From probabilistic inference to ‘Bayesian’ unfolding

(passing through a toy model)

Giulio D’Agostini

University and INFN Section of “Roma1”

Helmholtz School “Advanced Topics in Statistics”

Göttingen, 17-20 October 2010

Preamble

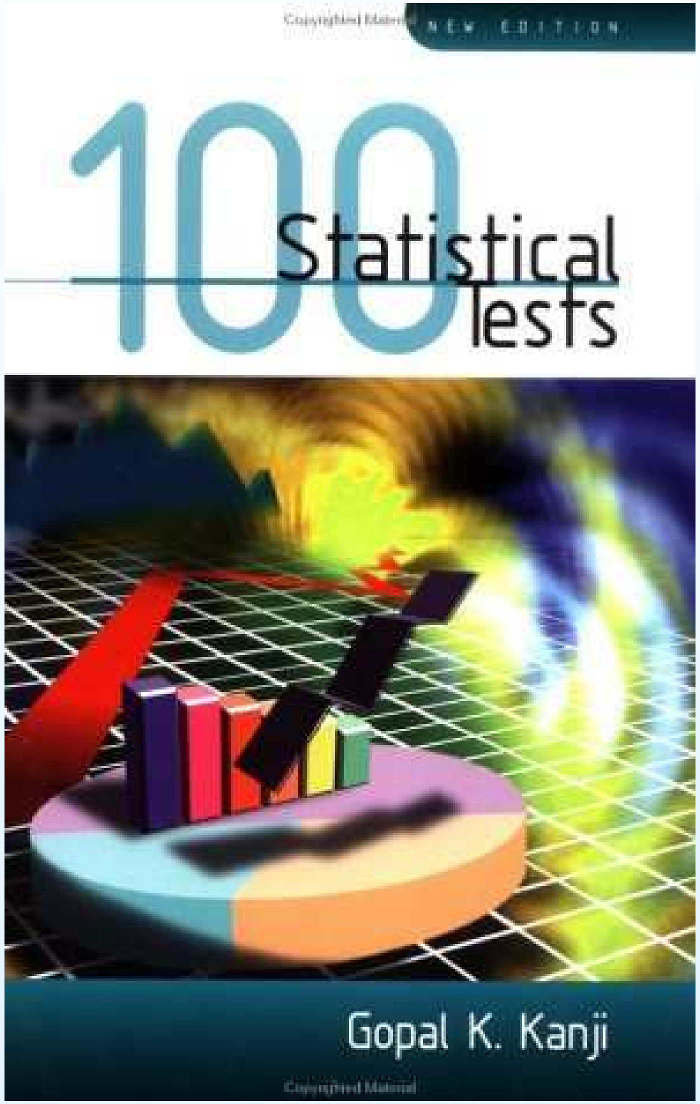
“Advanced topics”: ?

Preamble

“Advanced topics”: ?

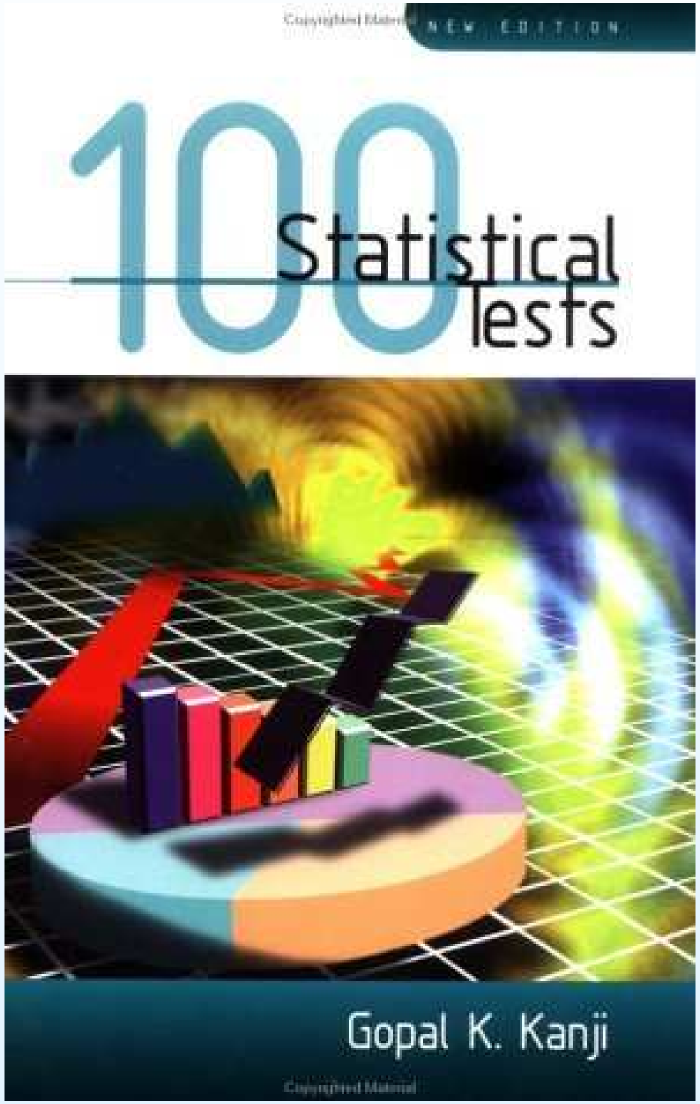
- Don’t expect fancy tests with Russian names

Preamble



Not exhaustive compilation. . .

Preamble



Not exhaustive compilation. . .

⇒ wikipedia.org/wiki/P-value#Frequent_misunderstandings

Preamble

"Advanced topics": ?

- Don't expect fancy tests with Russian names

$\Rightarrow$ An invitation to (re-)think on foundamental aspects, that help in developping applications

Preamble

"Advanced topics": ?

- Don't expect fancy tests with Russian names

⇒ An invitation to (re-)think on foundamental aspects, that help in developping applications

⇒ 'Forward to past'

Good and sane probabilistic reasoning
by Gauss, Laplace, etc.
(in contrast with XX century statisticians)

Preamble

"Advanced topics": ?

- Don't expect fancy tests with Russian names

$\Rightarrow$ An invitation to (re-)think on foundamental aspects, that help in developping applications

$\Rightarrow$ 'Forward to past'

$\Rightarrow$ Message to young people: improve quality of the teaching of probabilistic reasoning, recognized since centuries to be a weak point of the scholar system:

Preamble

"Advanced topics": ?

- Don't expect fancy tests with Russian names

⇒ An invitation to (re-)think on foundamental aspects, that help in developping applications

⇒ 'Forward to past'

⇒ Message to young people: improve quality of the teaching of probabilistic reasoning, recognized since centuries to be a weak point of the scholar system:

"The celebrated Monsieur Leibnitz *has observed it to be a defect in the common systems of logic, that they are very copious when they explain the operations of the understanding in the forming of demonstrations, but are too concise when they treat of probabilities, and those other measures of evidence on which life and action entirely depend, and which are our guides even in most of our philosophical speculations."* (D. Hume)

Preamble

"Advanced topics": ?

- Don't expect fancy tests with Russian names

⇒ An invitation to (re-)think on foundamental aspects, that help in developping applications

⇒ 'Forward to past'

⇒ Message to young people: improve quality of the teaching of probabilistic reasoning, recognized since centuries to be a weak point of the scholar system:

  ⇒ Not (magic) ad-hoc formulae, but a consistent probabilistic framework, capable to handle a large varity of problems

Preamble

"Advanced topics": ?

- Don't expect fancy tests with Russian names

$\Rightarrow$ An invitation to (re-)think on foundamental aspects, that help in developping applications

$\Rightarrow$ 'Forward to past'

$\Rightarrow$ Message to young people: improve quality of the teaching of probabilistic reasoning, recognized since centuries to be a weak point of the scholar system:

- Excellent philosophical introduction by Allen Caldwell

Preamble

"Advanced topics": ?

- Don't expect fancy tests with Russian names

$\Rightarrow$ An invitation to (re-)think on foundamental aspects, that help in developping applications

$\Rightarrow$ 'Forward to past'

$\Rightarrow$ Message to young people: improve quality of the teaching of probabilistic reasoning, recognized since centuries to be a weak point of the scholar system:

- Excellent philosophical introduction by Allen Caldwell . . .

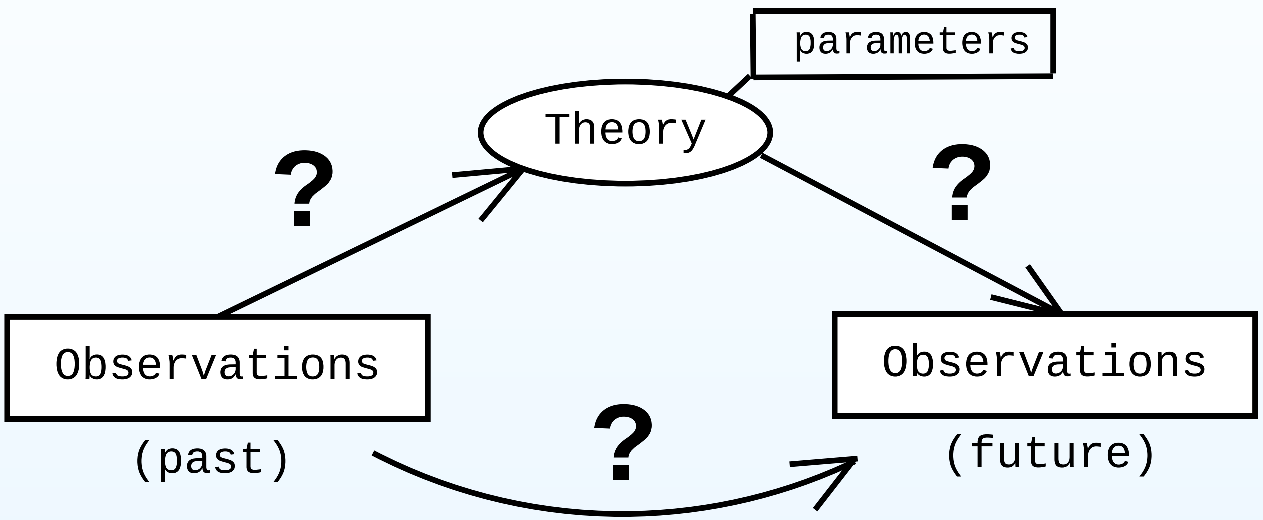
  that I will try to complement,
  before moving to a particular application.

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
  - Playing with 6 boxes and 30 balls
- 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)
- Some examples on toy models

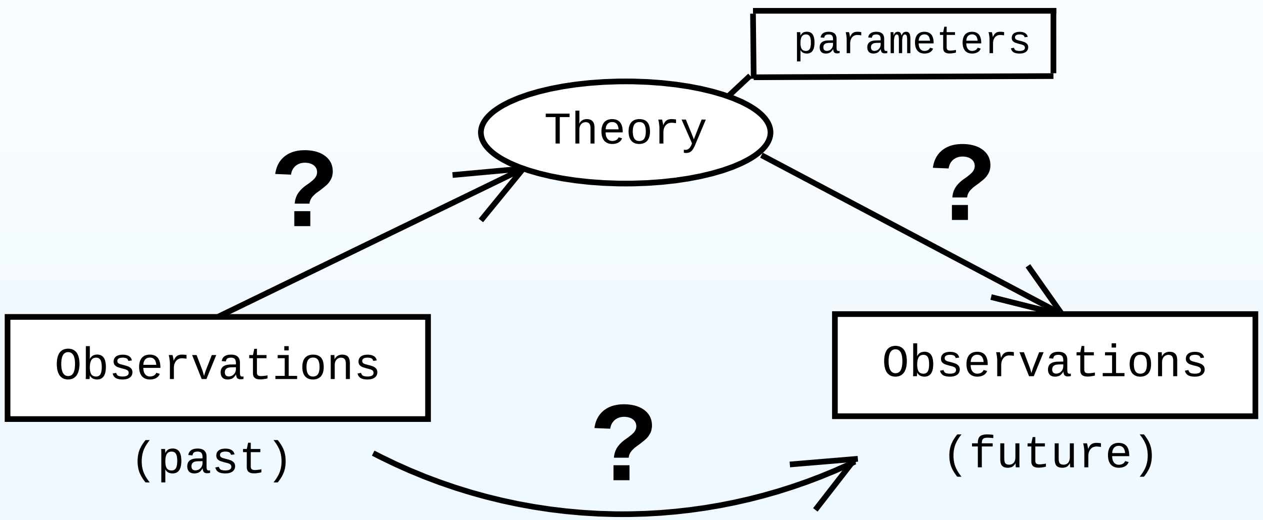
Learning from experience and source of uncertainty



Uncertainty:

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

Learning from experience and source of uncertainty



Uncertainty:

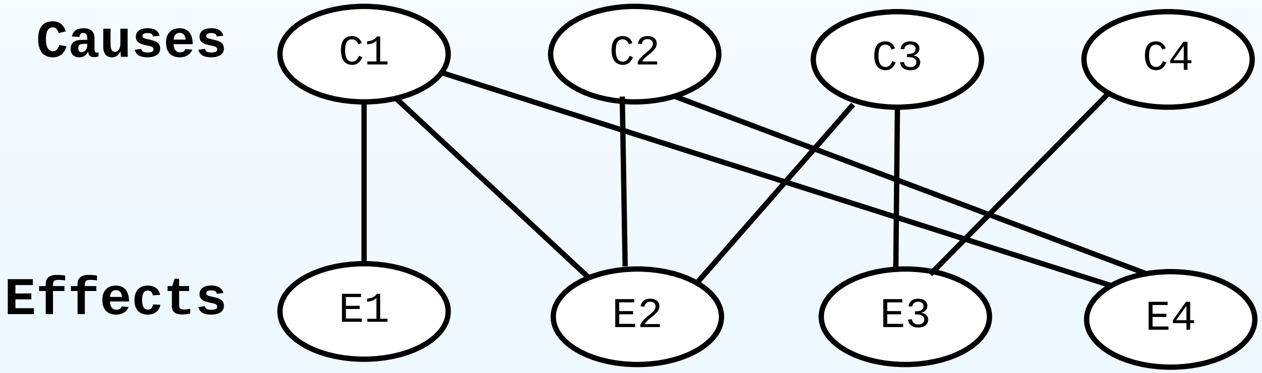
Theory	— ? ⟶	Future observations
Past observations	— ? ⟶	Theory
Past observations	— ? ⟶	Future observations

