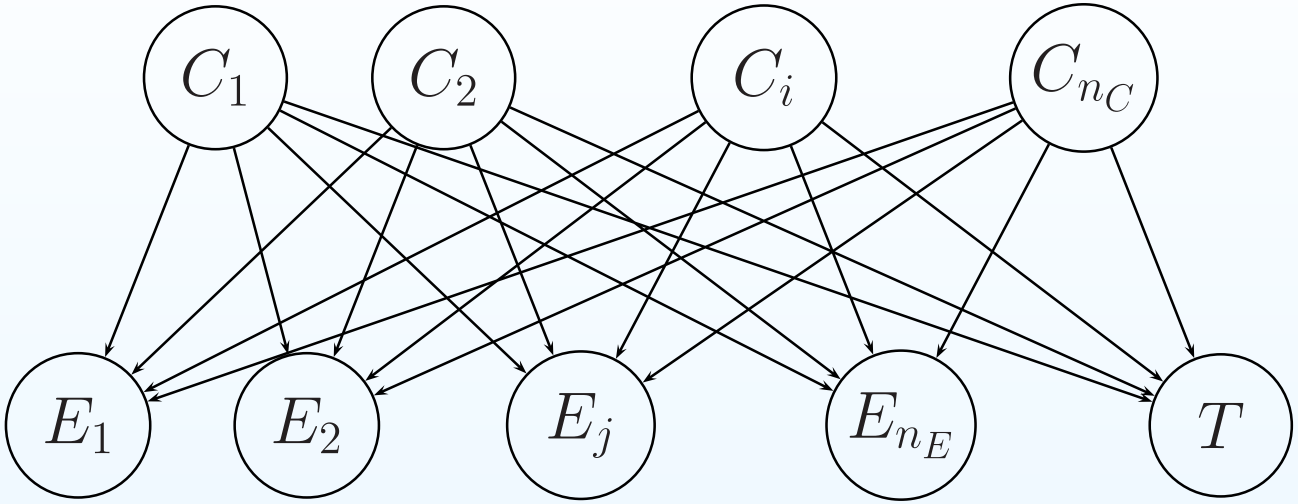


(T: 'trash')

$\boldsymbol{x}_C$ **:** *true spectrum* (nr of events in cause bins)

$\boldsymbol{x}_E$ **:** *observed spectrum* (nr of events in effect bins)

Discretized unfolding



(T: 'trash')

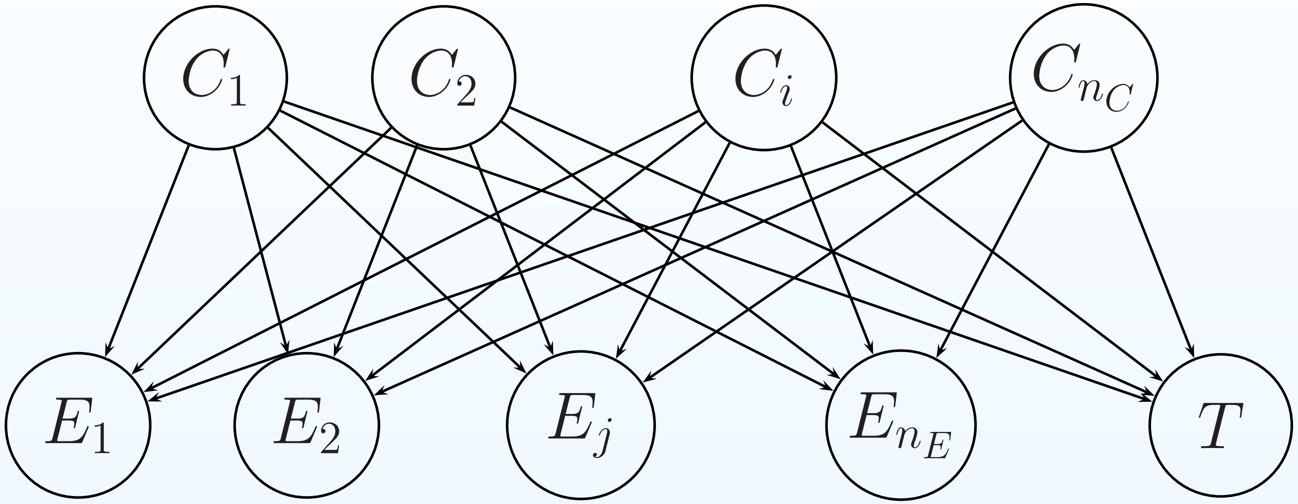
$\boldsymbol{x}_C$ **:** *true spectrum* (nr of events in cause bins)

$\boldsymbol{x}_E$ **:** *observed spectrum* (nr of events in effect bins)

Our aim:

- **not** to find *the* true spectrum
- but, more modestly, rank in beliefs all possible spectra that might have caused the observed one:
  $\Rightarrow P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I)$

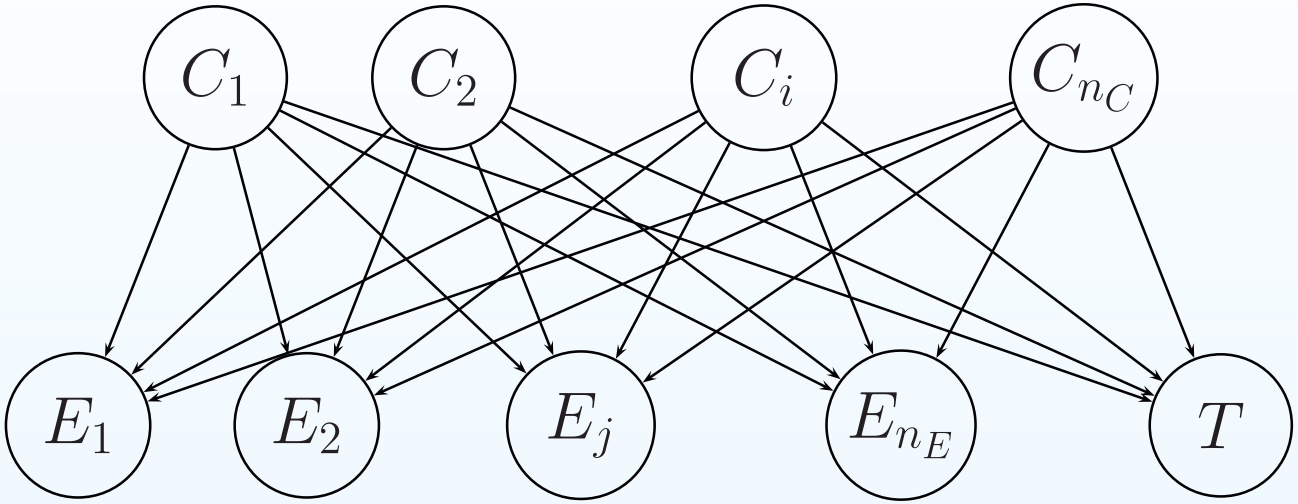
Discretized unfolding



(T: 'trash')

- $P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I)$ depends on the knowledge of *smearing matrix* Λ, with $\lambda_{ji} \equiv P(E_j \,|\, C_i,\, I)$.

