



REFLECTIONS ON ARMAX SYSTEMS

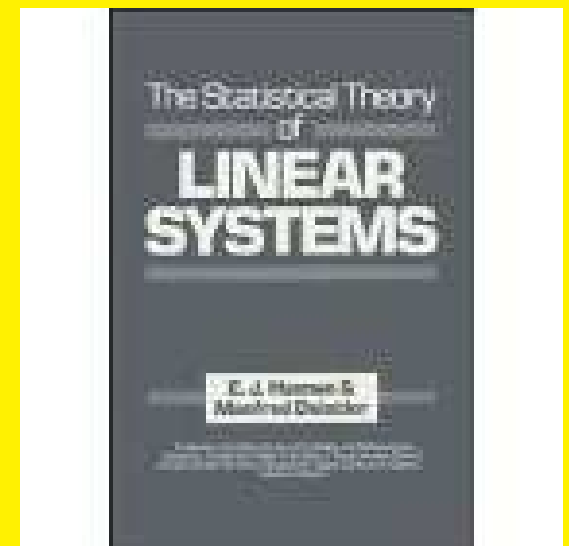
JAN C. WILLEMS
K.U. Leuven, Flanders, Belgium

Conference on Econometrics, Time Series Analysis and Systems Theory Vienna, June 18, 2009



In honor of [Manfred Deistler](#) on the occasion of his retirement

ARMAX



ARMAX systems

$$\begin{aligned} A_0 y(t) + A_1 y(t+1) + \dots + A_{L_1} y(t+L_1) \\ = X_0 u(t) + X_1 u(t+1) + \dots + X_{L_2} u(t+L_2) \\ + M_0 \varepsilon(t) + M_1 \varepsilon(t+1) + \dots + M_{L_3} \varepsilon(t+L_3) \end{aligned}$$

ARMAX systems

$$\begin{aligned} A_0 y(t) + A_1 y(t+1) + \dots + A_{L_1} y(t+L_1) \\ = X_0 u(t) + X_1 u(t+1) + \dots + X_{L_2} u(t+L_2) \\ + M_0 \varepsilon(t) + M_1 \varepsilon(t+1) + \dots + M_{L_3} \varepsilon(t+L_3) \end{aligned}$$

$$A(\sigma)y = X(\sigma)u + M(\sigma)\varepsilon$$

$\sigma =$ the **shift**, $\sigma f(t) := f(t+1)$

A, X, M : **real polynomial matrices**

ARMAX systems

$$A(\sigma)y = X(\sigma)u + M(\sigma)\varepsilon$$

$\sigma =$ the **shift**, $\sigma f(t) := f(t + 1)$

A, X, M : real polynomial matrices

$y, u : \mathbb{Z} \rightarrow \mathbb{R}^p, \mathbb{R}^m$, u input, y output

the variables whose dynamic relation is modeled

$\varepsilon : \mathbb{Z} \rightarrow \mathbb{R}^\ell$ disturbances, **'noise'**

A: **A**uto**R**egressive-part

M: **M**oving **A**verage-part

X: **E**Xogenous-part

Equivalent model class

$$\sigma x = Ax + Bu + G\varepsilon, y = Cx + Du + J\varepsilon$$

σ = the **shift**, $\sigma f(t) := f(t + 1)$

$y, u : \mathbb{Z} \rightarrow \mathbb{R}^p, \mathbb{R}^m$ u **input**, y **output**

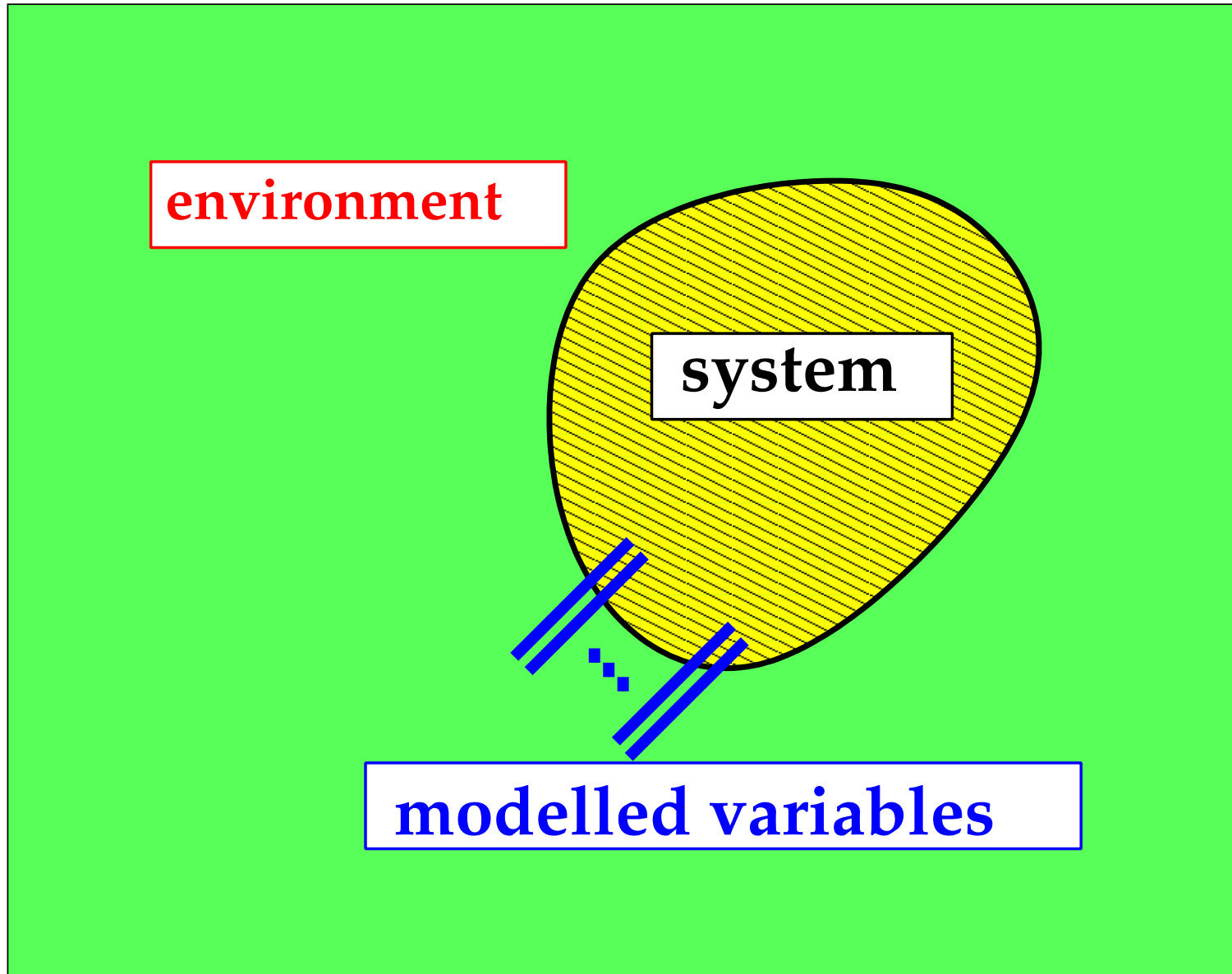
the variables whose dynamic behavior is modeled