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

CAUSE $\Longleftrightarrow$ **EFFECT**

Causes $\longrightarrow$ effects

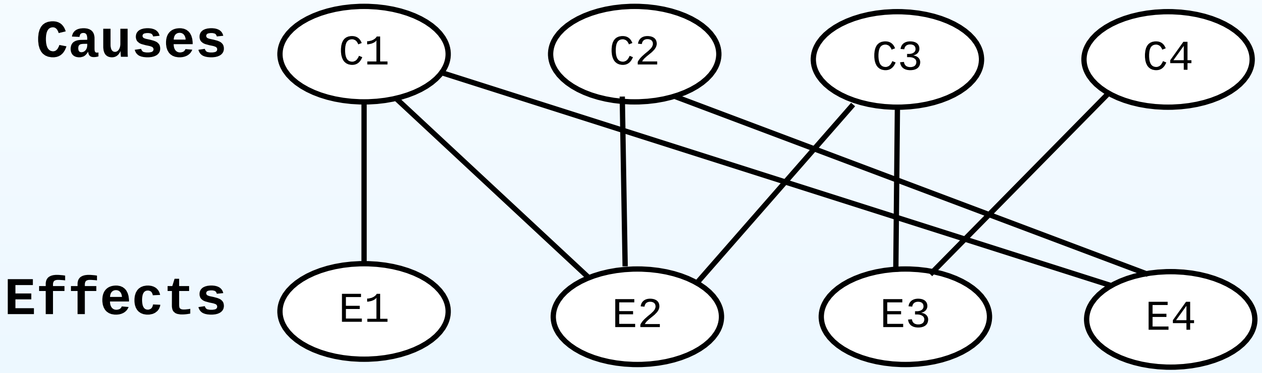
The same *apparent* cause might produce several,different effects



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

Causes $\longrightarrow$ effects

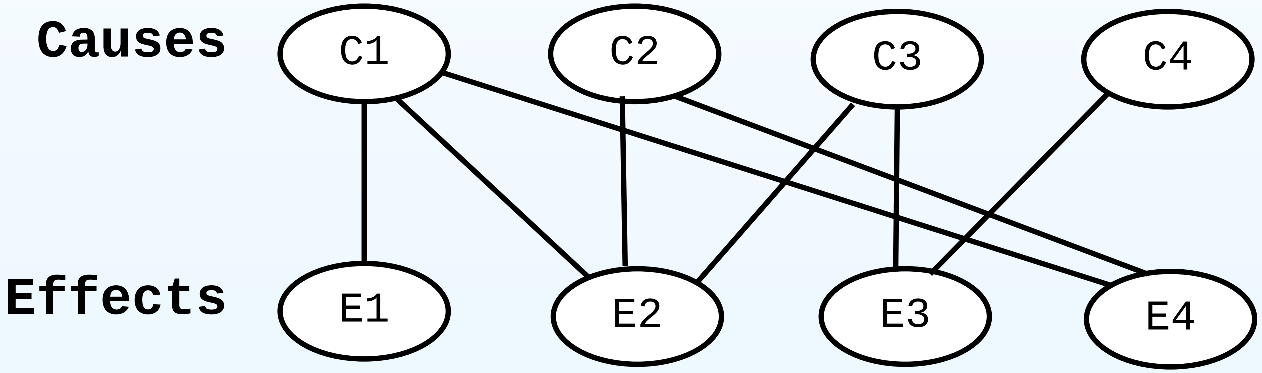
The same *apparent* cause might produce several,different effects



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

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.

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

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*)

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*)

- An essential problem of the experimental method would be expected to be thaught with special care in the first years of the physics curriculum. . .

Uncertainties in measurements

Having to perform a measurement:

Uncertainties in measurements

Having to perform a measurement:

> Which numbers shall come out from our device?

Having performed a measurement:

> What have we learned about the value of the quantity of interest?

How to quantify these kinds of uncertainty?

Uncertainties in measurements

Having to perform a measurement:

Which numbers shall come out from our device?

Having performed a measurement:

What have we learned about the value of the quantity of interest?

How to quantify these kinds of uncertainty?

Under well controlled conditions (calibration) we can make use of past frequencies to evaluate ‘somehow’ the detector response $P(x \mid \mu)$.

Uncertainties in measurements

Having to perform a measurement:

Which numbers shall come out from our device?

Having performed a measurement:

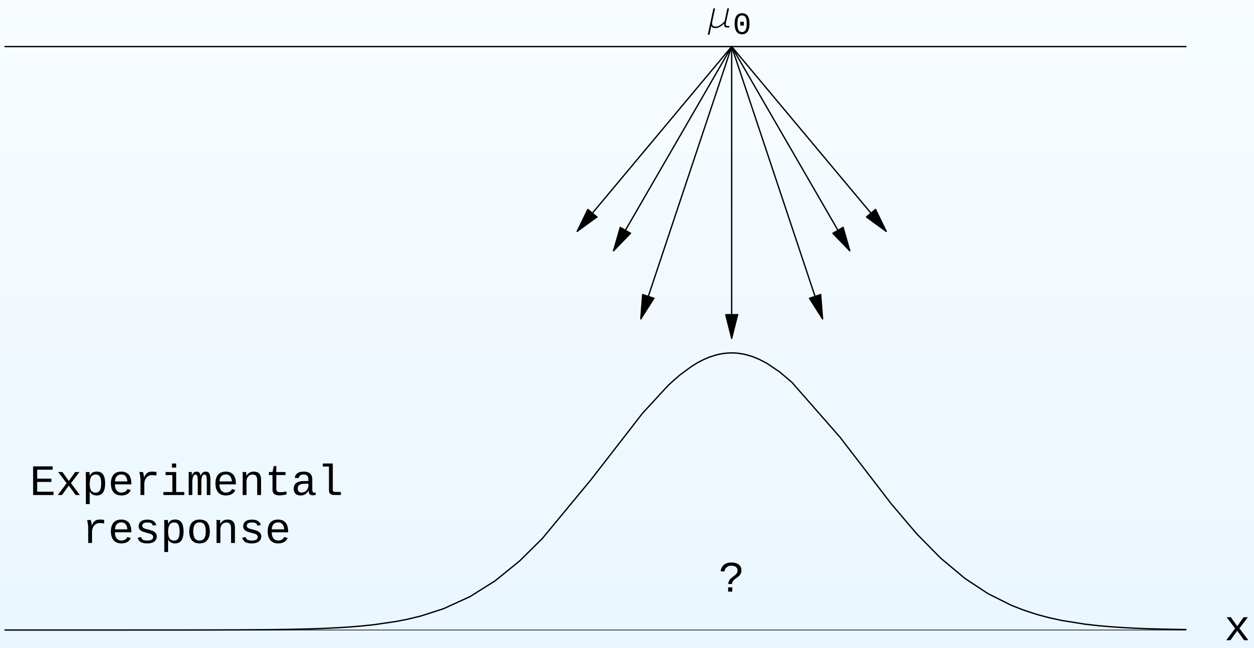
What have we learned about the value of the quantity of interest?

How to quantify these kinds of uncertainty?

Under well controlled conditions (calibration) we can make use of past frequencies to evaluate 'somehow' the detector response $P(x \mid \mu)$.

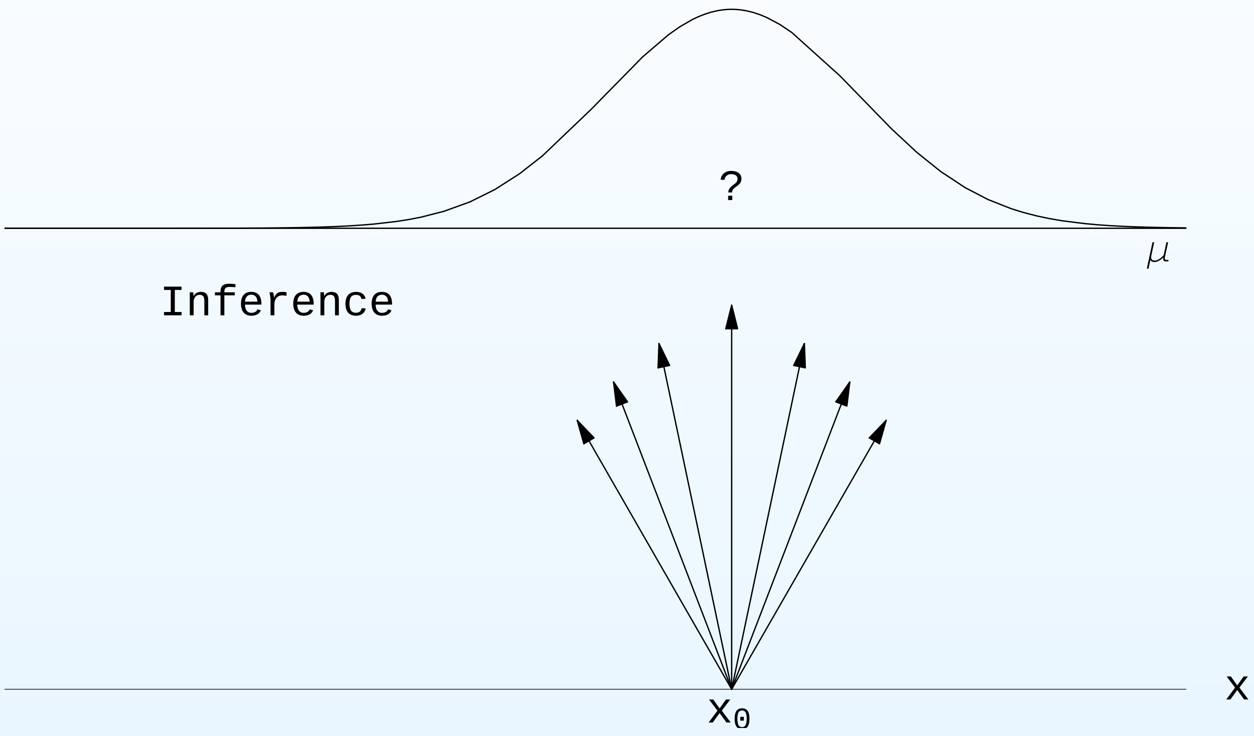
There is (in most cases) no way to get *directly* hints about $P(\mu \mid x)$.

Uncertainties in measurements



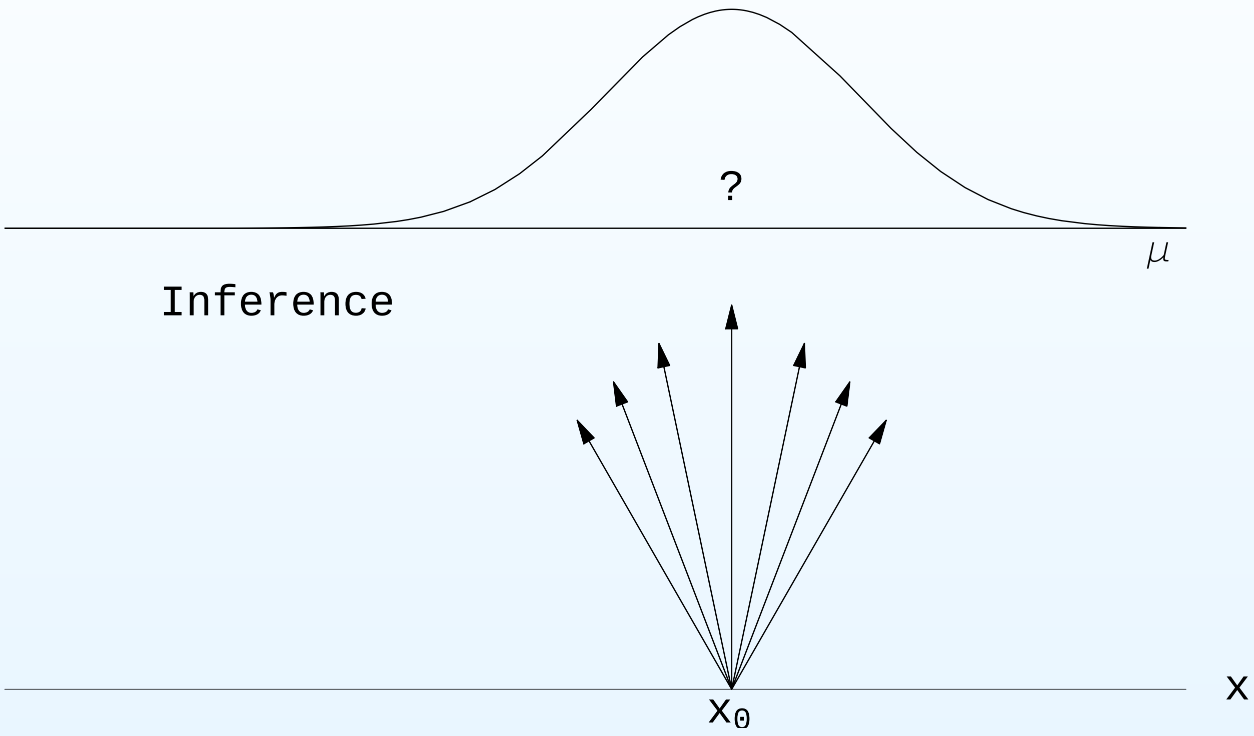
$P(x \mid \mu)$ experimentally accessible (though 'model filtered')

Uncertainties in measurements



$P(\mu \,|\, x)$ experimentally inaccessible

Uncertainties in measurements

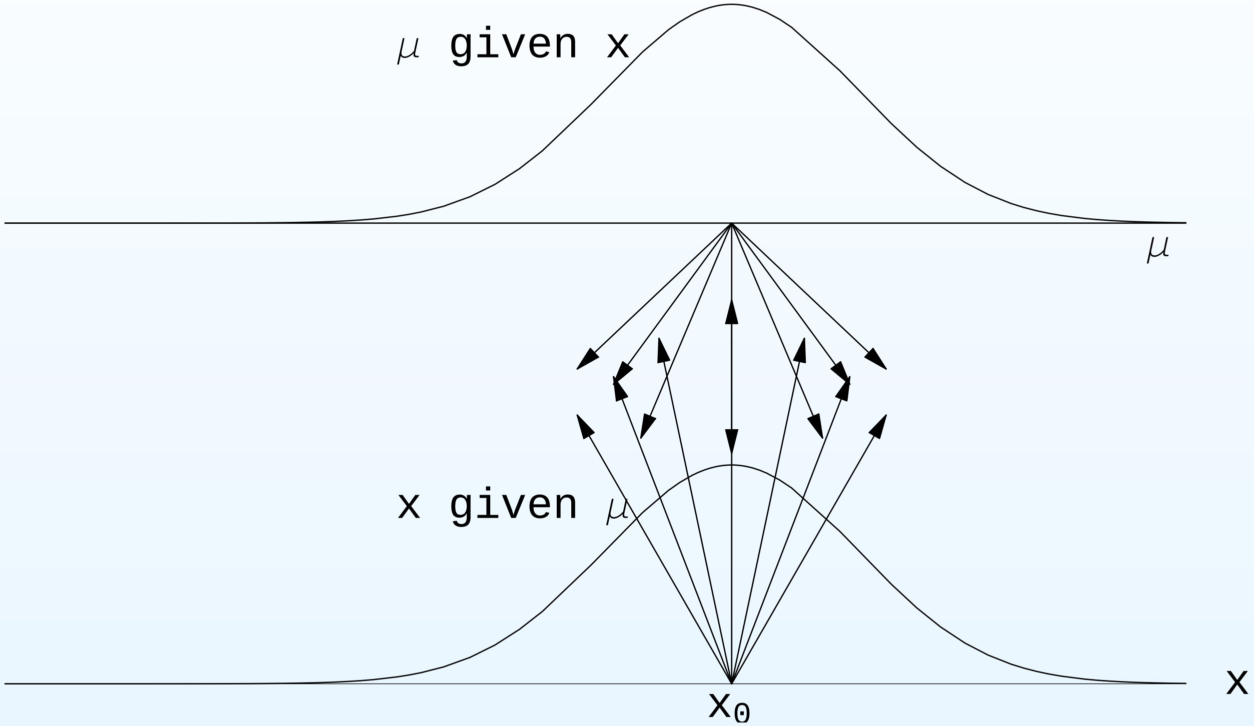


$P(\mu \mid x)$ experimentally inaccessible

but logically accessible!