Discretized unfolding

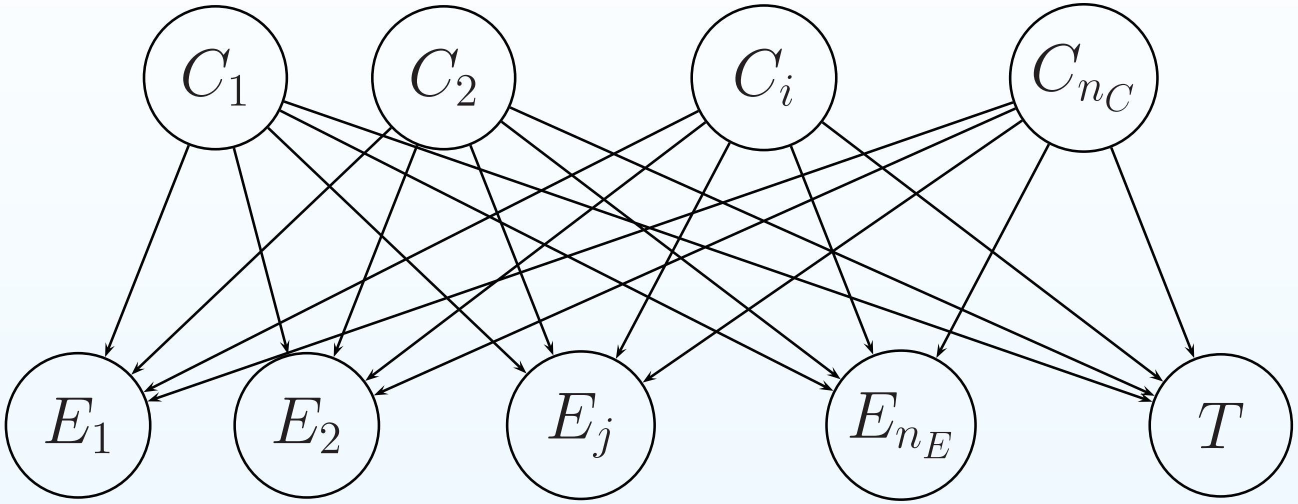


(T: 'trash')

- $P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I)$ depends on the knowledge of *smearing matrix* Λ, with $\lambda_{ji} \equiv P(E_j \,|\, C_i,\, I)$.
- but Λ is itself uncertain, because inferred from MC simulation:

$$\Rightarrow f(\Lambda \,|\, I)$$

Discretized unfolding



(T: 'trash')

- $P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I)$ depends on the knowledge of *smearing matrix* Λ,

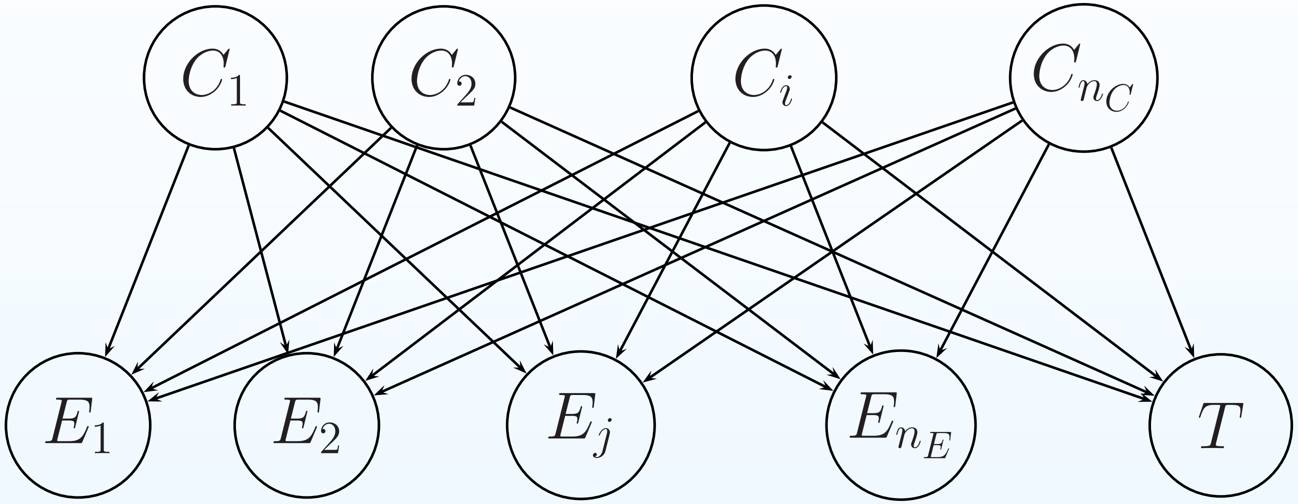
  with $\lambda_{ji} \equiv P(E_j \,|\, C_i,\, I)$.

- but Λ is itself uncertain, because inferred from MC simulation:

  $\Rightarrow f(\Lambda \,|\, I)$

- for each possible Λ we have a pdf of spectra:
  $\rightarrow P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda, I)$

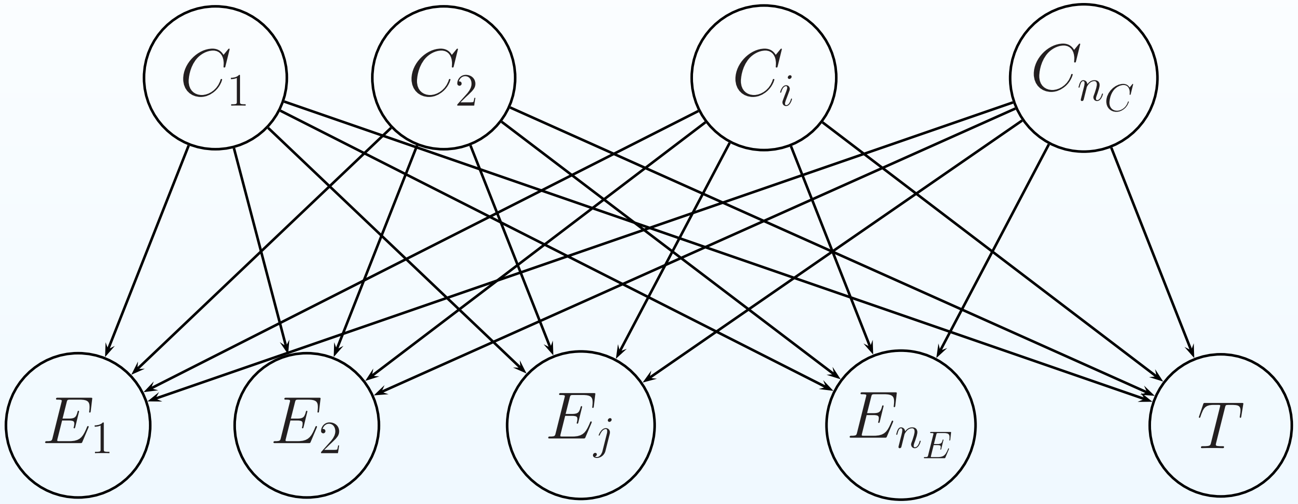
Discretized unfolding



(T: ‘trash’)