$\varepsilon : \mathbb{Z} \rightarrow \mathbb{R}^\ell$ **disturbance, 'noise'**

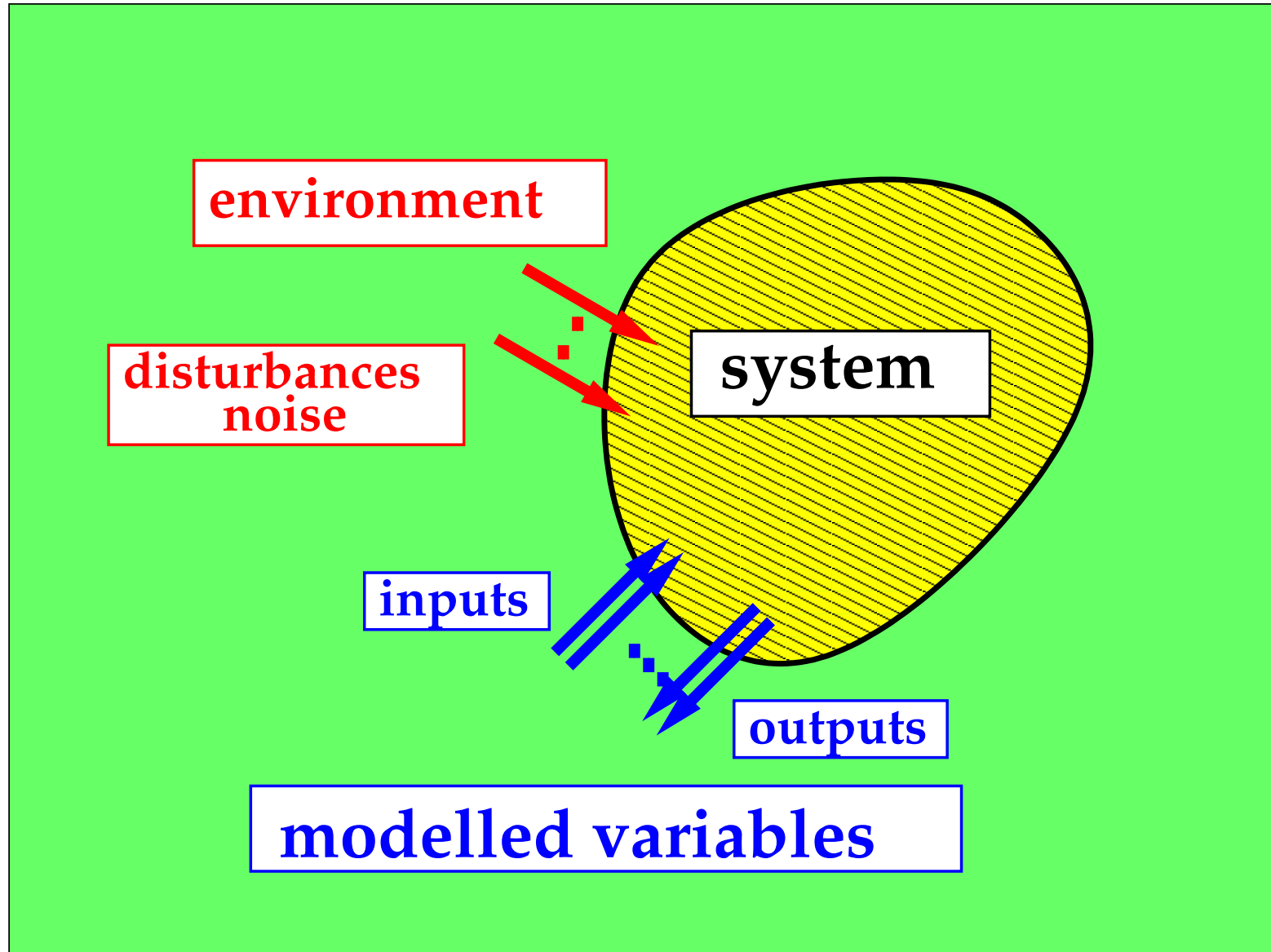
$x : \mathbb{Z} \rightarrow \mathbb{R}^n$ **auxiliary state variables**