$\longrightarrow$ we need to learn how to do it

Uncertainties in measurements



Symmetry in reasoning!

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

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

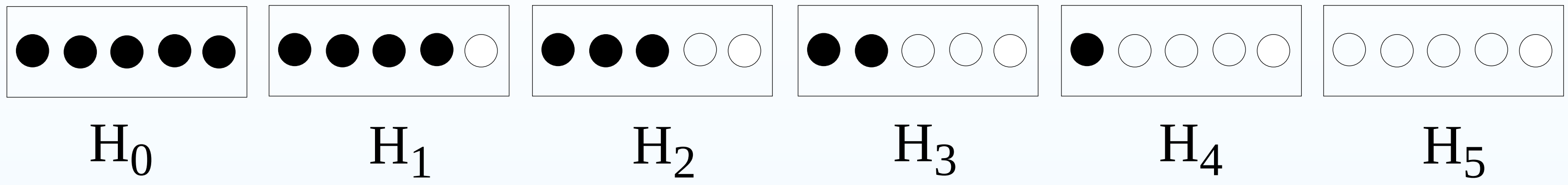
...although, such statements are considered

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

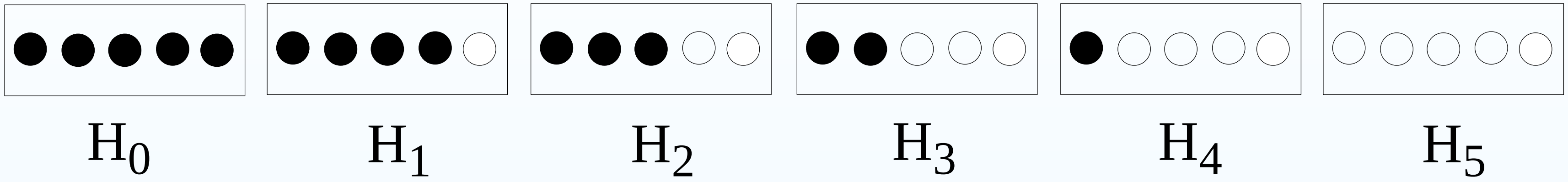
(see D. Hume, C. Darwin + modern researches)

The six box problem



Let us take randomly one of the boxes.

The six box problem



Let us take randomly one of the boxes.

We are in a state of uncertainty concerning several *events*, the most important of which correspond to the following questions:

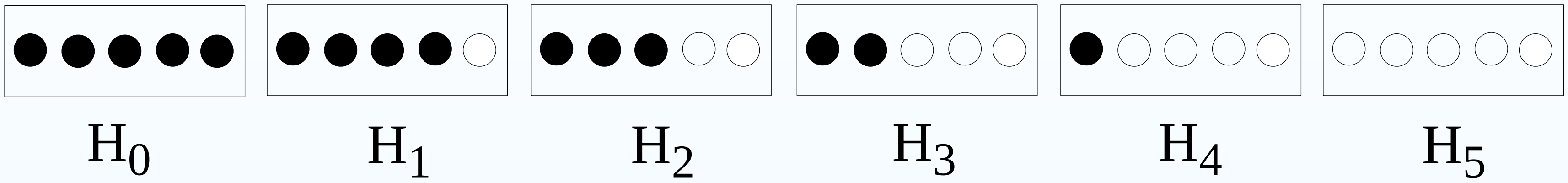
(a) Which box have we chosen, $H_0, H_1, \ldots, H_5$?

(b) If we extract randomly a ball from the chosen box, will we observe a white ($E_W \equiv E_1$) or black ($E_B \equiv E_2$) ball?

Our certainty:

$$\begin{aligned} \cup_{j=0}^{5} H_j &= \Omega \\ \cup_{i=1}^{2} E_i &= \Omega\,. \end{aligned}$$

The six box problem



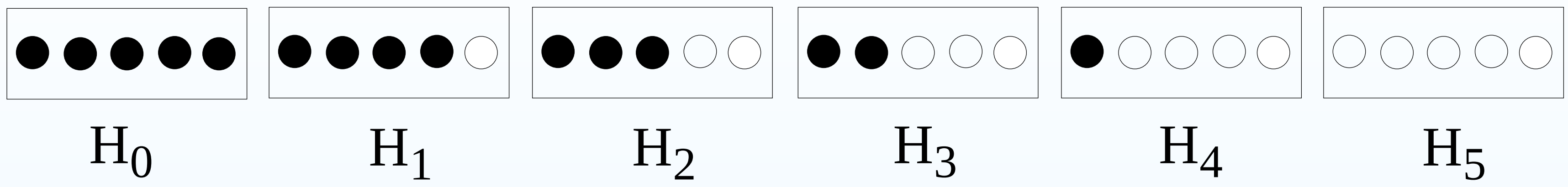
Let us take randomly one of the boxes.
We are in a state of uncertainty concerning several *events*, the most important of which correspond to the following questions:

(a) Which box have we chosen, $H_0, H_1, \ldots, H_5$?

(b) If we extract randomly a ball from the chosen box, will we observe a white ($E_W \equiv E_1$) or black ($E_B \equiv E_2$) ball?

- What happens after we have extracted one ball and looked its color?
  - Intuitively we now how to roughly change our opinion.
  - Can we do it quantitatively, in an objective way?

The six box problem



Let us take randomly one of the boxes.
We are in a state of uncertainty concerning several *events*, the most important of which correspond to the following questions:

(a) Which box have we chosen, $H_0, H_1, \ldots, H_5$?

(b) If we extract randomly a ball from the chosen box, will we observe a white ($E_W \equiv E_1$) or black ($E_B \equiv E_2$) ball?

- What happens after we have extracted one ball and looked its color?
  - Intuitively we now how to roughly change our opinion.
  - Can we do it quantitatively, in an objective way?
- And after a sequence of extractions?

Predicting sequences

Side remark/exercise

Imagine the four possible sequences resulting from the first two extractions from the misterious box:

BB, BW, WB and WW

- How likely do you consider them to occur?
  [$\longrightarrow$ If you could win a prize associated with the occurrence of one of them, on which sequence(s) would you bet?]

Predicting sequences

Side remark/exercise

Imagine the four possible sequences resulting from the first two extractions from the misterious box:

BB, BW, WB and WW

- How likely do you consider them to occur?
  [$\longrightarrow$ If you could win a prize associated with the occurrence of one of them, on which sequence(s) would you bet?]
- Or do you consider them equally likelly?

Predicting sequences

Side remark/exercise

Imagine the four possible sequences resulting from the first two extractions from the misterious box:

BB, BW, WB and WW

- How likely do you consider them to occur?
  [$\longrightarrow$ If you could win a prize associated with the occurrence of one of them, on which sequence(s) would you bet?]
- Or do you consider them equally likelly?
- ?

Predicting sequences

Side remark/exercise

Imagine the four possible sequences resulting from the first two extractions from the misterious box:

BB, BW, WB and WW

- How likely do you consider them to occur?
  [$\longrightarrow$ If you could win a prize associated with the occurrence of one of them, on which sequence(s) would you bet?]
- Or do you consider them equally likelly?
- ?
- No, they are not equally likelly!

Predicting sequences

Side remark/exercise

Imagine the four possible sequences resulting from the first two extractions from the misterious box:

BB, BW, WB and WW

- How likely do you consider them to occur?
  [$\longrightarrow$ If you could win a prize associated with the occurrence of one of them, on which sequence(s) would you bet?]
- Or do you consider them equally likelly?
- ?
- **No, they are not equally likelly!**

  Laplace new perfectly why
  $\longrightarrow$ If our logical abilities have regressed it is not a good sign!
  (Remember Leibnitz/Hume quote)

The toy inferential experiment

The aim of the experiment will be to guess the content of the box without looking inside it, only extracting a ball, recording its color and reintroducing it into the box

The toy inferential experiment

The aim of the experiment will be to guess the content of the box without looking inside it, only extracting a ball, recording its color and reintroducing it into the box

This toy experiment is conceptually very close to what we do in Physics

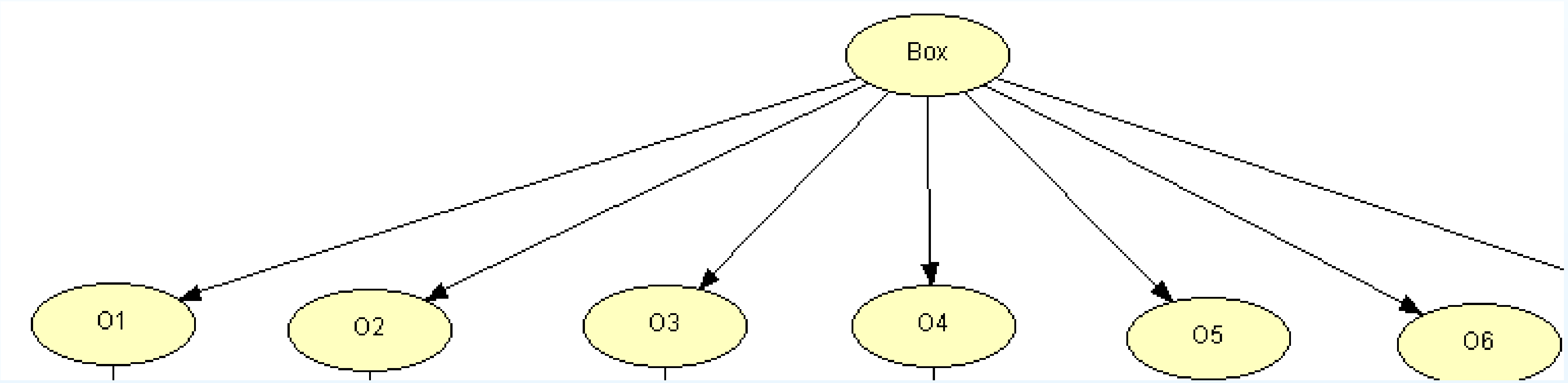
- try to guess what we cannot see (the electron mass, a branching ratio, etc)

  . . . from what we can see (somehow) with our senses.

The rule of the game is that we are not allowed to watch inside the box! (As we cannot open and electron and read its properties, like we read the MAC address of a PC interface)