- $P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I)$ depends on the knowledge of *smearing matrix* Λ,
  with $\lambda_{ji} \equiv P(E_j \,|\, C_i,\, I)$.
- but Λ is itself uncertain, because inferred from MC simulation:
  $$\Rightarrow f(\Lambda \,|\, I)$$
- for each possible Λ we have a pdf of spectra:
  $\longrightarrow P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda, I)$
  $\Rightarrow P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I) = \int P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda,\, I)\, f(\Lambda \,|\, I)\, \mathrm{d}\Lambda$ [by MC!]

Discretized unfolding

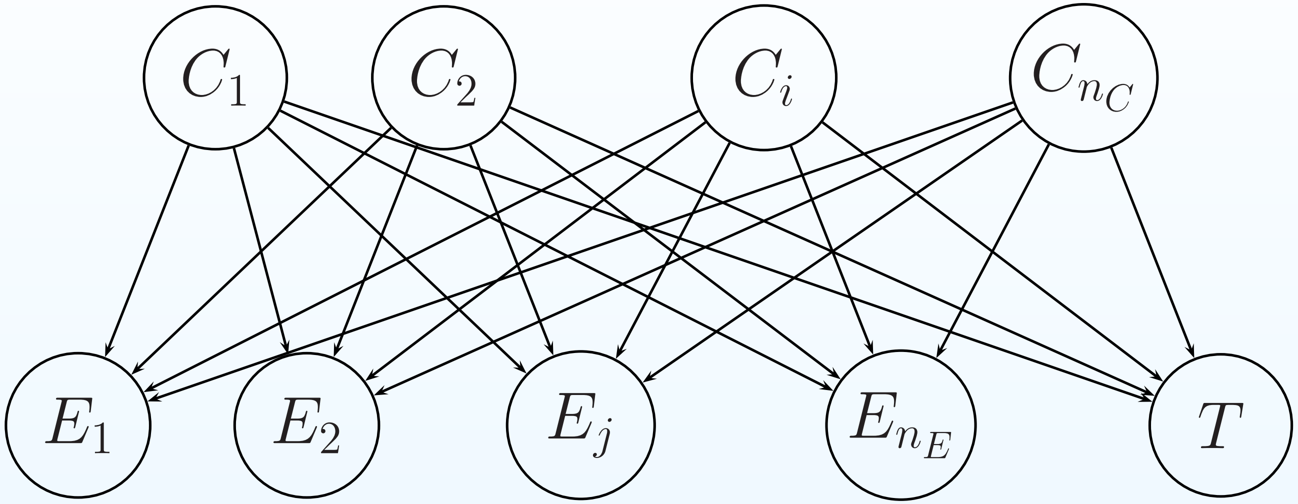


(T: 'trash')

- Bayes theorem:

$$P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \, \Lambda, \, I) \;\; \propto \;\; P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C, \, \Lambda, \, I) \cdot P(\boldsymbol{x}_C \,|\, I) \, .$$

Discretized unfolding



(T: 'trash')

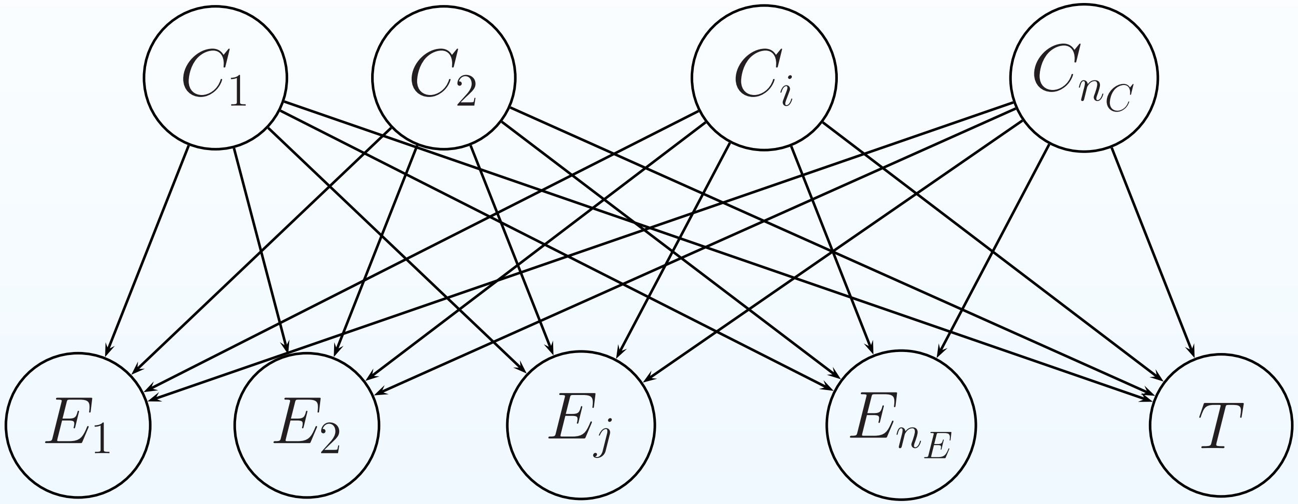
- Bayes theorem:

$$P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E,\, \Lambda,\, I) \;\propto\; P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I) \cdot P(\boldsymbol{x}_C \,|\, I)\,.$$

- Indifference w.r.t. all possible spectra

$$P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E,\, \Lambda,\, I) \;\propto\; P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I)$$