A, B, C, G, D, J : **real matrices**

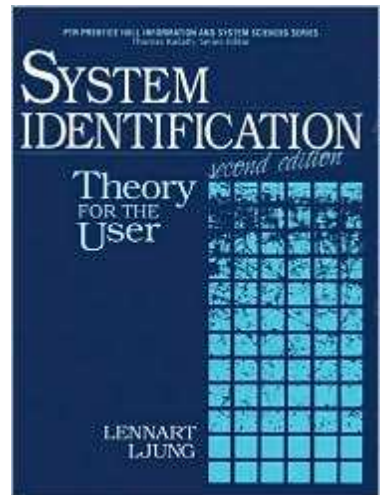
Modeling idea



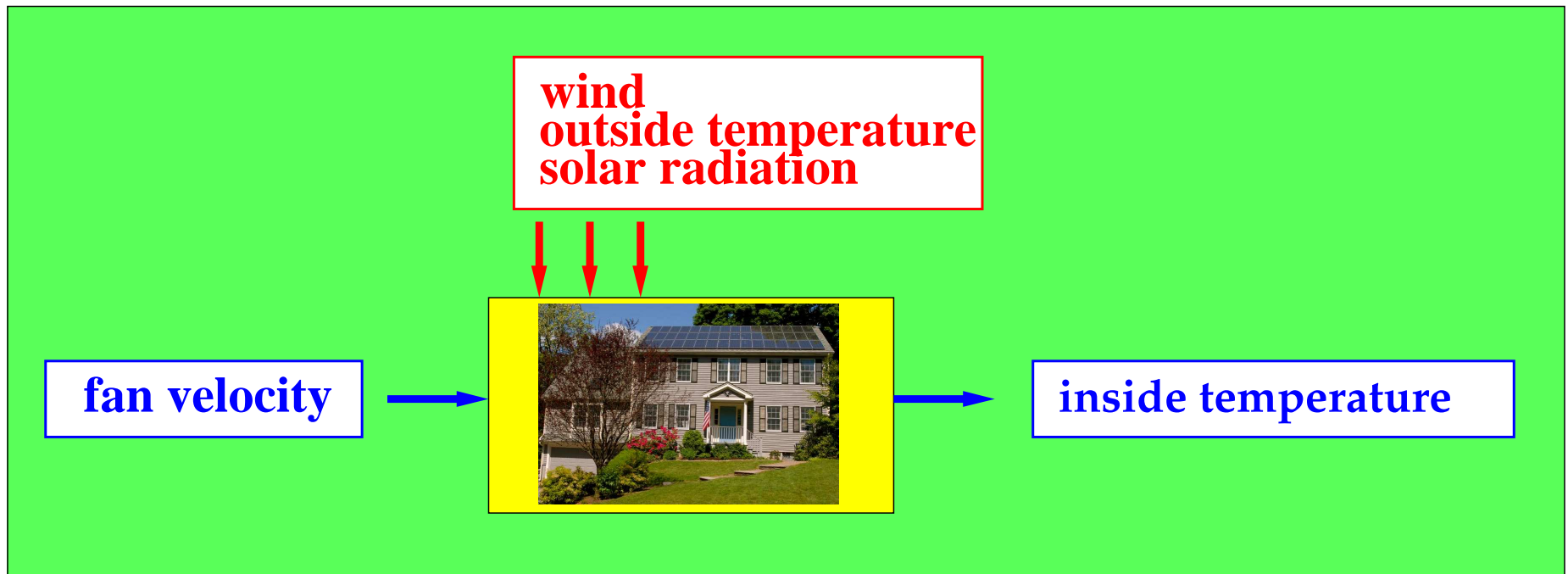
Modeling idea



Modeling idea

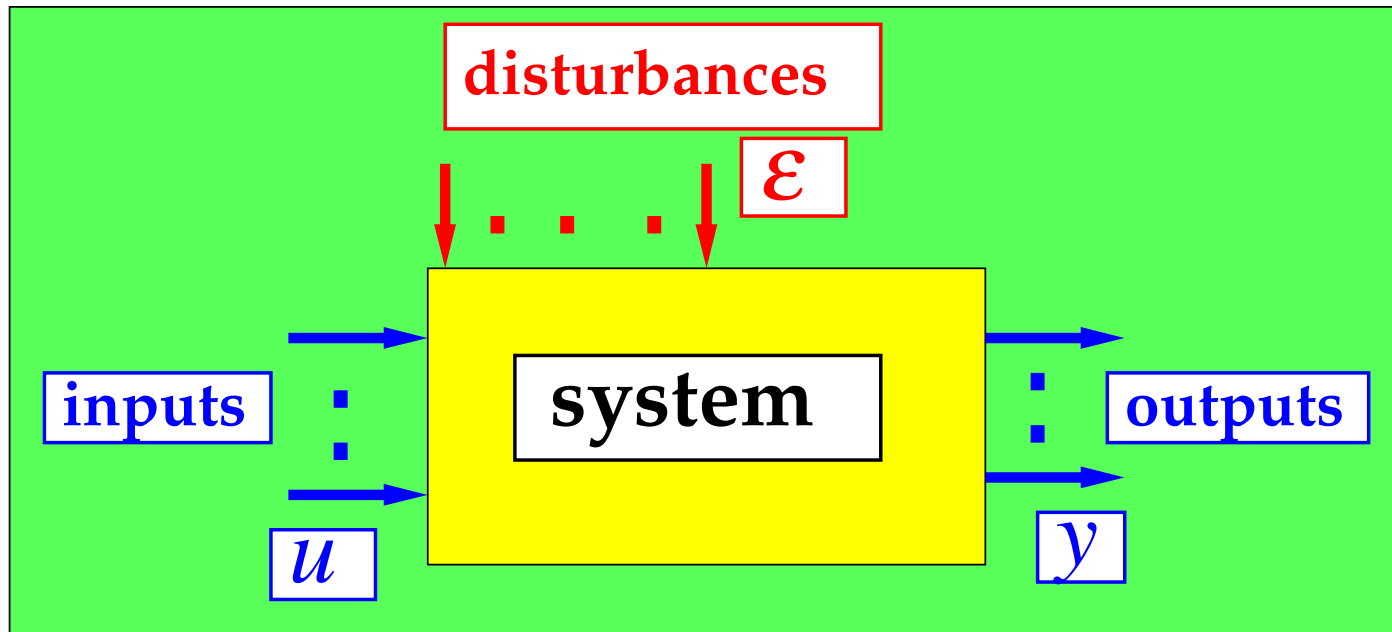


Example

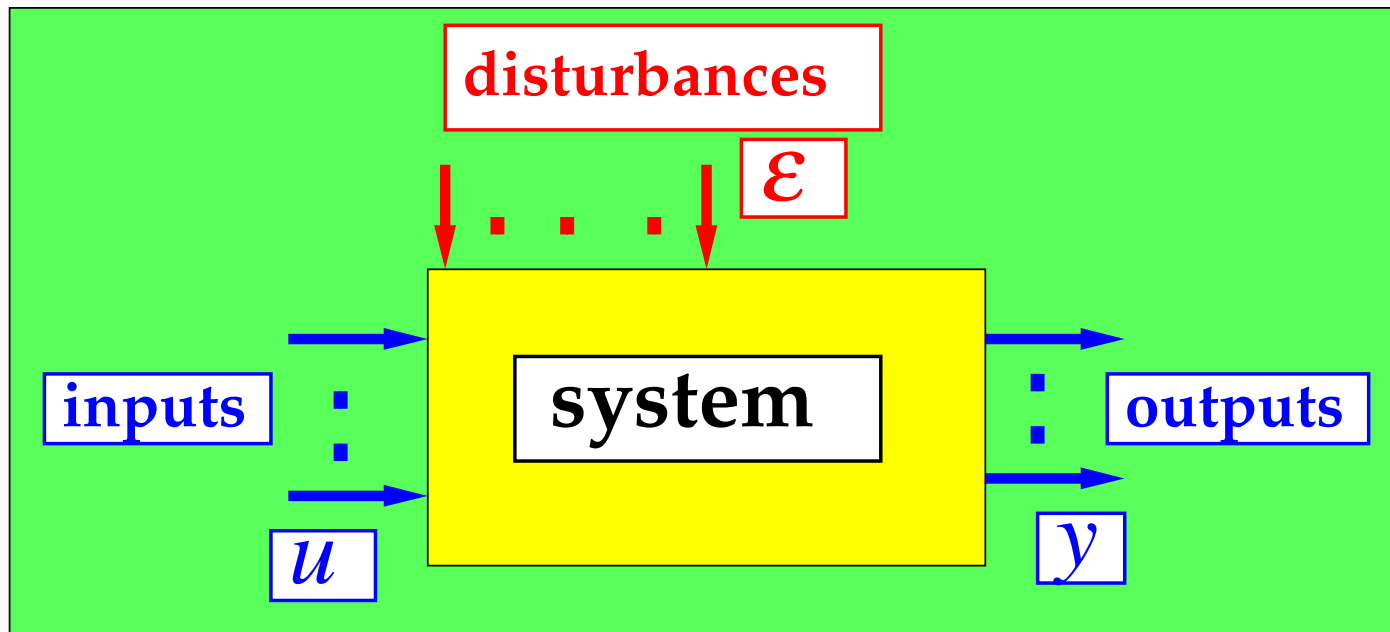


Inertia \rightsquigarrow difference equation with lags \Rightarrow ARMAX