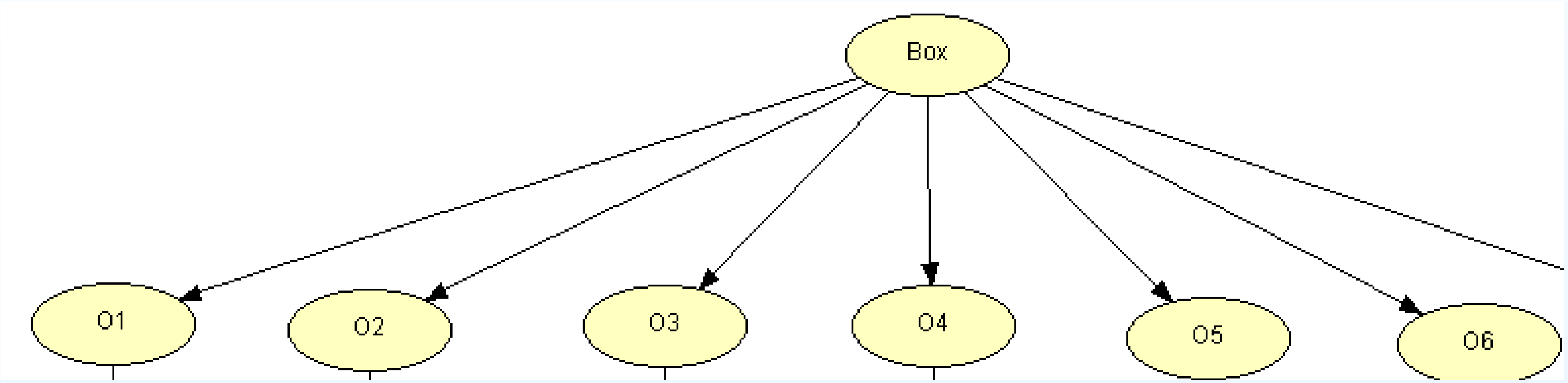
Cause-effect representation

box content → observed color



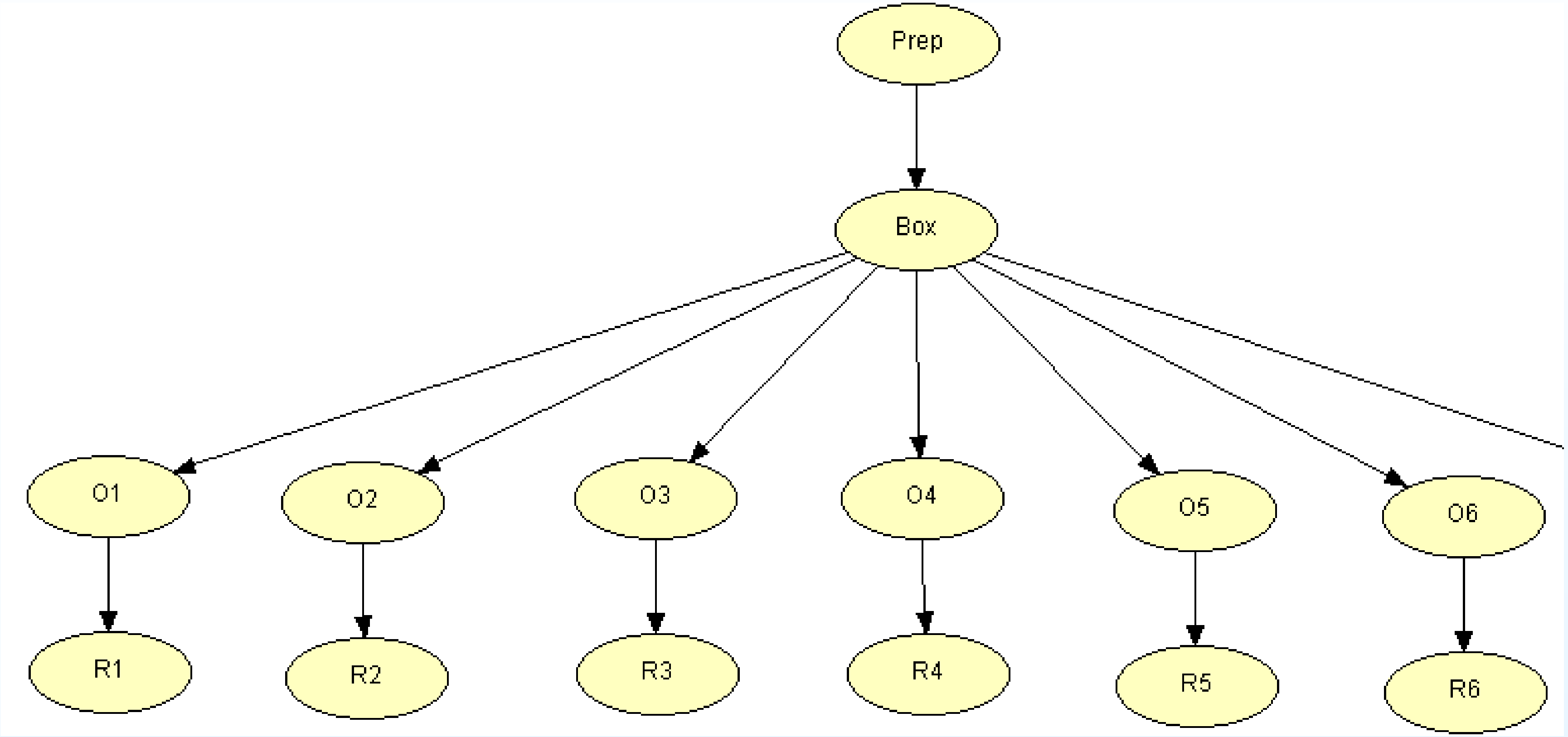
Cause-effect representation

box content $\longrightarrow$ observed color

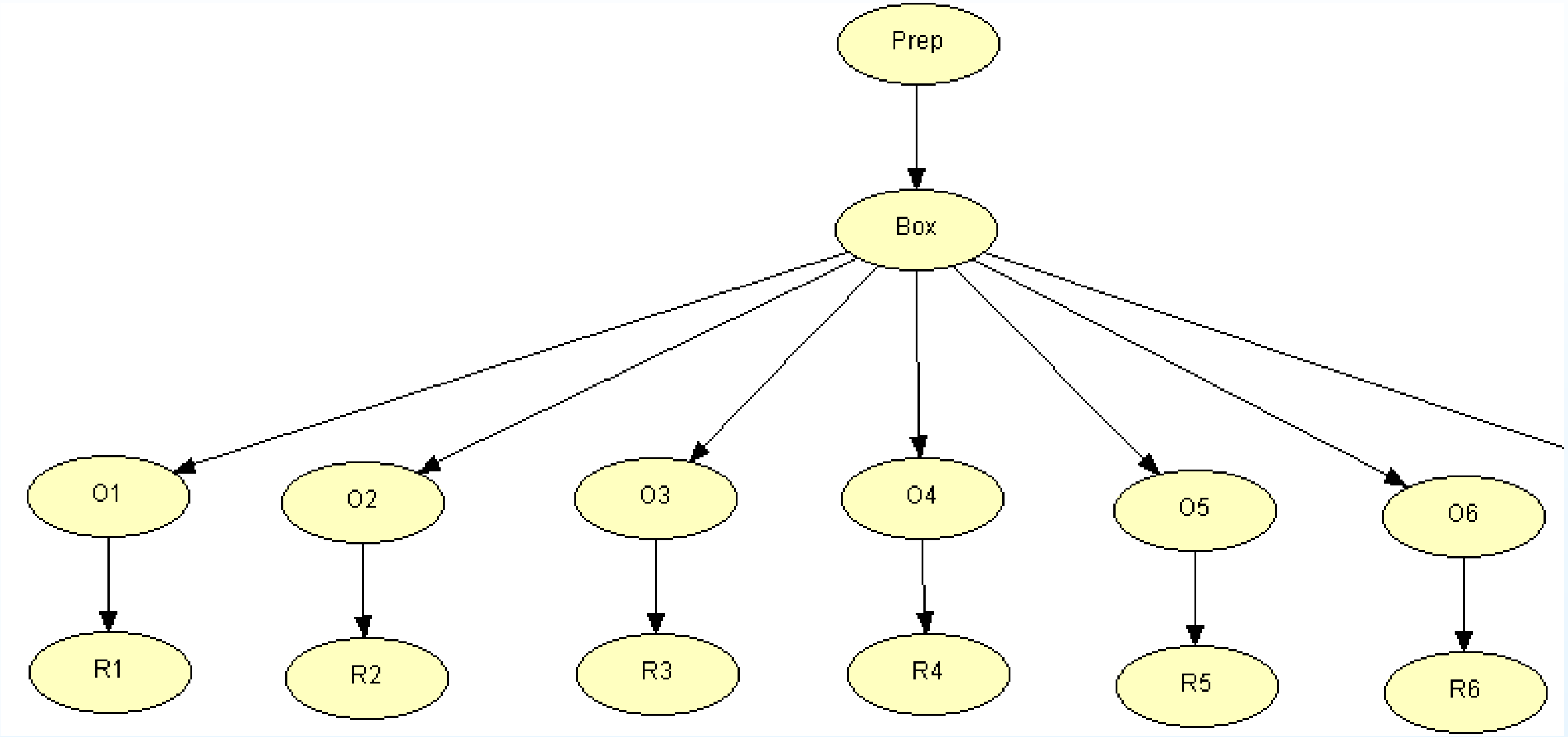


An effect might be the cause of another effect $\longrightarrow$

A network of causes and effects

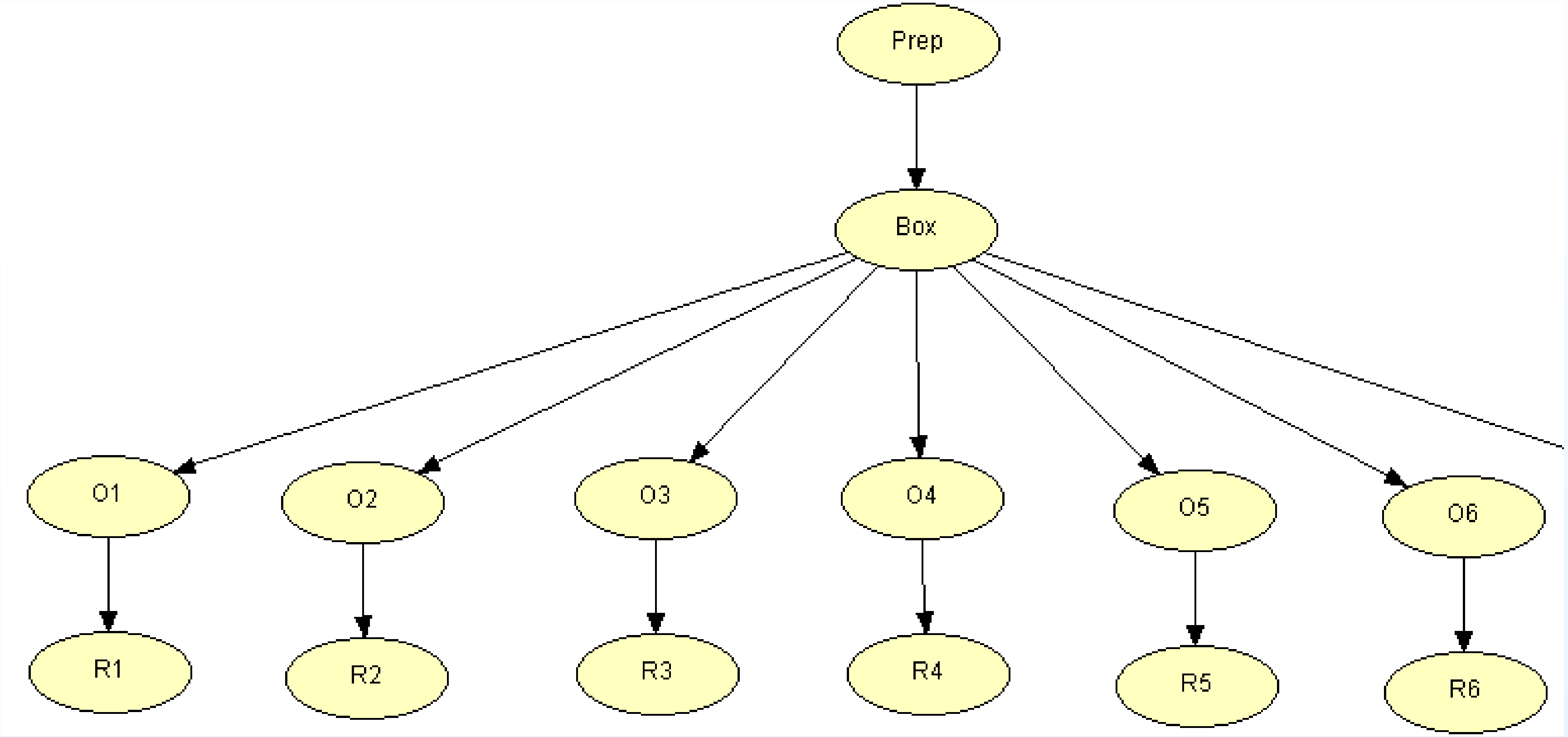


A network of causes and effects



A report (R_i) might not correspond exactly to what really happened (O_i)

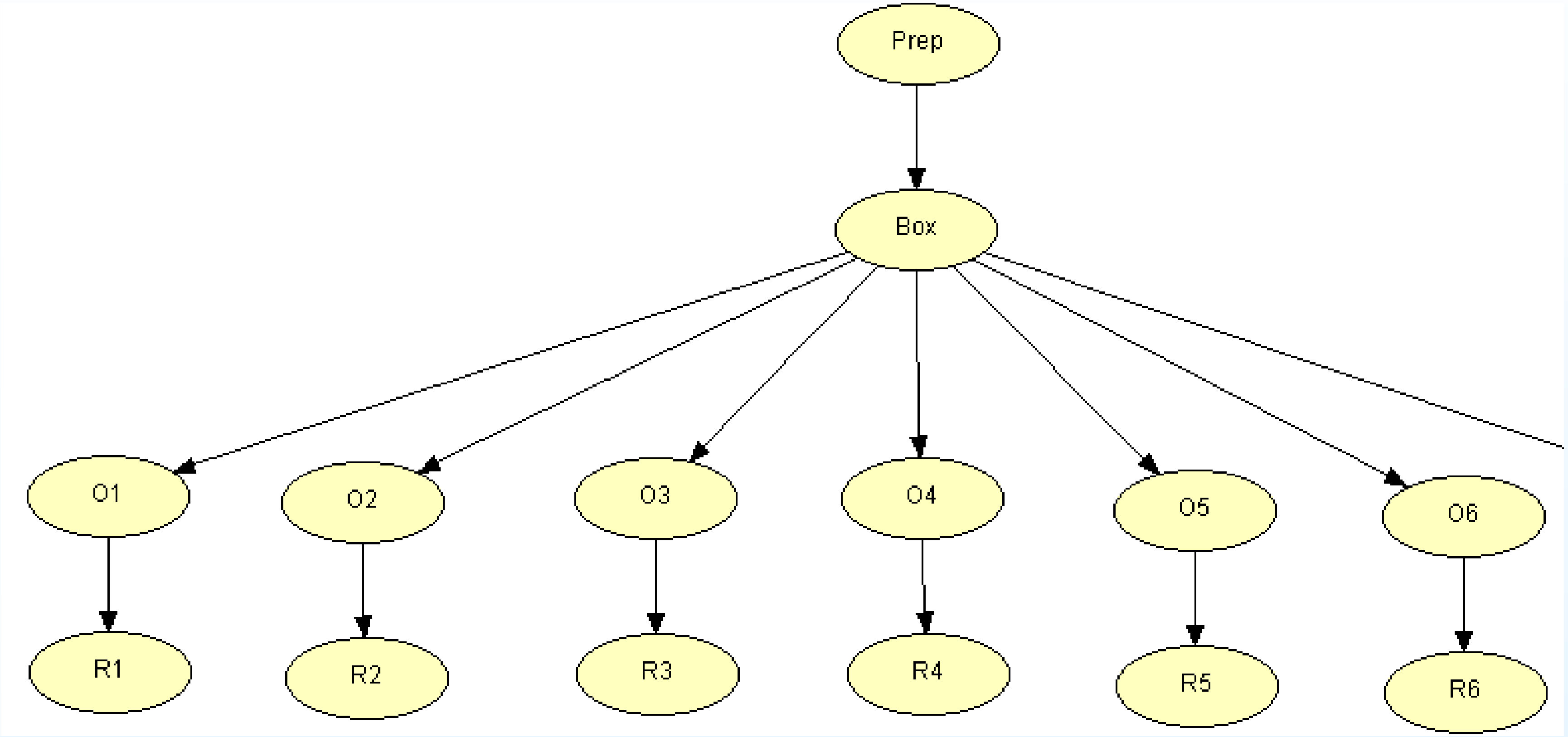
A network of causes and effects



Of crucial interest in Science!

$\Rightarrow$ Our devices seldom tell us 'the truth'.