$P(\boldsymbol{x}_E \,|\, x_{C_i},\, \Lambda,\, I)$



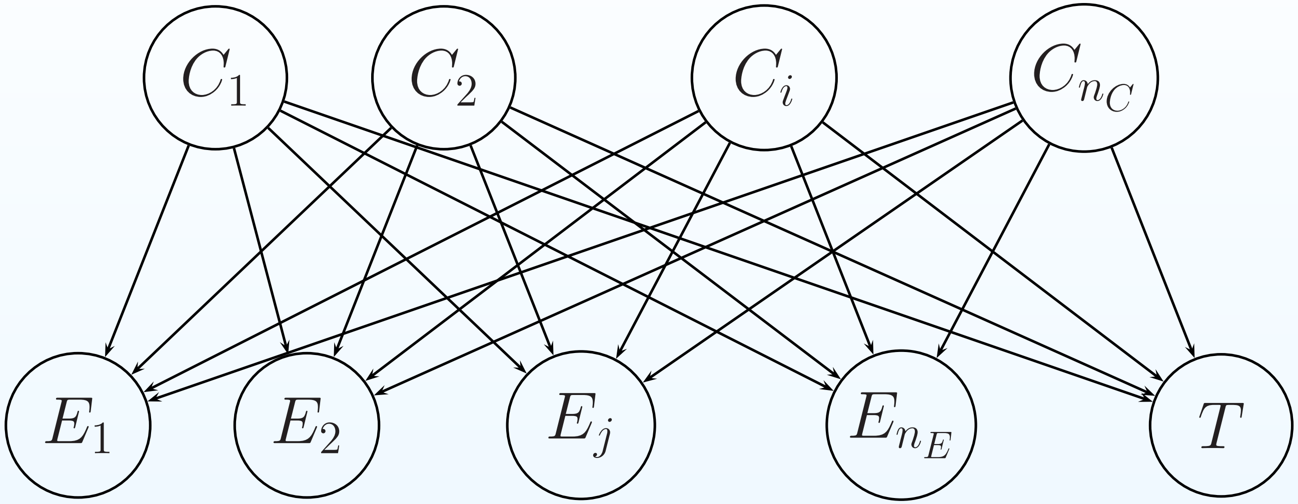
Given a certain number of events in a cause-bin $x(C_i)$, the number of events in the effect-bins, included the 'trash' one, is described by a multinomial distribution:

$$\boldsymbol{x}_E\big|_{x(C_i)} \;\sim\; \text{Mult}[x(C_i), \boldsymbol{\lambda}_i]\,,$$

with

$$\begin{aligned}\boldsymbol{\lambda}_i &= \{\lambda_{1,i},\, \lambda_{2,i}, \ldots,\, \lambda_{n_E+1,i}\} \\ &= \{P(E_1 \,|\, C_i,\, I),\, P(E_2 \,|\, C_i,\, I), \ldots,\, P(E_{n_E+1,i} \,|\, C_i,\, I)\}\end{aligned}$$

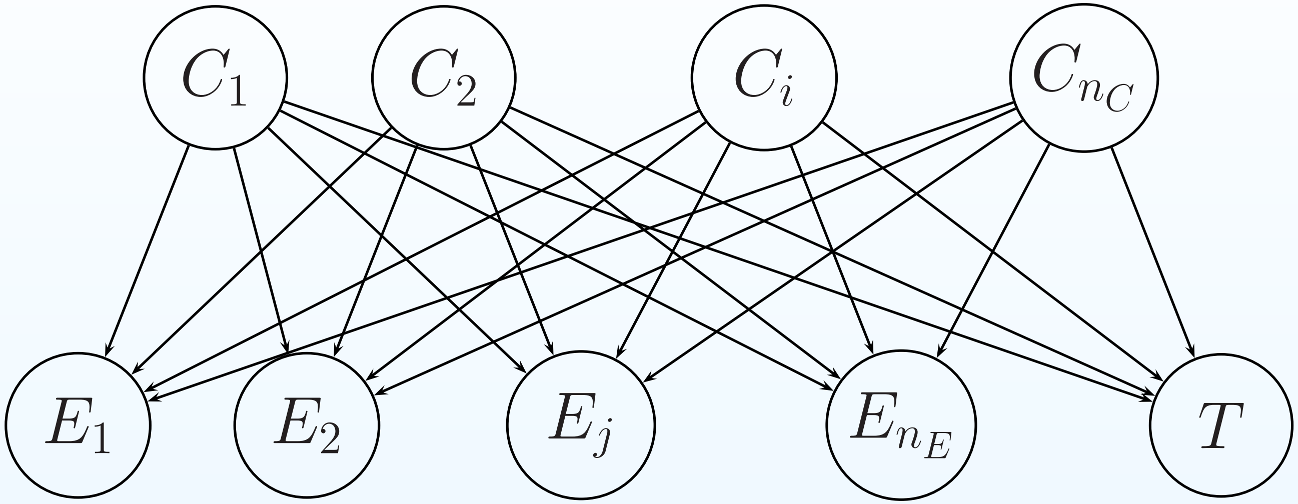
$P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I)$



$\boldsymbol{x}_E\big|_{x(C_i)}$ multinomial random vector,

$\Rightarrow \boldsymbol{x}_E\big|_{\boldsymbol{x}(C)}$ sum of several multinomials.

$P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I)$



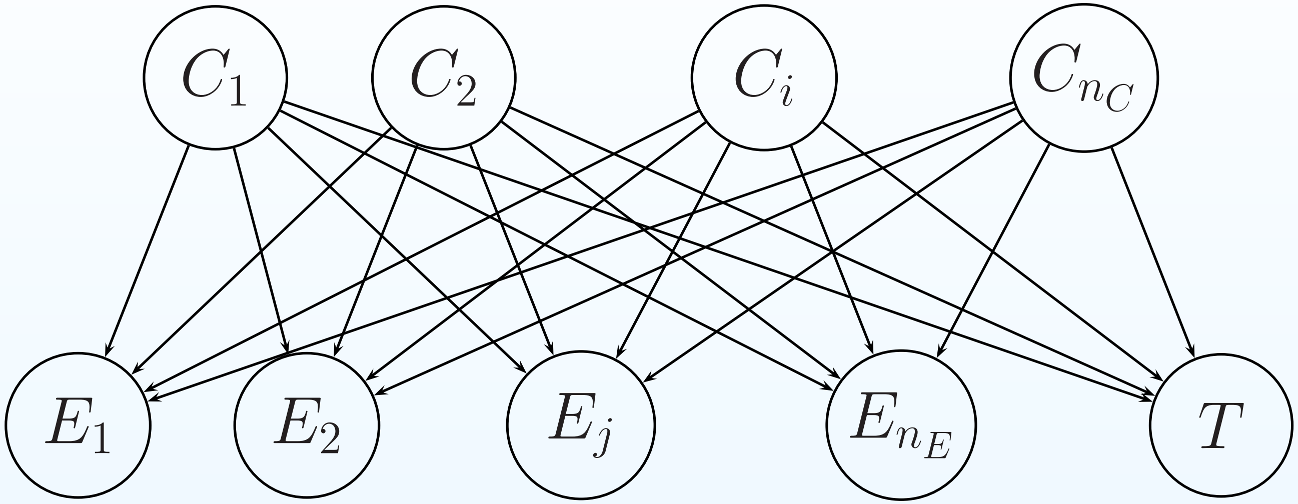
$\boldsymbol{x}_E|_{x(C_i)}$ multinomial random vector,

$\Rightarrow \boldsymbol{x}_E|_{\boldsymbol{x}(C)}$ sum of several multinomials.