Modeling idea



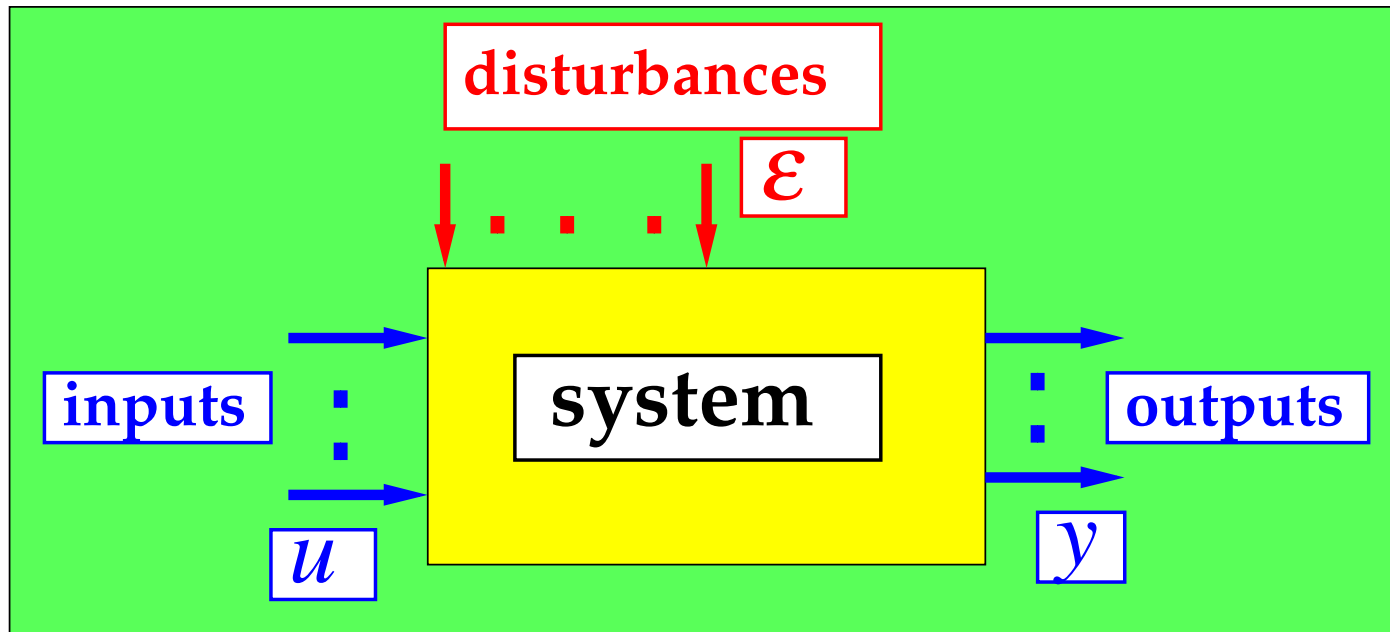
Modeling idea



Typical assumptions:

- ▶ ε a stationary stochastic (vector) process
- ▶ u a stochastic process, typically independent of ε
- ▶ suitable assumptions on A, M, X
- ▶ $\Rightarrow y$ stochastic process

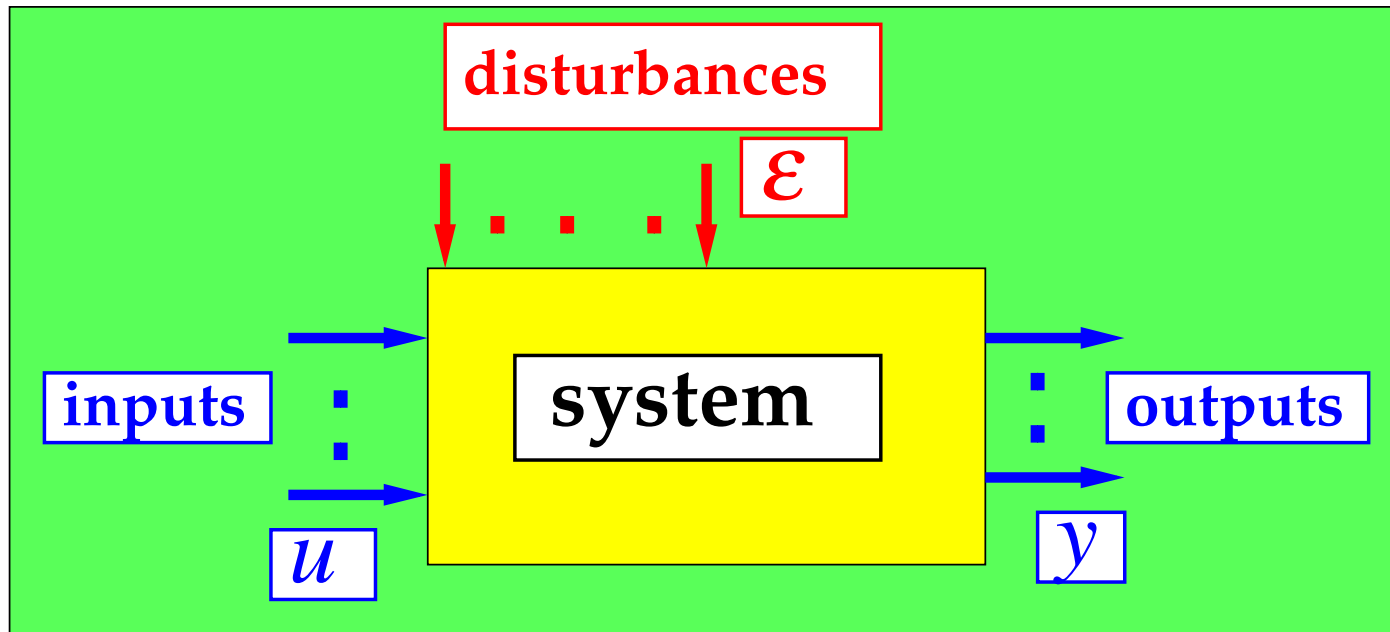
Modeling idea



Reflections:

- ▶ the separation of system variables
into **inputs** u and **outputs** y
- ▶ the stochastic nature of disturbance inputs ε
- ▶ the input nature of external disturbances

Modeling idea

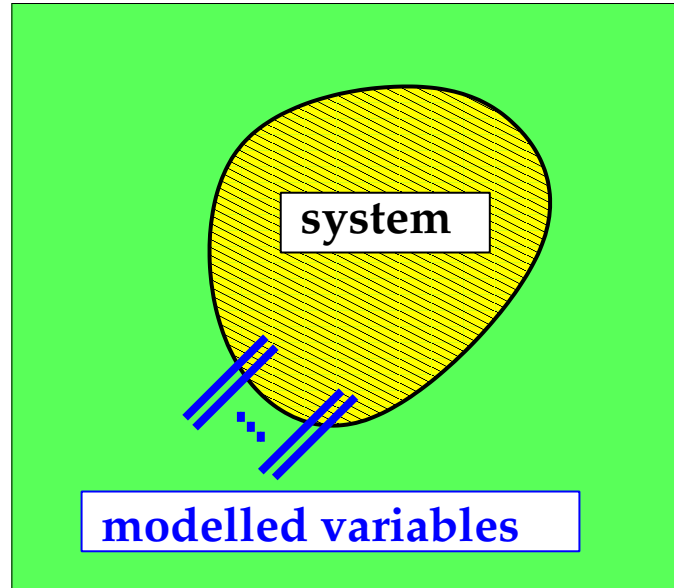


Reflections:

- ▶ **separation of system variables**
into **inputs u** and **outputs y**
- ▶ the **stochastic nature** of disturbance inputs ε
- ▶ the **input nature of external disturbances**

INPUTS and OUTPUTS

Closed systems



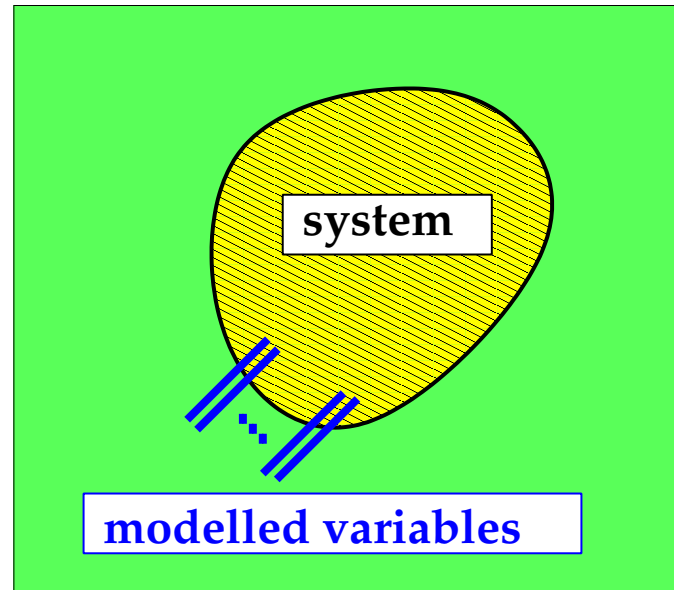
If the system variables are completely generated ‘internally’, we speak of **closed systems**.

Deterministic case: $x(t+1) = f(x(t))$ or $\frac{d}{dt}x = f(x)$, $w = h(x)$.

Stochastic case: $x(t+1) = f(x(t), \varepsilon(t))$,
or $dx = f(x) dt + h(x) d\varepsilon$, $w = h(x)$.

ε : internal noise

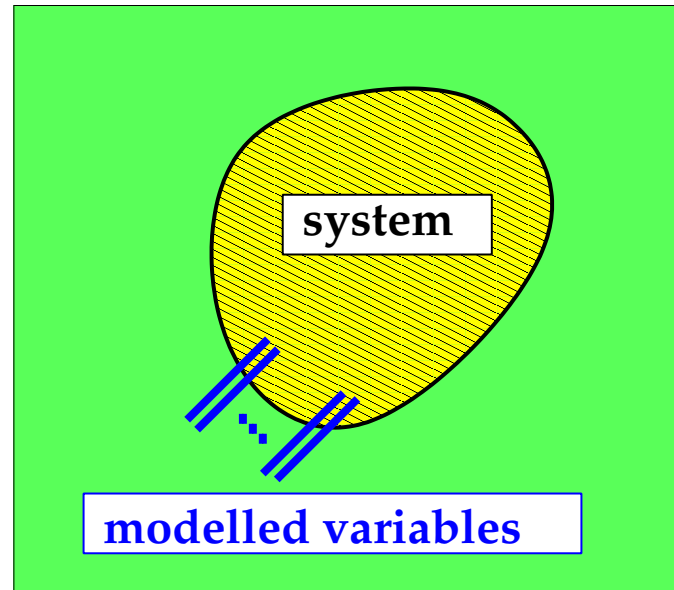
Closed systems



But closed systems do not form a good model class:

- ▶ **they do not cope with interconnection, with tearing**
- ▶ **the basic laws of physics are not closed systems**
- ▶ **implicitly forces to model the environment**