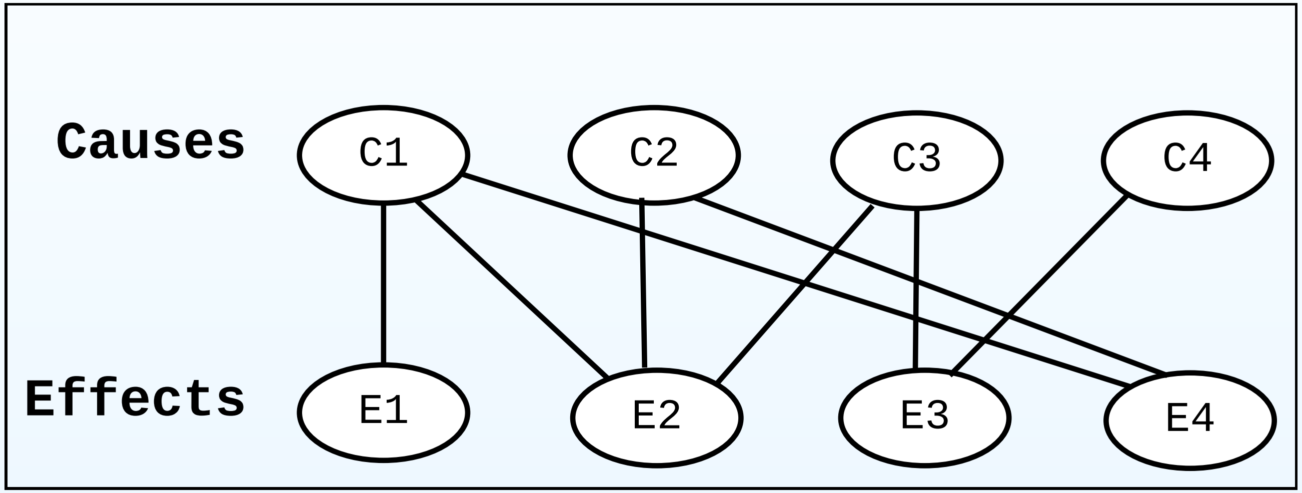
A network of causes and effects



$\Rightarrow$ Belief Networks (Bayesian Networks)

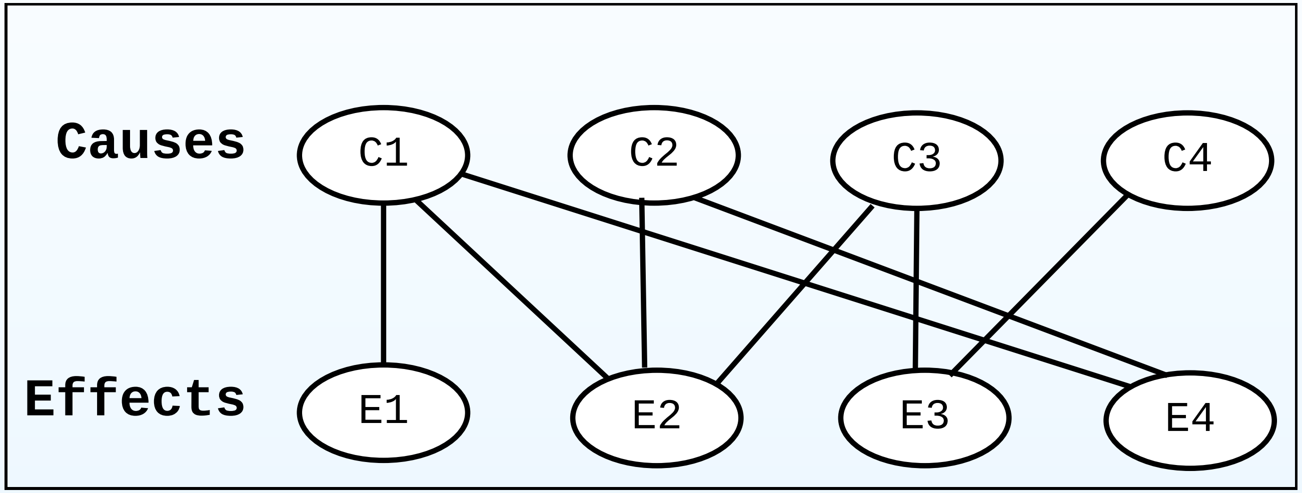
From causes to effects and back

Our original problem:



From causes to effects and back

Our original problem:

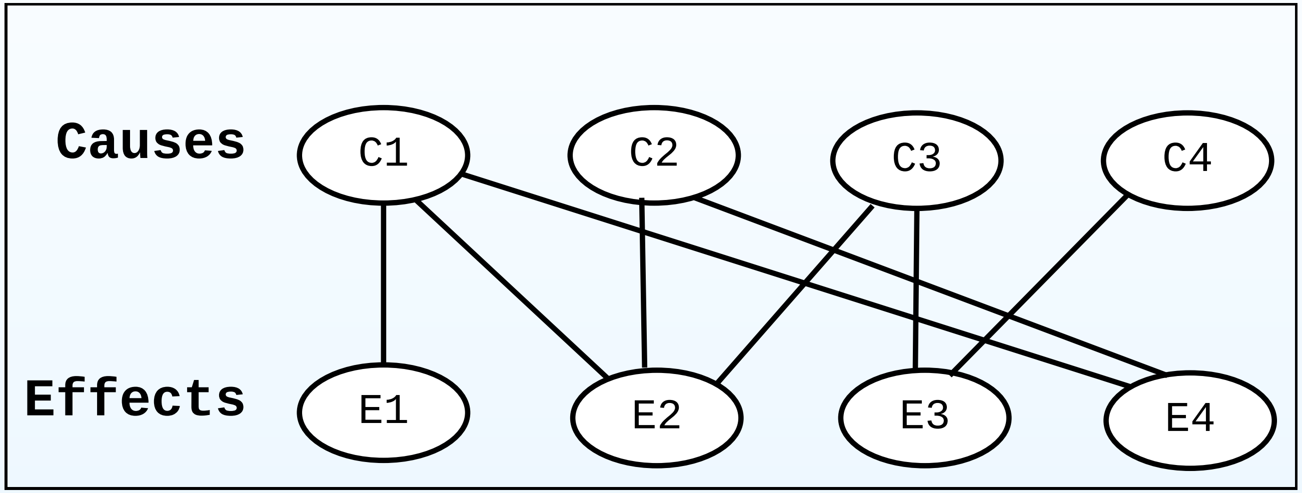


Our conditional view of probabilistic causation

$$P(E_i \,|\, C_j)$$

From causes to effects and back

Our original problem:



Our conditional view of probabilistic causation

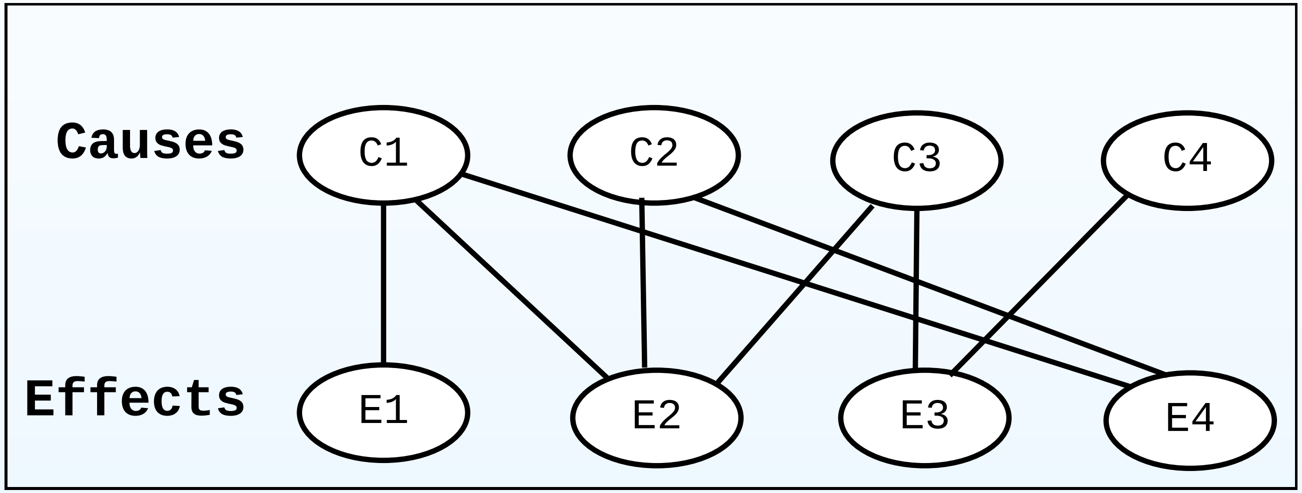
$$\boxed{P(E_i \,|\, C_j)}$$

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

$$P(E_i \,|\, C_j)$$

Our conditional view of probabilistic inference

$$P(C_j \,|\, E_i)$$

The fourth basic rule of probability:

$$P(C_j, E_i) = P(E_i \,|\, C_j)\, P(C_j) = P(C_j \,|\, E_i)\, P(E_i)$$

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

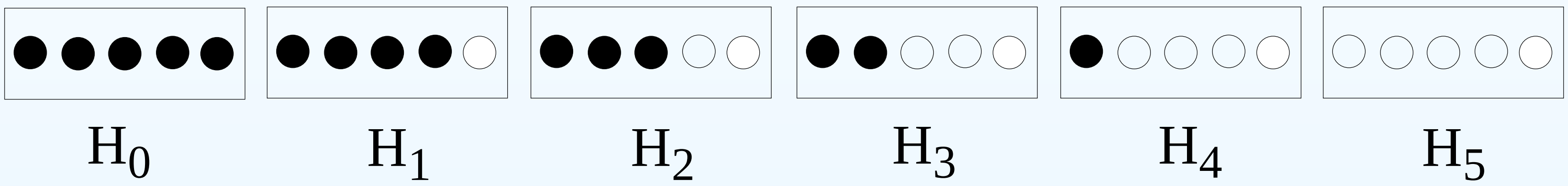
$$P(H_j\,|\,E_i) = \frac{P(E_i\,|\,H_j)}{P(E_i)}\,P(H_j)$$

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

(Gauss)

$\Rightarrow$ Bayes theorem

Application to the six box problem



Remind:

- $E_1 =$ White
- $E_2 =$ Black

Collecting the pieces of information we need

Our tool:

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

Collecting the pieces of information we need

Our tool:

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

- $P(H_j \,|\, I) = 1/6$

Collecting the pieces of information we need

Our tool:

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

- $P(H_j \mid I) = 1/6$
- $P(E_i \mid I) = 1/2$

Collecting the pieces of information we need

Our tool:

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

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j,\, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j,\, I) &= j/5 \\ P(E_2 \,|\, H_j,\, I) &= (5-j)/5 \end{aligned}$$

Collecting the pieces of information we need

Our tool:

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

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j,\, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j,\, I) &= j/5 \\ P(E_2 \,|\, H_j,\, I) &= (5-j)/5 \end{aligned}$$

Our prior belief about H_j

Collecting the pieces of information we need

Our tool:

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

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j, I) &= j/5 \\ P(E_2 \,|\, H_j, I) &= (5-j)/5 \end{aligned}$$

Probability of E_i under a well defined hypothesis H_j
It corresponds to the 'response of the apparatus in measurements.
→ likelihood (traditional, rather confusing name!)

Collecting the pieces of information we need

Our tool:

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

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j,\, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j,\, I) &= j/5 \\ P(E_2 \,|\, H_j,\, I) &= (5-j)/5 \end{aligned}$$

Probability of E_i taking account all possible H_j
$\rightarrow$ How much we are confident that E_i will occur.

Collecting the pieces of information we need

Our tool:

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

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j,\, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j,\, I) &= j/5 \\ P(E_2 \,|\, H_j,\, I) &= (5-j)/5 \end{aligned}$$

Probability of E_i taking account all possible H_j
$\rightarrow$ How much we are confident that E_i will occur.
Easy in this case, because of the symmetry of the problem.
But already after the first extraction of a ball our opinion about the box content will change, and symmetry will break.

Collecting the pieces of information we need

Our tool:

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

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j, I) &= j/5 \\ P(E_2 \,|\, H_j, I) &= (5-j)/5 \end{aligned}$$

But it easy to prove that $P(E_i \,|\, I)$ is related to the other ingredients, usually easier to ‘measure’ or to assess somehow, though vaguely

Collecting the pieces of information we need

Our tool:

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

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j, \, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j, \, I) &= j/5 \\ P(E_2 \,|\, H_j, \, I) &= (5-j)/5 \end{aligned}$$

But it easy to prove that $P(E_i \,|\, I)$ is related to the other ingredients, usually easier to 'measure' or to assess somehow, though vaguely

'decomposition law': $P(E_i \,|\, I) = \sum_j P(E_i \,|\, H_j, \, I) \cdot P(H_j \,|\, I)$

($\rightarrow$ Easy to check that it gives $P(E_i \,|\, I) = 1/2$ in our case).

Collecting the pieces of information we need

Our tool:

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

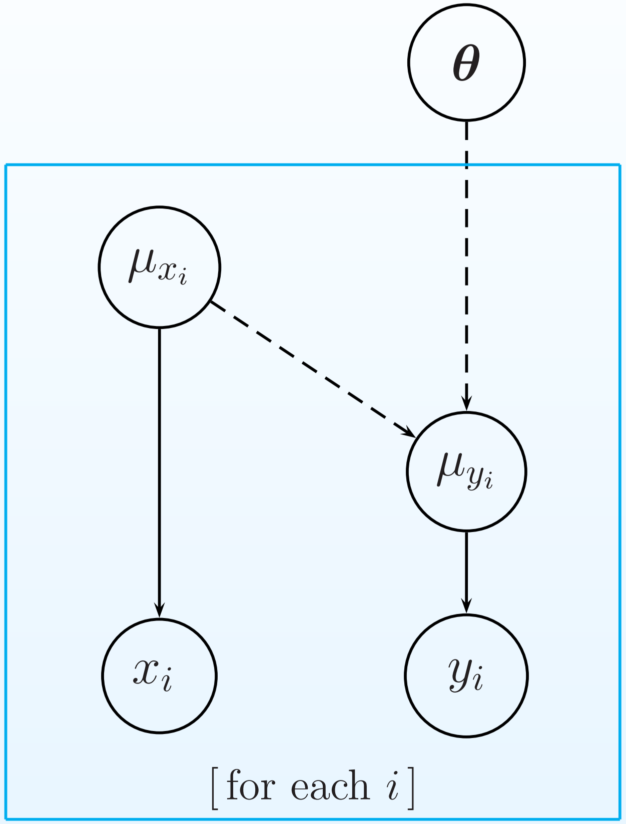
- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = \sum_j P(E_i \,|\, H_j, I) \cdot P(H_j \,|\, I)$
- $P(E_i \,|\, H_j, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j, I) &= j/5 \\ P(E_2 \,|\, H_j, I) &= (5-j)/5 \end{aligned}$$

We are ready

$\longrightarrow$ Let's play!