BUT

no 'easy' expression for $P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C, \Lambda, I)$

$P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I)$



$\boldsymbol{x}_E|_{x(C_i)}$ multinomial random vector,

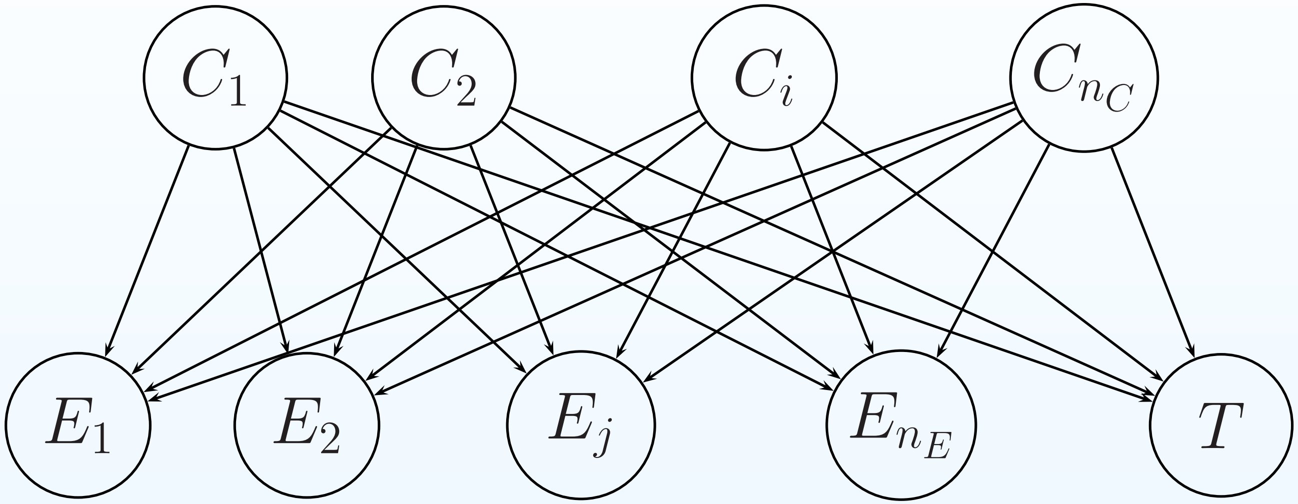
$\Rightarrow$ $\boldsymbol{x}_E|_{\boldsymbol{x}(C)}$ sum of several multinomials.

BUT

no 'easy' expression for $P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C, \Lambda, I)$

$\Rightarrow$ STUCK!

$P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I)$



$\boldsymbol{x}_E\big|_{x(C_i)}$ multinomial random vector,

$\Rightarrow$ $\boldsymbol{x}_E\big|_{\boldsymbol{x}(C)}$ sum of several multinomials.

BUT

no 'easy' expression for $P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C, \Lambda, I)$

$\Rightarrow$ STUCK!

$\Rightarrow$ Change strategy

The rescue trick

Instead of using the original probability inversion (applied directly) to spectra

$$P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E,\, \Lambda,\, I) \quad \propto \quad P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I) \cdot P(\boldsymbol{x}_C \,|\, I)\,,$$

we restart from

$$P(C_i \,|E_j,\, I) \quad \propto \quad P(E_j \,|\, C_i,\, I) \cdot P(C_i \,|\, I).$$

The rescue trick

Instead of using the original probability inversion (applied directly) to spectra

$$P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E,\, \Lambda,\, I) \;\;\propto\;\; P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I) \cdot P(\boldsymbol{x}_C \,|\, I)\,,$$

we restart from

$$P(C_i \,|E_j,\, I) \;\;\propto\;\; P(E_j \,|\, C_i,\, I) \cdot P(C_i \,|\, I).$$

Consequences:

1. the sharing of observed events among the cause bins needs to be performed 'by hand';

The rescue trick

Instead of using the original probability inversion (applied directly) to spectra

$$P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda, I) \;\;\propto\;\; P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C, \Lambda, I) \cdot P(\boldsymbol{x}_C \,|\, I)\,,$$

we restart from

$$P(C_i \,|E_j, I) \;\;\propto\;\; P(E_j \,|\, C_i, I) \cdot P(C_i \,|\, I).$$

Consequences:

1. the sharing of observed events among the cause bins needs to be performed 'by hand';
2. a uniform prior $P(C_i \,|\, I) = k$ does not mean indifference over all possible spectra.

   ⇒ $P(C_i \,|\, I) = k$ is a well precise spectrum (in most cases far from the physical one)

   ⇒ VERY STRONG prior that biases the result!

The rescue trick

Instead of using the original probability inversion (applied directly) to spectra

$$P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E,\, \Lambda,\, I) \;\; \propto \;\; P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I) \cdot P(\boldsymbol{x}_C \,|\, I)\,,$$

we restart from

$$P(C_i \,|E_j,\, I) \;\; \propto \;\; P(E_j \,|\, C_i,\, I) \cdot P(C_i \,|\, I).$$

Consequences:

1. the sharing of observed events among the cause bins needs to be performed 'by hand';
2. a uniform prior $P(C_i \,|\, I) = k$ does not mean indifference over all possible spectra.

   $\Rightarrow$ $P(C_i \,|\, I) = k$ is a well precise spectrum (in most cases far from the physical one)

   $\Rightarrow$ VERY STRONG prior that biases the result! $\longrightarrow$ iterations

Old algorithm

1. [*] λ_{ij} estimated by MC simulation as

$$\lambda_{ji} \approx x(E_j)^{MC} / x(C_i)^{MC} \; ;$$

Old algorithm

1. [*] λ_{ij} estimated by MC simulation as

$$\lambda_{ji} \approx x(E_j)^{MC}/x(C_i)^{MC}\,;$$

2. $P(C_i\,|E_j,\,I)$ from Bayes theorem; [$\theta_{ij} \equiv P(C_i\,|E_j,\,I)$]

$$P(C_i\,|E_j,\,I) \;=\; \frac{P(E_j\,|\,C_i,\,I)\cdot P(C_i\,|\,I)}{\sum_i P(E_j\,|\,C_i,\,I)\cdot P(C_i\,|\,I)}\,,$$

or

$$\theta_{ij} \;=\; \frac{\lambda_{ji}\cdot P(C_i\,|\,I)}{\sum_i \lambda_{ji}\cdot P(C_i\,|\,I)}\,,$$