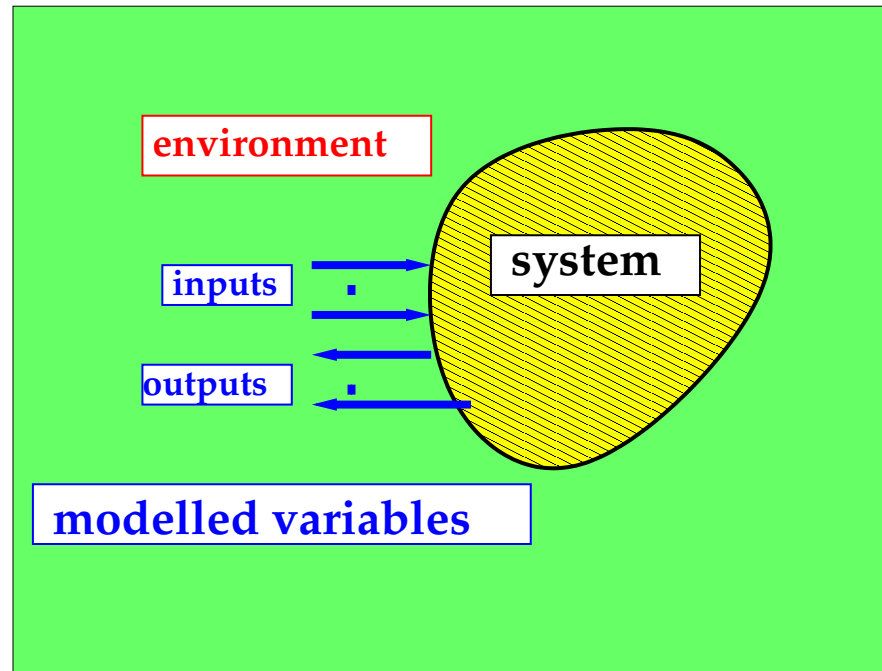
Closed systems



How to model interaction with the environment?

Open systems

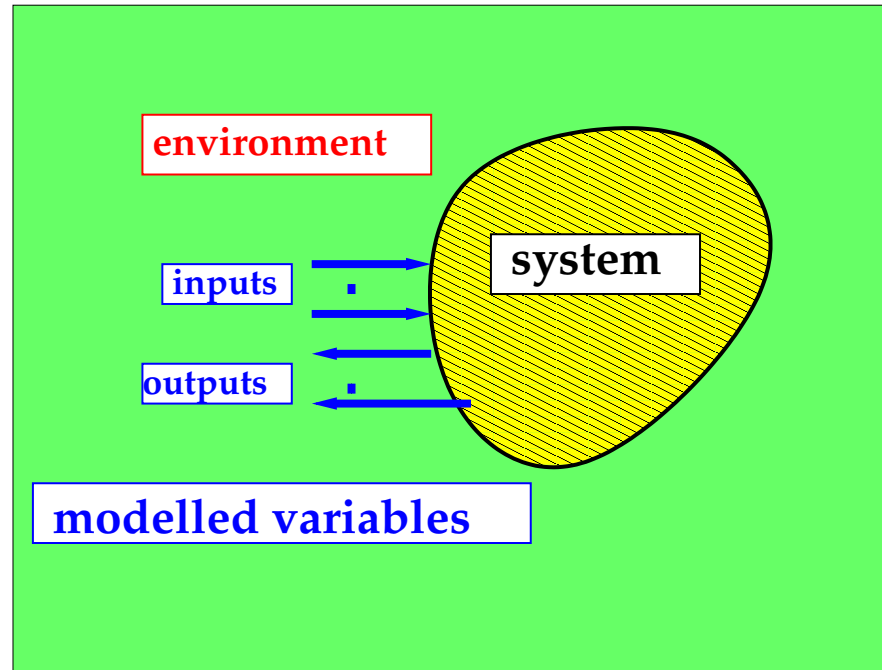
Classical approach:



$\rightsquigarrow x(t+1) = f(x(t), u(t)), y(t) = h(x(t), u(t)), w = (u, y)$, **or**
 $x(t+1) = f(x(t), u(t), \varepsilon(t)), y(t) = h(x(t), u(t), \varepsilon(t)), w = (u, y)$,
or transfer functions, or ARMAX systems,...

Open systems

Classical approach:



$\rightsquigarrow x(t+1) = f(x(t), u(t)), y(t) = h(x(t), u(t)), w = (u, y)$, **or**
 $x(t+1) = f(x(t), u(t), \varepsilon(t)), y(t) = h(x(t), u(t), \varepsilon(t)), w = (u, y)$,
or transfer functions, or ARMAX systems,...

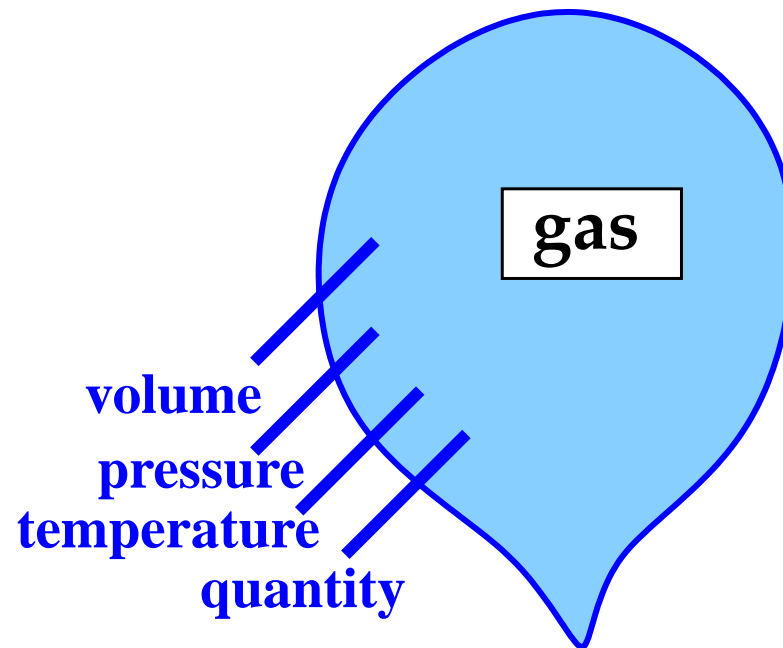
Does this input/output partition respect the physics?

Input/output thinking

The input/output view as the **primary and universal** concept for open systems is a **misconception**
It fails in the first examples.

Input/output thinking

The input/output view as the primary and universal concept for open systems is a misconception. It fails in the first examples.