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)

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

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

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)

Bin to bin?

En passant:

- OK if the are no migrations:
  $\longrightarrow$ each bin is an ‘independent issue’,
  treated with a binomial process, given some efficiencies.

Bin to bin?

En passant:

- OK if the are no migrations:
  $\longrightarrow$ each bin is an 'independent issue',
  treated with a binomial process, given some efficiencies.
- Otherwise
  - 'error analysis' troublesome
    (just imagine e.g. that a bin has an 'efficiency' > 1,
    because of migrations from other bins);
  - iteration is important
    (efficiencies depend on 'true distribution')

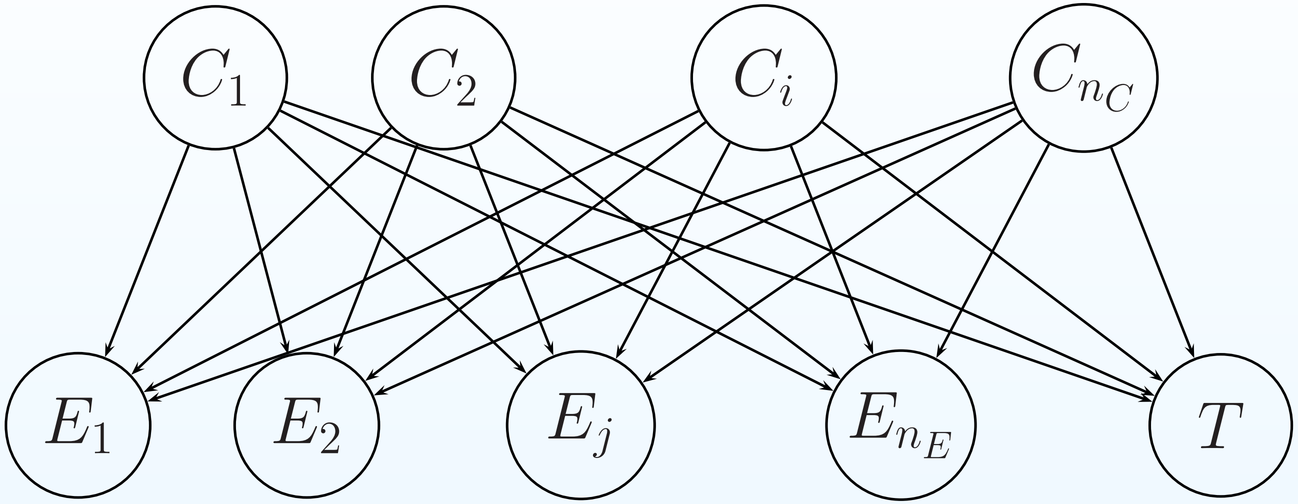
Bin to bin?

En passant:

- OK if the are no migrations:
  $\longrightarrow$ each bin is an 'independent issue',
  treated with a binomial process, given some efficiencies.
- Otherwise
  - 'error analysis' troublesome
    (just imagine e.g. that a bin has an 'efficiency' > 1,
    because of migrations from other bins);
  - iteration is important
    (efficiencies depend on 'true distribution')

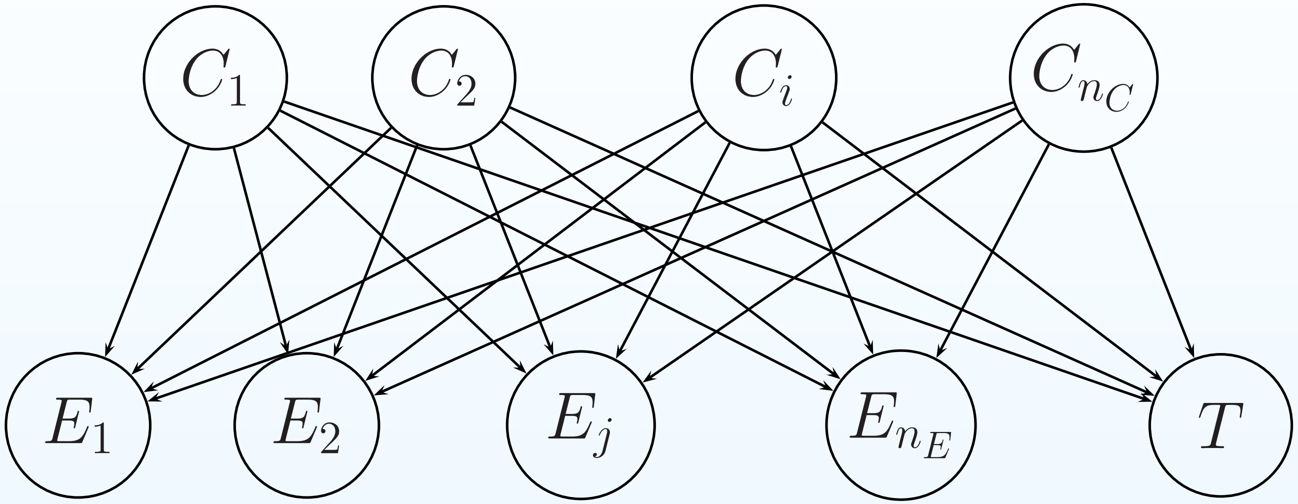
[Anyway, one might set up a procedure for a specific problem, test it with simulations and apply it to real data (the frequentistic way – if ther is *the way*...)]

Discretized unfolding



(T: ‘trash’)

Discretized unfolding

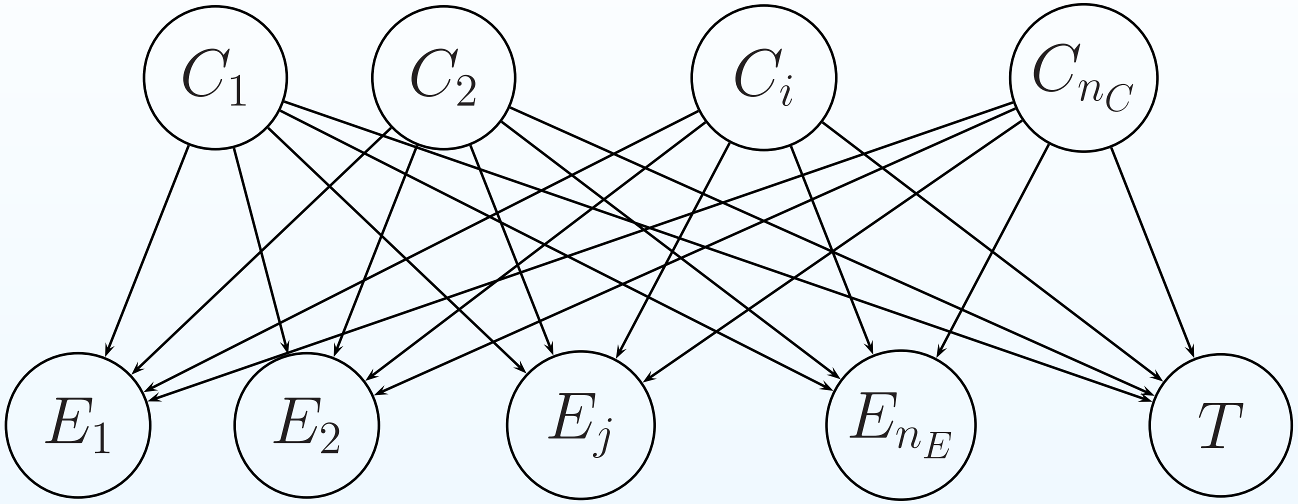


(T: 'trash')

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

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

Discretized unfolding



(T: 'trash')

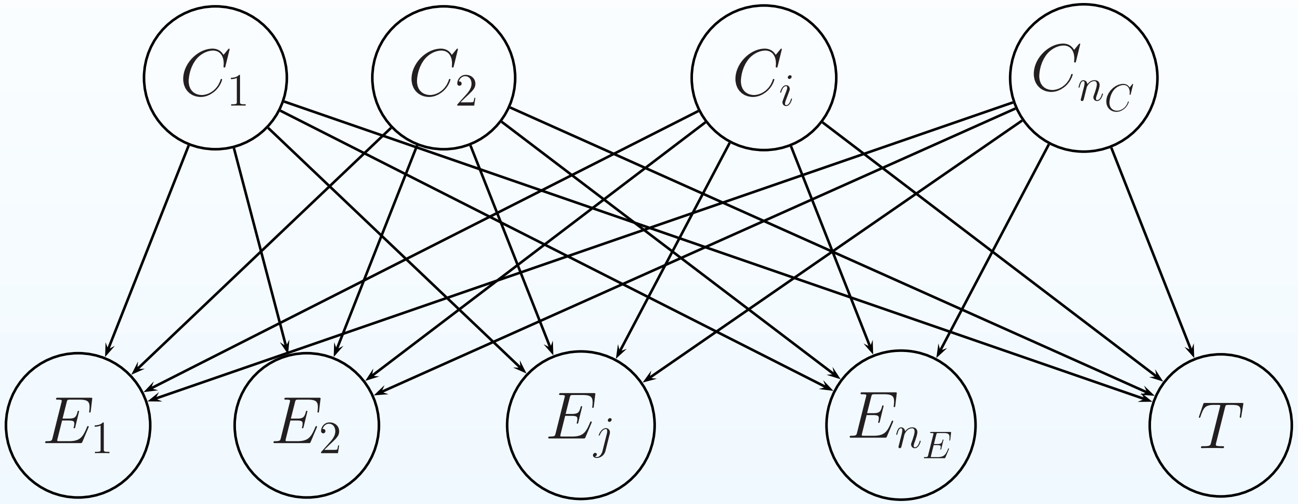
$\boldsymbol{x}_C$: *true spectrum* (nr of events in cause bins)

$\boldsymbol{x}_E$: *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:
  $\Rightarrow P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I)$

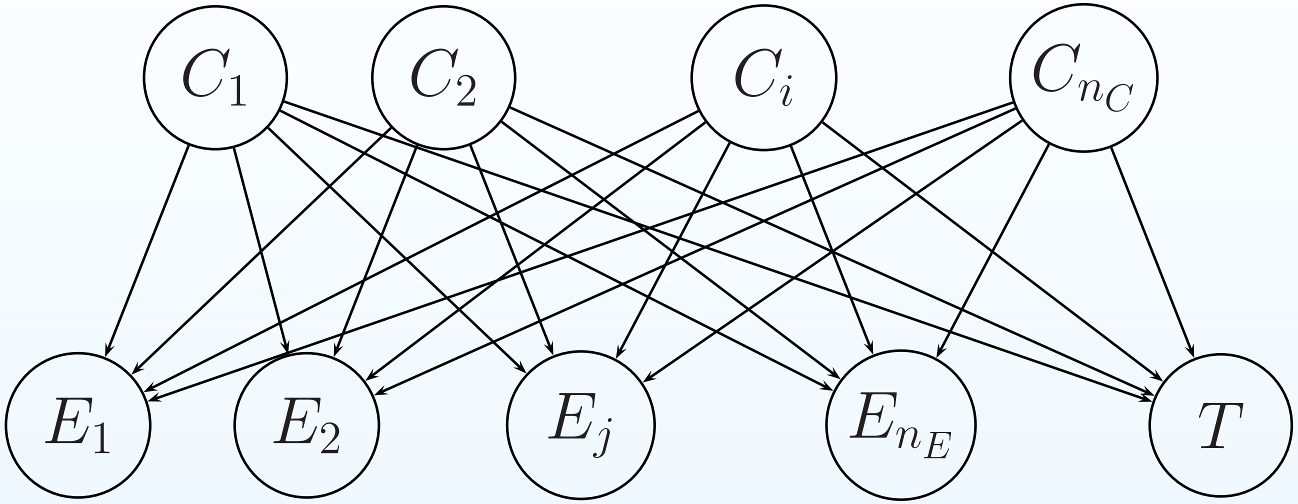
Discretized unfolding



(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

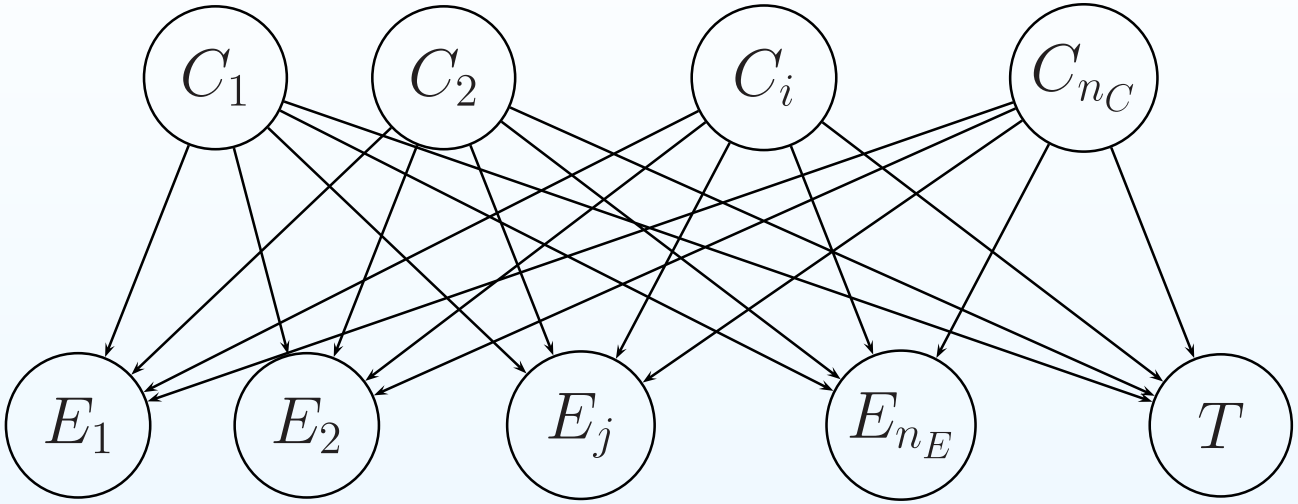


(T: ‘trash’)