Old algorithm

1. [$\ast$] λ_{ij} estimated by MC simulation as

$$\lambda_{ji} \approx x(E_j)^{MC} / x(C_i)^{MC} \,;$$

2. $P(C_i \,|\, E_j,\, I)$ from Bayes theorem; [$\theta_{ij} \equiv P(C_i \,|\, E_j,\, I)$]
3. [$\ast$] Assignement of events to cause bins:

$$\begin{aligned} x(C_i)|_{x(E_j)} &\approx P(C_i \,|\, E_j,\, I) \cdot x(E_j) \\ x(C_i)|_{\boldsymbol{x}_E} &\approx \sum_{j=1}^{n_E} P(C_i \,|\, E_j,\, I) \cdot x(E_j) \\ x(C_i) &\approx \frac{1}{\epsilon_i}\, x(C_i)|_{\boldsymbol{x}_E} \,, \end{aligned}$$

with $\epsilon_i = \sum_{j=1}^{n_E} P(E_j \,|\, C_i,\, I)$

Old algorithm

1. [∗] λ_{ij} estimated by MC simulation as

$$\lambda_{ji} \approx x(E_j)^{MC}/x(C_i)^{MC}\,;$$

2. $P(C_i\,|E_j,\,I)$ from Bayes theorem; $[\theta_{ij} \equiv P(C_i\,|E_j,\,I)]$

3. [∗] Assignement of events to cause bins:

$$\begin{aligned} x(C_i)|_{x(E_j)} &\approx P(C_i\,|E_j,\,I)\cdot x(E_j) \\ x(C_i)|_{\boldsymbol{x}_E} &\approx \sum_{j=1}^{n_E} P(C_i\,|E_j,\,I)\cdot x(E_j) \\ x(C_i) &\approx \frac{1}{\epsilon_i}\,x(C_i)|_{\boldsymbol{x}_E}\,, \end{aligned}$$

with $\epsilon_i = \sum_{j=1}^{n_E} P(E_j\,|\,C_i,\,I)$

4. [∗] Uncertainty by ‘standard error propagation’

Improvements

1. $\boldsymbol{\lambda}_i$: having each element λ_{ji} the meaning of "p_j" of a Multinomial distribution, their distribution can easily (and conveniently and realistically) modelled by a Dirichlet:

$$\boldsymbol{\lambda}_i \quad \sim \quad \text{Dir}\left[\boldsymbol{\alpha}_{prior} + \boldsymbol{x}_E^{MC}\Big|_{x(C_i)^{MC}}\right],$$

(The Dirichlet is the prior conjugate of the Multinomial)

Improvements

1. $\boldsymbol{\lambda}_i$:
$$\boldsymbol{\lambda}_i \sim \text{Dir}[\boldsymbol{\alpha}_{prior} + \boldsymbol{x}_E^{MC}\big|_{x(C_i)^{MC}}]\,,$$

2. uncertainty on $\boldsymbol{\lambda}_i$:
taken into account by sampling $\Rightarrow$ equivalent to integration

$$\Rightarrow P(\boldsymbol{x}_C\,|\,\boldsymbol{x}_E, I) \;=\; \int P(\boldsymbol{x}_C\,|\,\boldsymbol{x}_E, \Lambda,\, I)\, f(\Lambda\,|\,I)\, \mathrm{d}\Lambda$$

Improvements

1. $\boldsymbol{\lambda}_i$:

$$\boldsymbol{\lambda}_i \sim \text{Dir}[\boldsymbol{\alpha}_{prior} + \boldsymbol{x}_E^{MC}\big|_{x(C_i)^{MC}}],$$

2. uncertainty on $\boldsymbol{\lambda}_i$:
   taken into account by sampling

3. sharing $x_{E_j} \rightarrow \boldsymbol{x}_C$: done by a Multinomial:

$$\boldsymbol{x}_C\big|_{x(E_j)} \sim \text{Mult}[x(E_j), \boldsymbol{\theta}_j],$$

Improvements

1. $\boldsymbol{\lambda}_i$:

$$\boldsymbol{\lambda}_i \sim \text{Dir}[\boldsymbol{\alpha}_{prior} + \boldsymbol{x}_E^{MC}\big|_{x(C_i)^{MC}}]\,,$$

2. uncertainty on $\boldsymbol{\lambda}_i$:
   taken into account by sampling
3. sharing $x_{E_j} \rightarrow \boldsymbol{x}_C$: done by a Multinomial:

$$\boldsymbol{x}_C\big|_{x(E_j)} \sim \text{Mult}[x(E_j),\, \boldsymbol{\theta}_j]\,,$$

4. $x(E_j) \rightarrow \mu_j$: what needs to be shared is not the observed number $x(E_j)$, but rather the estimated true value μ_j: remember $x(E_j) \sim \text{Poisson}[\mu_j]$

$$\mu_j \sim \text{Gamma}[c_j + x(E_j),\, r_j + 1]\,,$$

(Gamma is prior conjugate of Poisson)

Improvements

1. $\boldsymbol{\lambda}_i$:

$$\boldsymbol{\lambda}_i \sim \text{Dir}\left[\boldsymbol{\alpha}_{prior} + \boldsymbol{x}_E^{MC}\Big|_{x(C_i)^{MC}}\right],$$

2. uncertainty on $\boldsymbol{\lambda}_i$:
   taken into account by sampling

3. sharing $x_{E_j} \to \boldsymbol{x}_C$: done by a Multinomial:

$$\boldsymbol{x}_C\Big|_{x(E_j)} \sim \text{Mult}\left[x(E_j), \boldsymbol{\theta}_j\right],$$

4. $x(E_j) \to \mu_j$:

$$\mu_j \sim \text{Gamma}\left[c_j + x(E_j), r_j + 1\right],$$

   BUT μ_i is real, while the the number of event parameter of a multinomial must be integer $\Rightarrow$ solved with interpolation

5. uncertainty on μ_i: taken into account by sampling

Iteration and (intermediate) smoothing

instead of using a flat prior over the possible spectra
we are using a particular (flat) spectrum as prior

Iteration and (intermediate) smoothing

instead of using a flat prior over the possible spectra
we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $\boldsymbol{x}_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

Iteration and (intermediate) smoothing

instead of using a flat prior over the possible spectra
we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $x_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

$\Rightarrow$ problem worked around by ITERATIONS

$\Rightarrow$ posterior becomes prior of next iteration

Iteration and (intermediate) smoothing

> instead of using a flat prior over the possible spectra
> we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $x_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

$\Rightarrow$ problem worked around by ITERATIONS

$\Rightarrow$ posterior becomes prior of next iteration

$\Rightarrow$ **Usque tandem?**

Iteration and (intermediate) smoothing

instead of using a flat prior over the possible spectra
we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $x_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

$\Rightarrow$ problem worked around by ITERATIONS

$\Rightarrow$ posterior becomes prior of next iteration

$\Rightarrow$ **Usque tandem?**

- Empirical approach (with help of simulation):
  - 'True spectrum' recovered in a couple of steps
  - Then the solution starts to diverge towards a wildy oscillating spectrum (any unavoidable fluctuation is believed more and more. . .)
    $\Rightarrow$ find empirically an optimum

Iteration and (intermediate) smoothing

instead of using a flat prior over the possible spectra we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $x_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

$\Rightarrow$ problem worked around by ITERATIONS

$\Rightarrow$ posterior becomes prior of next iteration

$\Rightarrow$ Usque tandem?