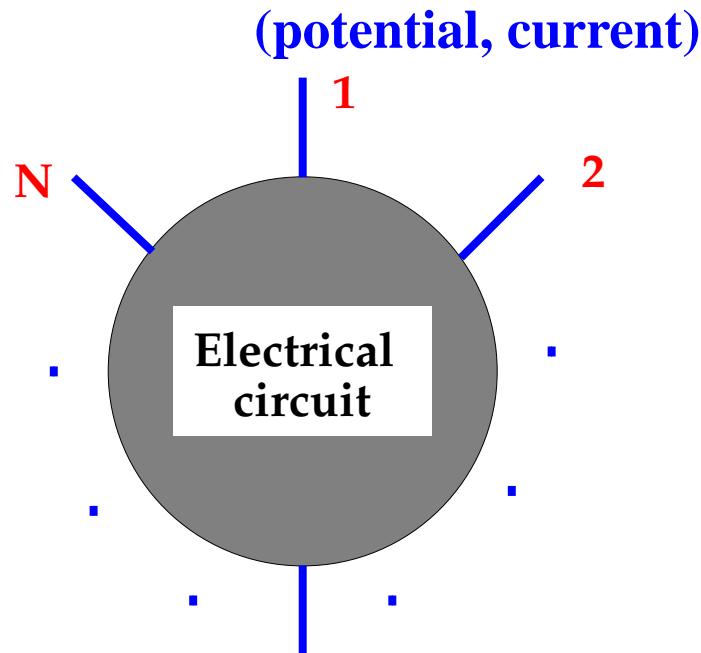


The gas law imposes the relation on $PV = NT$.

It makes no sense to view this in an input/output way.

Input/output thinking

The input/output view as the primary and universal concept for open systems is a misconception. It fails in the first examples.



The circuit imposes a relation on

$$V_1, I_1, V_2, I_2, \dots, V_N, I_N$$

Only after modeling \Rightarrow voltage or current driven terminals.

Input/output thinking

Maxwell's equations



$$\begin{aligned}\nabla \cdot \vec{E} &= \frac{1}{\epsilon_0} \rho, \\ \nabla \times \vec{E} &= -\frac{\partial}{\partial t} \vec{B}, \\ \nabla \cdot \vec{B} &= 0, \\ c^2 \nabla \times \vec{B} &= \frac{1}{\epsilon_0} \vec{j} + \frac{\partial}{\partial t} \vec{E}.\end{aligned}$$

10 variables, 8 equations, $\Rightarrow \exists$ free variables.

But it makes no sense to declare some variables as inputs...

Input/output thinking

The input/output view as the primary and universal concept for open systems is a **misconception**
It fails in the first examples.

The strongest argument against input/output thinking comes from **system interconnection**
****variable sharing** not **output-to-input assignment****
is the mechanism to interconnect systems.

BEHAVIORS

Behavioral systems - deterministic case

A (static) model is a subset \mathcal{B} of the universum \mathcal{U} of possible outcomes of a phenomenon.

\mathcal{B} is the behavior of the model.

A **dynamical system** $:\Leftrightarrow (\mathbb{T}, \mathbb{W}, \mathcal{B})$, with

$\mathbb{T} \subseteq \mathbb{R}$ the **time set**

\mathbb{W} the **signal space**

$\mathcal{B} \subseteq \mathbb{W}^{\mathbb{T}}$ the **behavior**

Behavioral systems - deterministic case

A **dynamical system** $:\Leftrightarrow (\mathbb{T}, \mathbb{W}, \mathcal{B})$, with

$\mathbb{T} \subseteq \mathbb{R}$ the **time set**

\mathbb{W} the **signal space**

$\mathcal{B} \subseteq \mathbb{W}^{\mathbb{T}}$ the **behavior**

So, a dynamical system is merely a family of time-trajectories taking values in a signal space.

If $\mathbb{W} = \mathbb{R}^w$, then all variables are treated on the same level. When analyzing \mathcal{B} , some components of $w \in \mathcal{B}$ may be ‘free’, in a sense ‘inputs’.

Behavioral systems - deterministic case

A **dynamical system** $:\Leftrightarrow (\mathbb{T}, \mathbb{W}, \mathcal{B})$, with

$\mathbb{T} \subseteq \mathbb{R}$ the **time set**

\mathbb{W} the **signal space**

$\mathcal{B} \subseteq \mathbb{W}^{\mathbb{T}}$ the **behavior**

A rich theory has been developed in this deterministic case, featuring new viewpoints, e.g. about LTIDSs, about controllability, etc.

Linear time-invariant difference systems

The dynamical system $\Sigma = (\mathbb{Z}, \mathbb{R}^w, \mathcal{B})$ is said to be

▶ **linear** $:\Leftrightarrow \mathcal{B} \subseteq (\mathbb{R}^w)^\mathbb{Z}$ is linear

▶ **time-invariant** $:\Leftrightarrow \mathcal{B} = \sigma \mathcal{B}$

▶ **complete** $:\Leftrightarrow$

$[[w \in \mathcal{B}]] \Leftrightarrow [[w|_{[t_1, t_2]} \in \mathcal{B}|_{[t_1, t_2]} \text{ for all } t_1, t_2 \in \mathbb{Z}]]$

Linear time-invariant difference systems

The following are equivalent for $\Sigma = (\mathbb{Z}, \mathbb{R}^w, \mathcal{B})$

- ▶ Σ is linear, time-invariant, **complete**
- ▶ $\mathcal{B} \subseteq (\mathbb{R}^w)^{\mathbb{Z}}$ linear, shift-invariant, and **closed**
- ▶ \exists a polynomial matrix $R \in \mathbb{R}^{\bullet \times w}[\xi]$ such that

$$\mathcal{B} = \{w : \mathbb{Z} \rightarrow \mathbb{R}^w \mid R(\sigma)w = 0\}$$

that is, \mathcal{B} is the solution set of

$$R_0 w(t) + R_1 w(t+1) + \cdots + R_L w(t+L) = 0 \text{ for all } t \in \mathbb{Z}$$

‘kernel representation’

Linear time-invariant difference systems

The following are equivalent for $\Sigma = (\mathbb{Z}, \mathbb{R}^w, \mathcal{B})$

▶ \mathcal{B} is the solution set of

$$R_0 w(t) + R_1 w(t+1) + \cdots + R_L w(t+L) = 0 \text{ for all } t \in \mathbb{Z}$$

‘kernel representation’

▶ R' and R'' define the same system

\Leftrightarrow their rows generate
the same $\mathbb{R}[\xi, \xi^{-1}]$ -module

Linear time-invariant difference systems

The following are equivalent for $\Sigma = (\mathbb{Z}, \mathbb{R}^w, \mathcal{B})$

▶ \mathcal{B} is the solution set of

$$R_0 w(t) + R_1 w(t+1) + \cdots + R_L w(t+L) = 0 \text{ for all } t \in \mathbb{Z}$$

‘kernel representation’