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

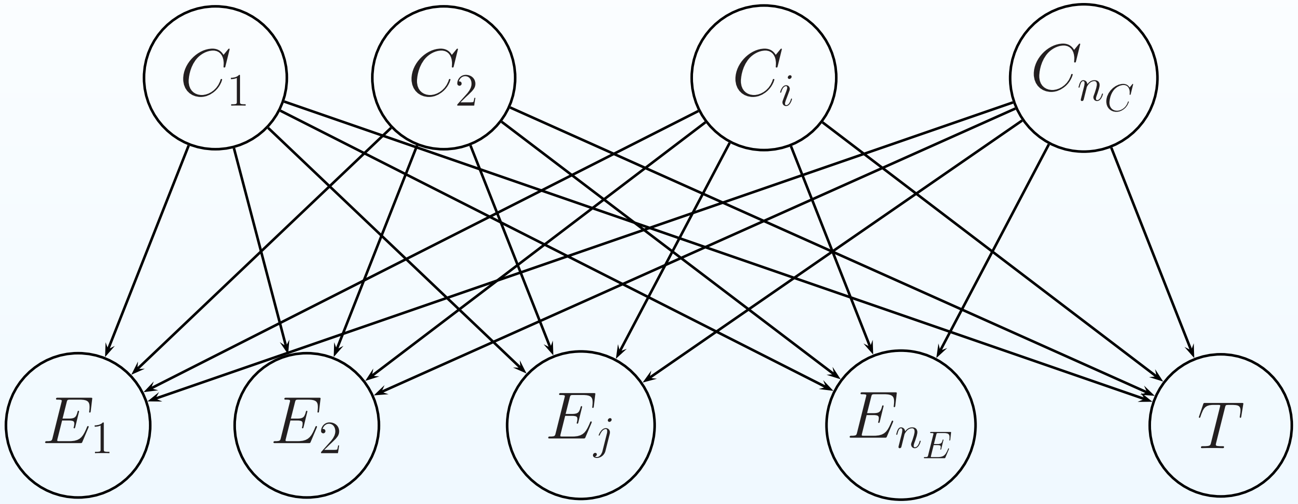


(T: 'trash')

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

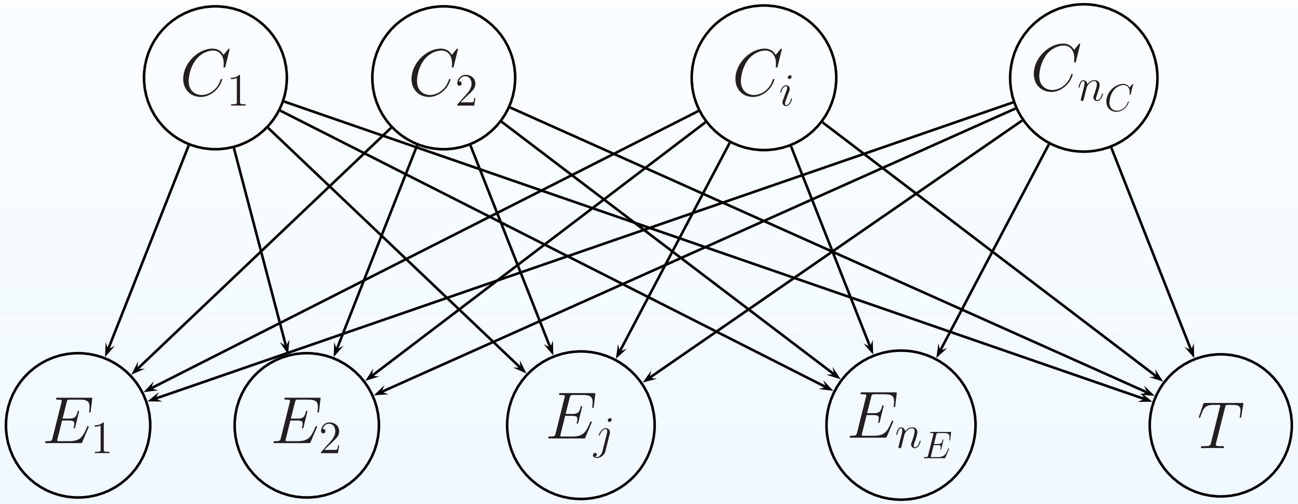
Discretized unfolding



(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:
  $\rightarrow P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda, I)$
  $\Rightarrow P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, I) = \int P(\boldsymbol{x}_C \,|\, \boldsymbol{x}_E, \Lambda,\, I)\, f(\Lambda \,|\, I)\, \mathrm{d}\Lambda$ [by MC!]

Discretized unfolding

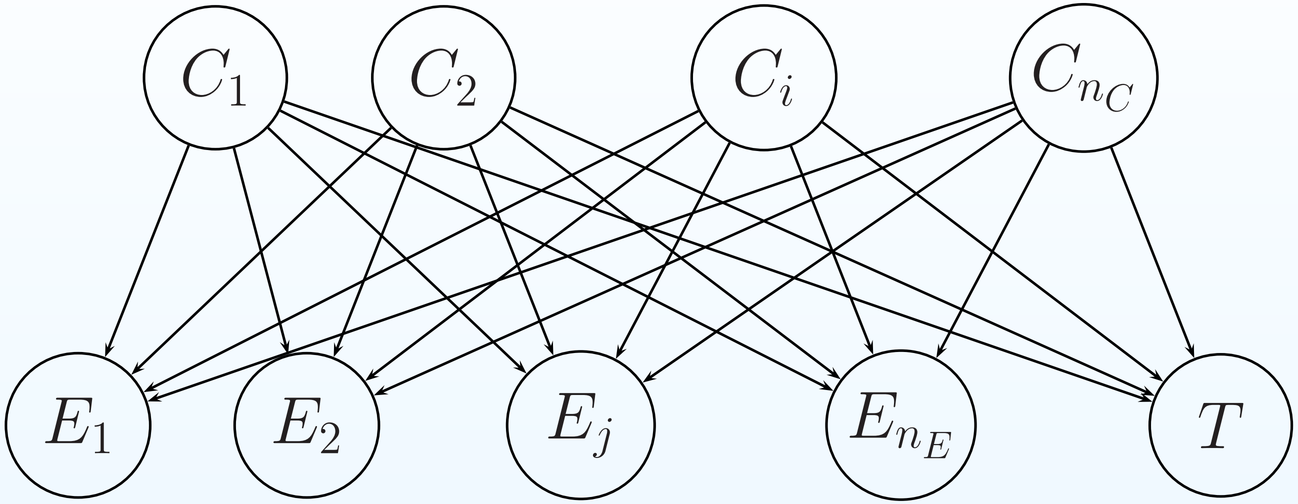


(T: ‘trash’)

- 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')

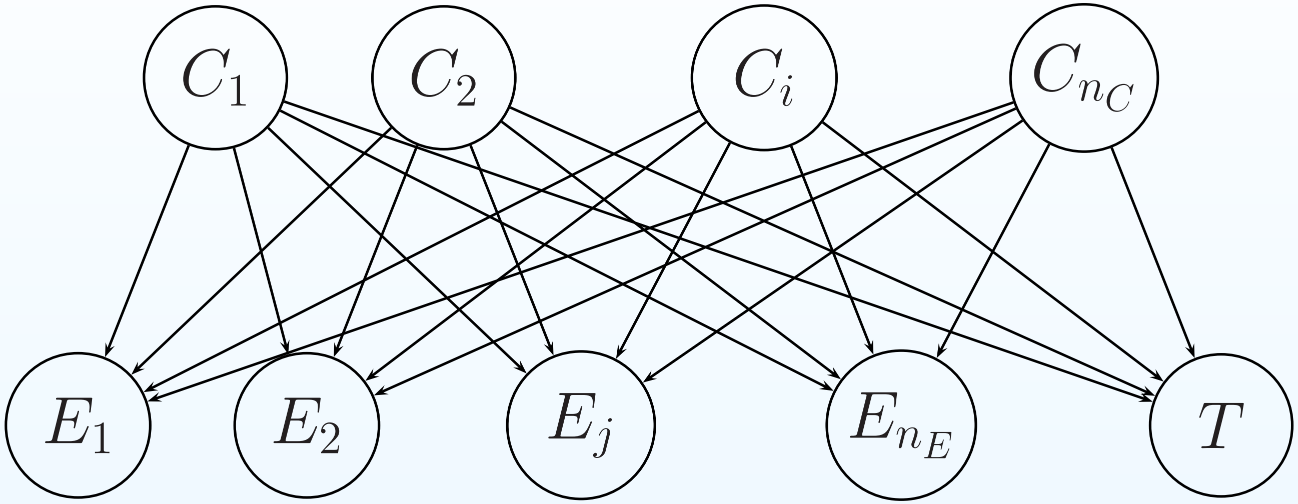
- Bayes theorem:

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

- Indifference w.r.t. all possible spectra

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

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



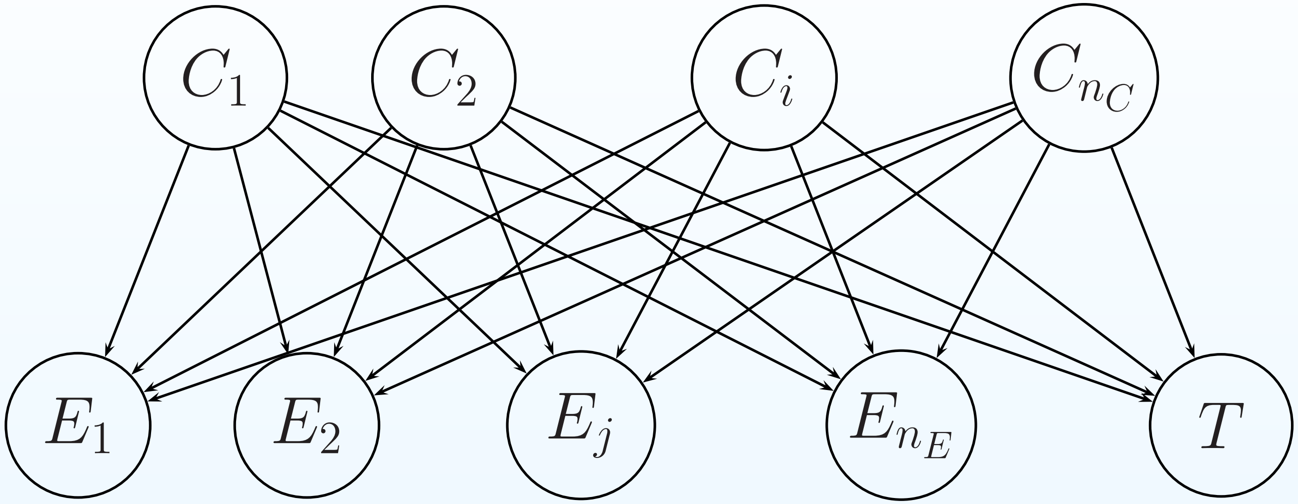
Given a certain number of events in a cause-bin $x(C_i)$, the number of events in the effect-bins, included the 'trash' one, is described by a multinomial distribution:

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

with

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

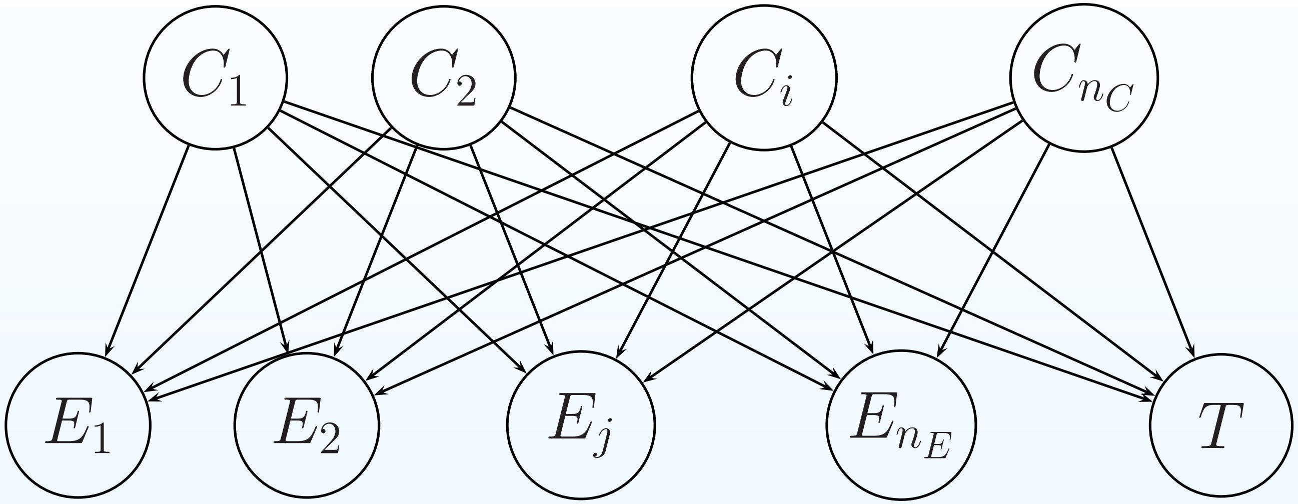
$P(\boldsymbol{x}_E \,|\, \boldsymbol{x}_C,\, \Lambda,\, I)$



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

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

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



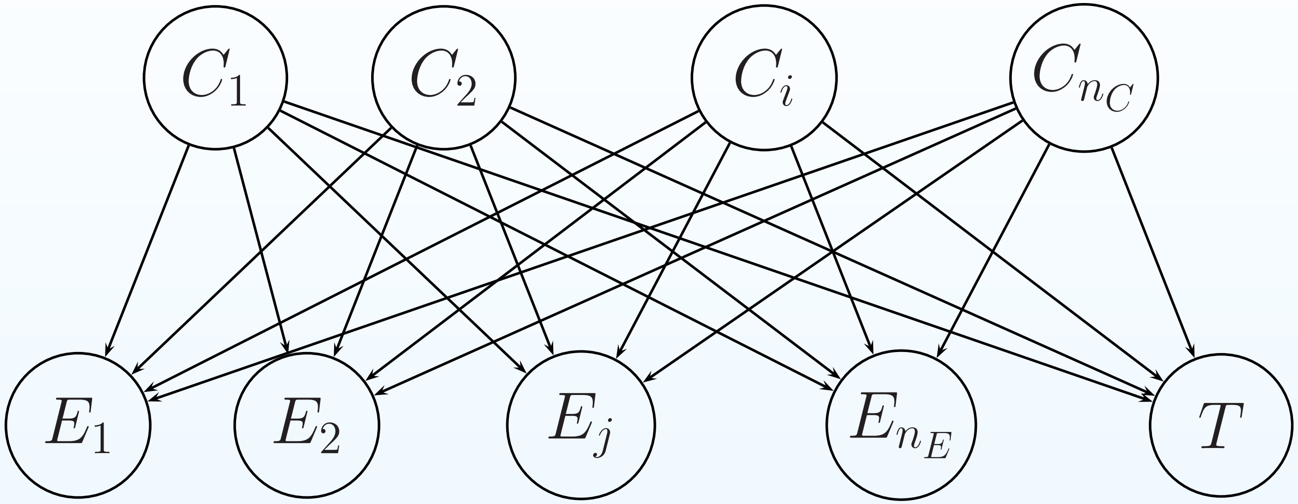
$\boldsymbol{x}_E|_{x(C_i)}$ multinomial random vector,