- **regularization** (a subject by itself)
  my preferred approach
  - regularize the posterior before using as next prior

Iteration and (intermediate) smoothing

> instead of using a flat prior over the possible spectra
> we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $x_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

$\Rightarrow$ problem worked around by ITERATIONS

$\Rightarrow$ posterior becomes prior of next iteration

$\Rightarrow$ **Usque tandem?**

- **regularization** (a subject by itself)
  my preferred approach
  - regularize the posterior before using as next prior
  - intermediate smoothing $\Rightarrow$ we belief physics is 'smooth'
  - . . . but 'irregularities' of the data are not washed out

($\Rightarrow$ unfolding Vs parametric inference)

Iteration and (intermediate) smoothing

instead of using a flat prior over the possible spectra
we are using a particular (flat) spectrum as prior

$\Rightarrow$ the posterior [i.e. the ensemble of $x_C^{(t)}$ obtained by sampling] is affected by this quite strong assumption, that seldom holds in real cases.

$\Rightarrow$ problem worked around by ITERATIONS

$\Rightarrow$ posterior becomes prior of next iteration

$\Rightarrow$ Usque tandem?

- **regularization** (a subject by itself)
  my preferred approach
  - regularize the posterior before using as next prior

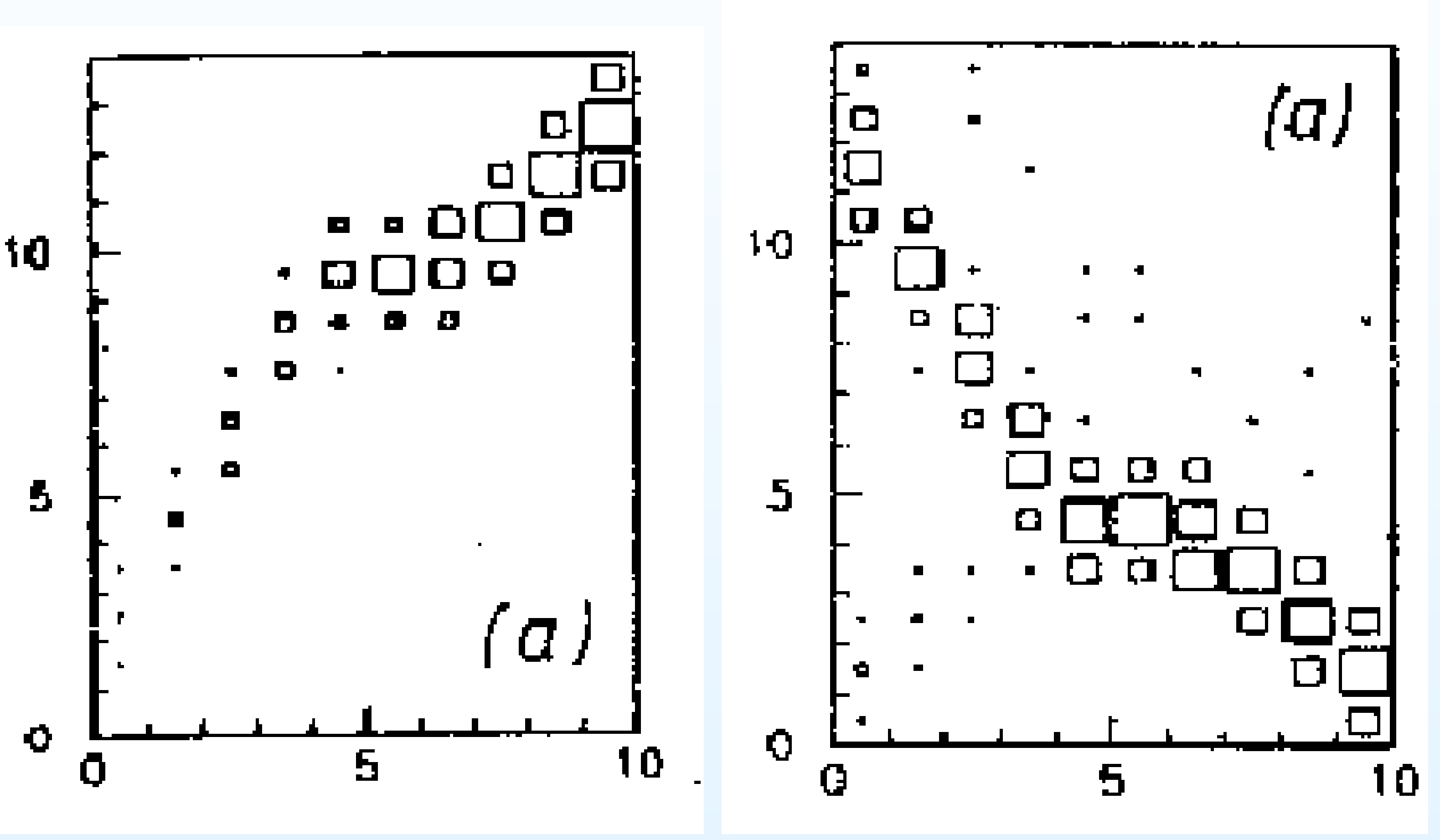
  $\Rightarrow$ Good compromize and good results

  $\Rightarrow$ Very 'Bayesian'

  $\Rightarrow$ No oscillations for $n_{steps} \rightarrow \infty$

Examples

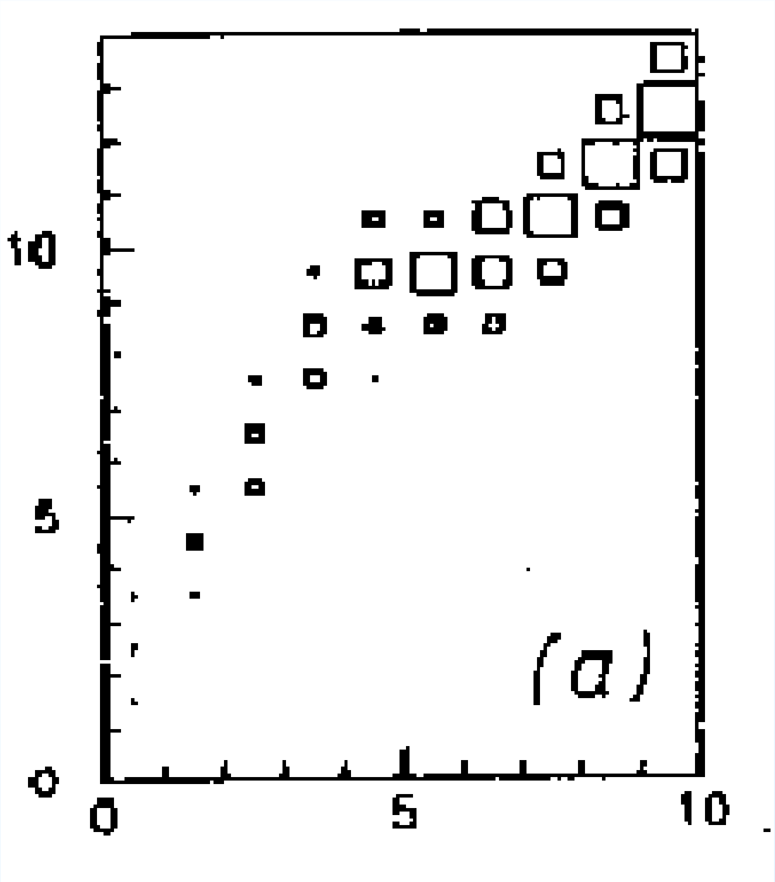
smearing matrix (from 1995 NIM paper)

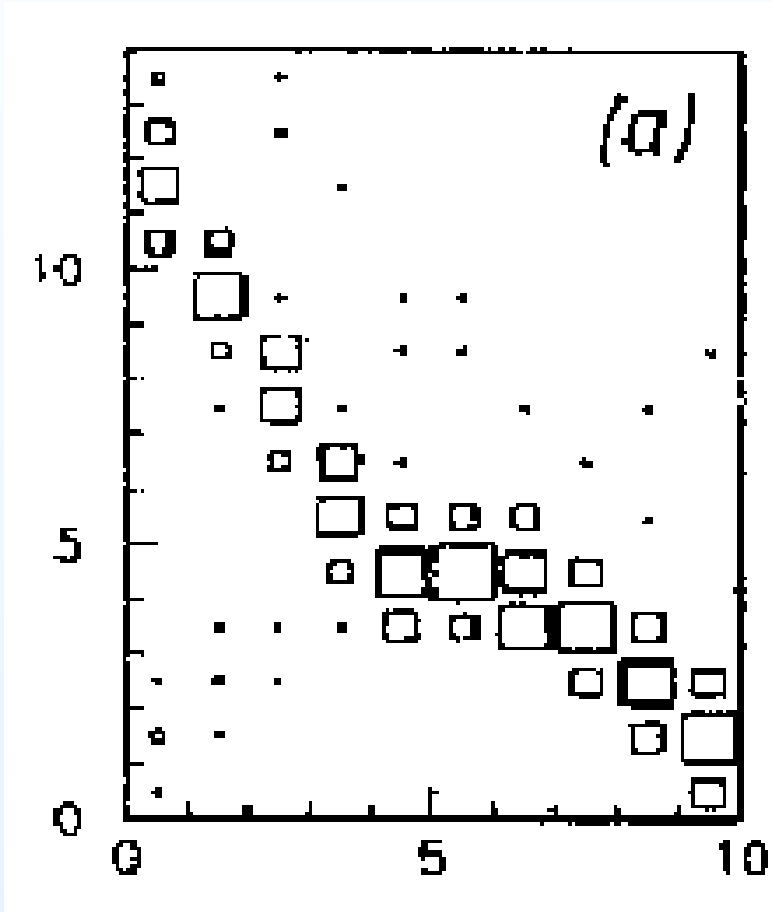


quite bad! (real cases are usually more gentle)

Examples

smearing matrix (from 1995 NIM paper)





quite bad! (real cases are usually more gentle)

$\Rightarrow$ watch DEMO

Conclusions

- left to users
  1. “non chiedere all’oste com’è il vino”. . .
  2. if I knew how (and was able) to do it better,
     I had already done it. . .

Conclusions

- left to users
  1. “non chiedere all’oste com’è il vino”. . .
  2. if I knew how (and was able) to do it better,
     I had already done it. . .
- still quite used because of simplicity of reasoning and code

Conclusions

- left to users
  1. “non chiedere all’oste com’è il vino”. . .
  2. if I knew how (and was able) to do it better,
     I had already done it. . .
- still quite used because of simplicity of reasoning and code
- new version improves
  - evaluation of uncertainties
  - handling of small numbers

Conclusions

- left to users
  1. “non chiedere all’oste com’è il vino”. . .
  2. if I knew how (and was able) to do it better,
     I had already done it. . .
- still quite used because of simplicity of reasoning and code
- new version improves
  - evaluation of uncertainties
  - handling of small numbers

$\longrightarrow$ Some notes follow $\Longrightarrow$

Notes added

1. “iterative” put within parentheses in title
   (motivated by Zech’ classification of methods)
   (a) the spirit of the method is Bayesian
   (b) the iteration issue is secondary

Notes added

1. “iterative” put within parentheses in title
2. An interesting book:
   (thanks to Blobel)
   - J. Kaipio and E. Somersalo
     Statistical and Computational Inverse Problems
     Springer, 2004

Notes added

1. "iterative" put within parentheses in title
2. An interesting book:
3. Uncertainty due the possible choice among several smearing models, Λ_1, Λ_2, etc.
   (triggered by Marisa Sandhoff's talk)
   - the $\boldsymbol{\theta}_i$ sampling can be done at random form either matrix,
     with weights depending on our beliefs in the different unfolding models
     (obviously not yet implemented in the R code, and I am not sure I will do it, but it can be implemented in C/C++ versions)

Notes added

1. "iterative" put within parentheses in title
2. An interesting book:
3. Uncertainty due the possible choice among several smearing models, Λ_1, Λ_2, etc.

Notes added

1. “iterative” put within parentheses in title
2. An interesting book:
3. Uncertainty due the possible choice among several smearing models, Λ_1, Λ_2, etc.
4. Extending an “anonymous” citation (Blobel’s talk)

   “...it gives the best results (in terms of its ability to reproduce the true distribution) if one make a realistic guess about the distribution that the true values follow...
   but, in case of total ignorance, satisfactory results are obtained even starting from a uniform distribution;”

   GdA, NIM A362 (1995) 487

Notes added

1. "iterative" put within parentheses in title
2. An interesting book:
3. Uncertainty due the possible choice among several smearing models, Λ_1, Λ_2, etc.
4. Extending an "anonymous" citation (Blobel's talk)

   "...it gives the best results (in terms of its ability to reproduce the true distribution) if one make a realistic guess about the distribution that the true values follow...
   but, in case of total ignorance, satisfactory results are obtained even starting from a uniform distribution;"

   GdA, NIM A362 (1995) 487

$\Rightarrow$ just a honest statement: what is wrong with it?

Notes added

1. "iterative" put within parentheses in title
2. An interesting book:
3. Uncertainty due the possible choice among several smearing models, Λ_1, Λ_2, etc.
4. Extending an "anonymous" citation

Buon divertimento!