$\Rightarrow \boldsymbol{x}_E|_{\boldsymbol{x}(C)}$ 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,

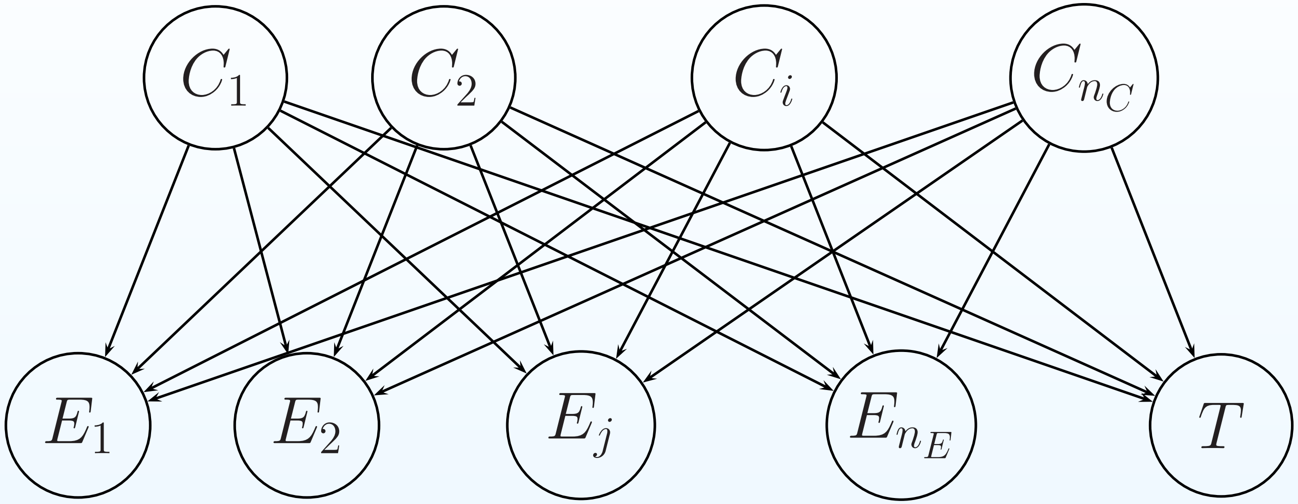
$\Rightarrow \boldsymbol{x}_E|_{\boldsymbol{x}(C)}$ sum of several multinomials.

BUT

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

$\Rightarrow$ STUCK!

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



$\boldsymbol{x}_E|_{x(C_i)}$ 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!

$\Rightarrow$ Change strategy

The rescue trick

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

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

we restart from

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

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.

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

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

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. [*] 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 $\epsilon_i = \sum_{j=1}^{n_E} P(E_j\,|\,C_i,\,I)$

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:

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

(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 $x_{E_j} \to \boldsymbol{x}_C$: done by a Multinomial:

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

4. $x(E_j) \to \mu_j$: 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. $\boldsymbol{\lambda}_i$:

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

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. $x(E_j) \rightarrow \mu_j$:

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

   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

(⇒ 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

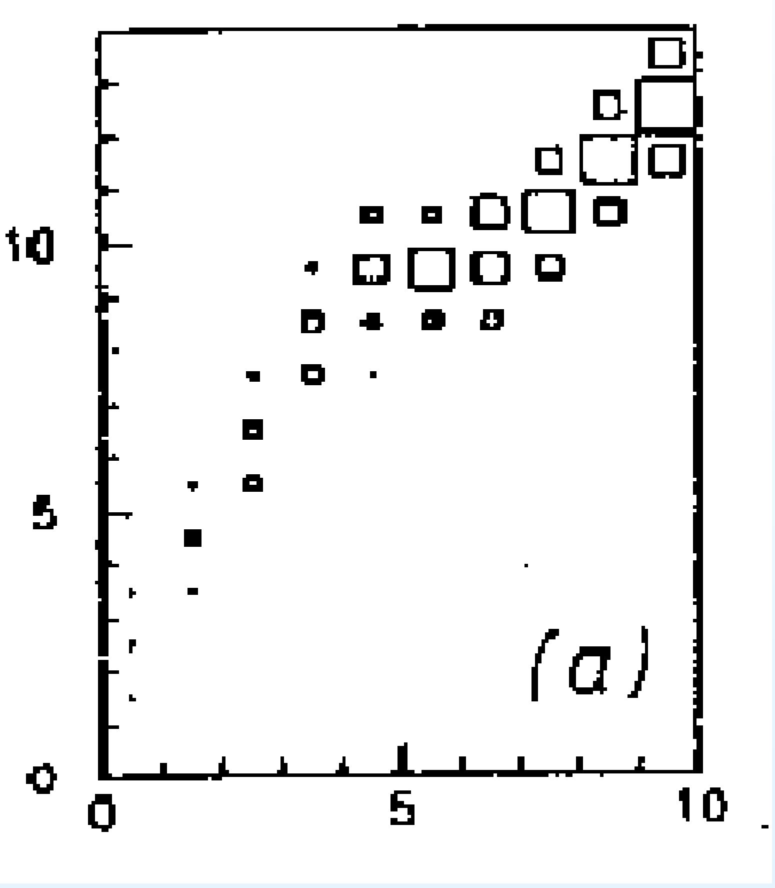
  $\Rightarrow$ Good compromize and good results

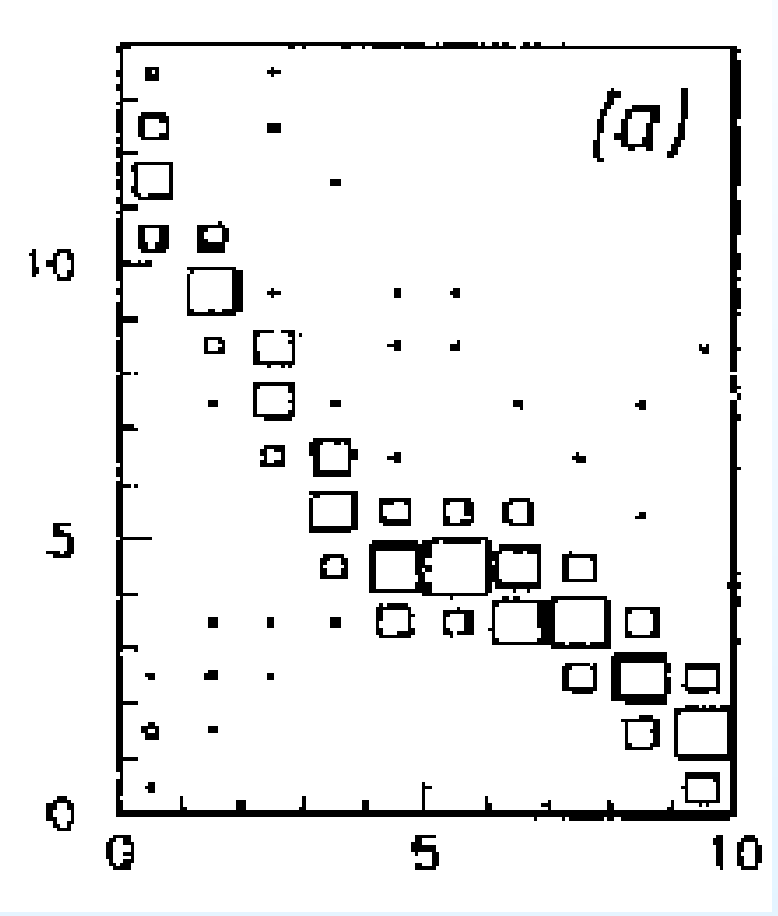
  $\Rightarrow$ Very 'Bayesian'

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

Examples

smearing matrix (from 1995 NIM paper)

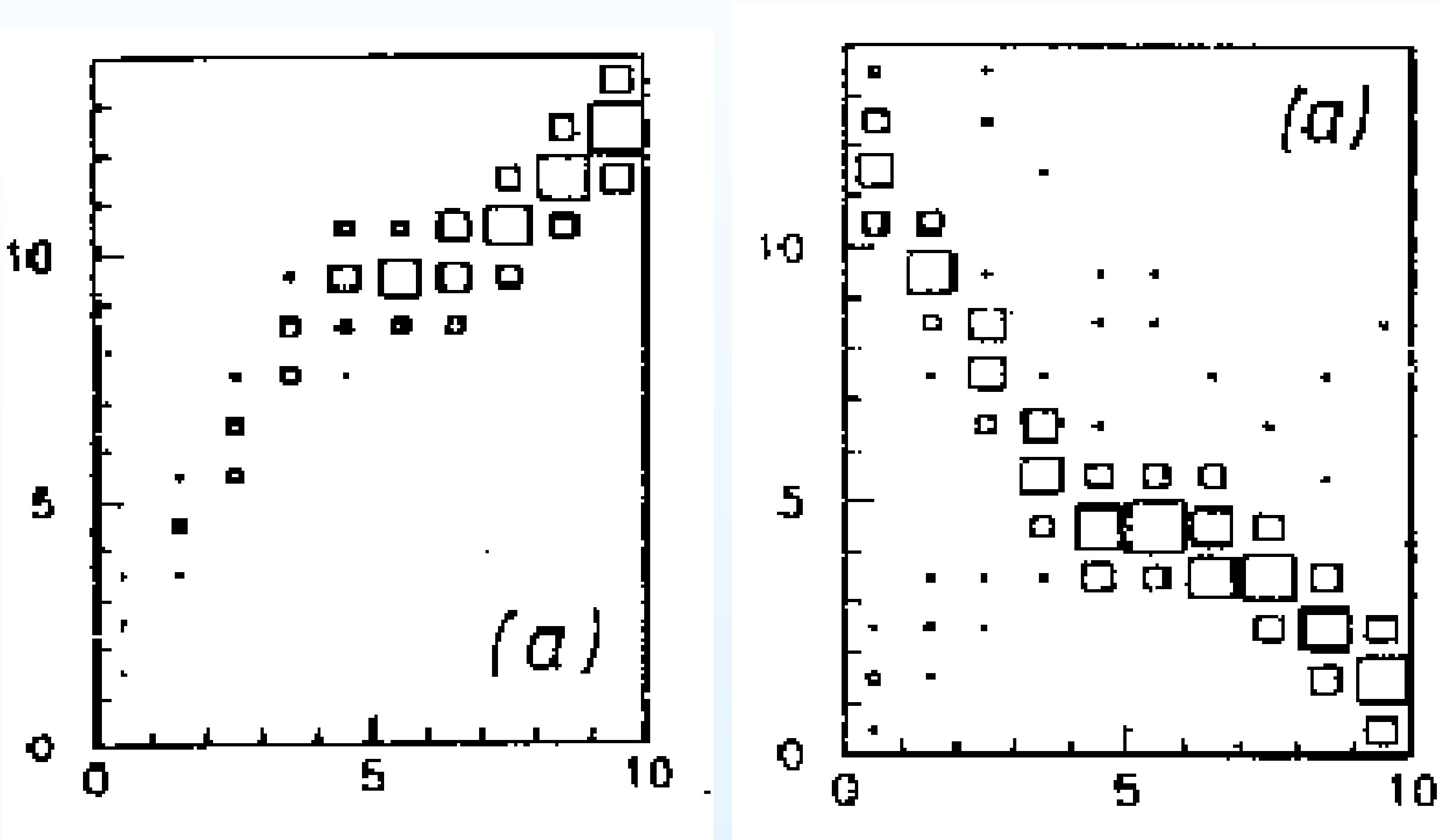




quite bad! (real cases are usually more gentle)

Examples

smearing matrix (from 1995 NIM paper)



quite bad! (real cases are usually more gentle)

$\Rightarrow$ watch DEMO

Conclusions

In general:

- A probabilistic approach ('Bayesian') offers a consistent framework to handle consistently a large variaty of problem
- Easy to use (at least conceptually), unless you have some ideological biases against it.

Conclusions

In general:

- A **probabilistic approach** ('Bayesian') offers a consistent framework to handle consistently a large variaty of problem
- Easy to use (at least conceptually), unless you have some ideological biases against it.

Concerning unfolding:

- **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

In general:

- A probabilistic approach ('Bayesian') offers a consistent framework to handle consistently a large variaty of problem
- Easy to use (at least conceptually), unless you have some ideological biases against it.

Concerning unfolding:

- 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

In general:

- A probabilistic approach ('Bayesian') offers a consistent framework to handle consistently a large variaty of problem
- Easy to use (at least conceptually), unless you have some ideological biases against it.

Concerning unfolding:

- 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

In general:

- A probabilistic approach ('Bayesian') offers a consistent framework to handle consistently a large variaty of problem
- Easy to use (at least conceptually), unless you have some ideological biases against it.

Concerning unfolding:

- 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

Extra references (including on yesterday comments) $\Longrightarrow$

References

['BR' stands for "GdA, Bayesian Reasoning in Data Analysis"]

- new unfolding: arXiv:1010.0632v1;
- for a multilevel introduction to probabilistic reasoning, including a short introduction to Bayesian networks: arXiv:1003.2086v2;
- ISO sources of uncertainties: BR, sec. 1.2;
- on uncertainties due to systematics: BR, secs. 6.8-6.10, 8.6-8.14, 12.2.2;
- 'asymmetric errors' and their potential dangers: physics/0403086;
- about the Gauss' derivation of the 'Gaussian': BR, 6.12; web site on "Fermi, Bayes and Gauss"
- box and ball 'game': AJP 67, issue 12 (1999) 1260-1268;

References

- upper/lower limits Vs sensitivity bounds: BR, secs. 13.16-13.18;
- fits from a Bayesian network perpective: physics/0511182;
- criticisms about 'tests': BR, 1.8;
- . . . but why "do they often work?": BR, 10.8;
- on the reason why 'standard' confidence intervals and confidence levels do not tell how much we are confident on something: BR, 1.7; arXiv:physics/0605140v2 (see also talk by A. Caldwell);
- on how to subtract the expected background in a probabilistics way: BR, 7.7.5;
- for a nice introduction to MCMC: C. Andrieu at al. "An introduction to MCMC for Machine Learning", downloadable pdf.