▶ R' and R'' define the same system

\Leftrightarrow their rows generate
the same $\mathbb{R}[\xi, \xi^{-1}]$ -module

▶ \exists one-to-one relation between

LTIDSs and $\mathbb{R}[\xi, \xi^{-1}]$ -modules

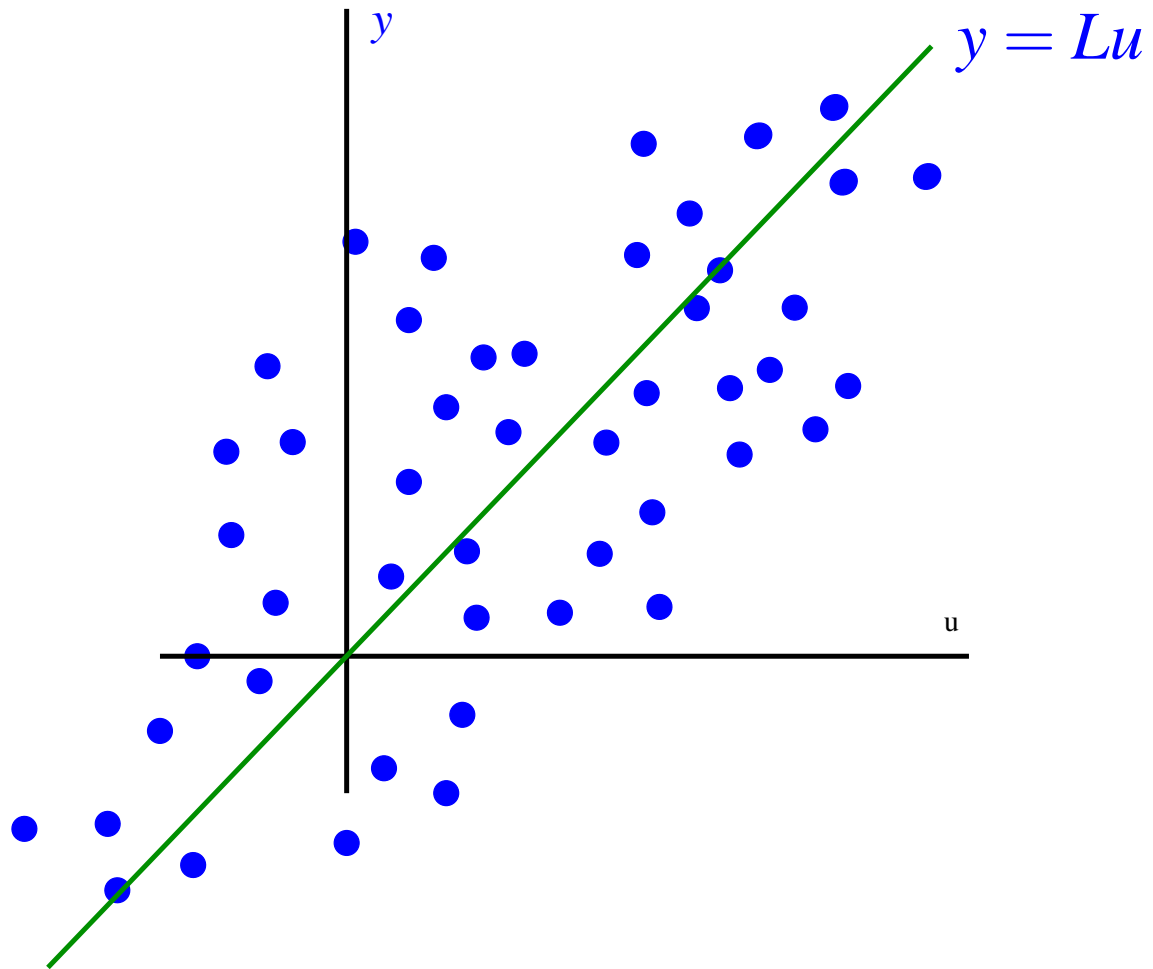
STOCHASTIC BEHAVIORS

STATIC CASE

Static case

‘Regression’

$$y = Lu + \varepsilon$$



Static case

‘Regression’

$$y = Lu + \varepsilon$$

ε models the uncertainty of the ‘law’

$$y = Lu$$

Classical: ε is a random vector.

But what should one assume about u ?

And about the relation between u and ε ?

Static case

‘Regression’

$$y = Lu + \varepsilon$$

Classical: ε is a random vector.

But what should one assume about u ?

And about the relation between u and ε ?

Since u is ‘external’, generated by the environment, one should not state anything about u .

Modeling a system should not require modeling the environment!

We also want to treat u and y on the same level

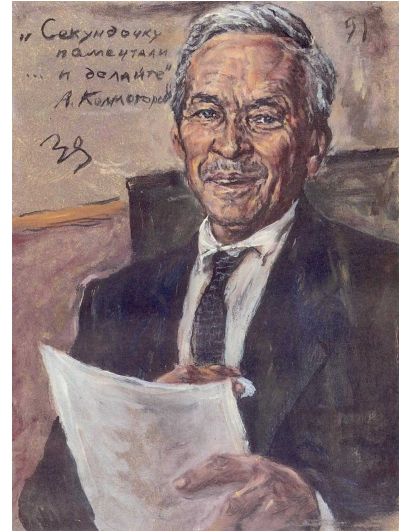
Stochastic static linear system

Recall the classical definition of an abstract random variable $(\mathbb{A}, \mathcal{A}, P)$ with

\mathbb{A} the space of elementary events

\mathcal{A} a sigma-algebra of subsets of \mathbb{A}

$P : \mathcal{A} \rightarrow [0, 1]$ a probability measure



In what is called an n -dimensional real random vector, we obtain $(\mathbb{R}^n, \mathcal{A}, P)$

with \mathcal{A} the sigma-algebra of Borel subsets of \mathbb{R}^n .

Our proposal is that (even for regression!), we should not take the Borel sigma-algebra.

Stochastic static linear system

Definition: A **stochastic static linear system** is a random variable

$$(\mathbb{R}^n, \mathcal{A}, P)$$

with \mathcal{A} the sigma-algebra of subsets of \mathbb{R}^n defined as follows in terms of a linear subspace $\mathbb{L} \subseteq \mathbb{R}^n$

$$\mathcal{A} = \{S \subseteq \mathbb{R}^n \mid S = S' + \mathbb{L}, S' \subseteq \mathbb{R}^n \text{ Borel}\}$$

Stochastic static linear system

Definition: A **stochastic static linear system** is a random variable

$$(\mathbb{R}^n, \mathcal{A}, P)$$

with \mathcal{A} the sigma-algebra of subsets of \mathbb{R}^n defined as follows in terms of a linear subspace $\mathbb{L} \subseteq \mathbb{R}^n$

$$\mathcal{A} = \{S \subseteq \mathbb{R}^n \mid S = S' + \mathbb{L}, S' \subseteq \mathbb{R}^n \text{ Borel}\}$$

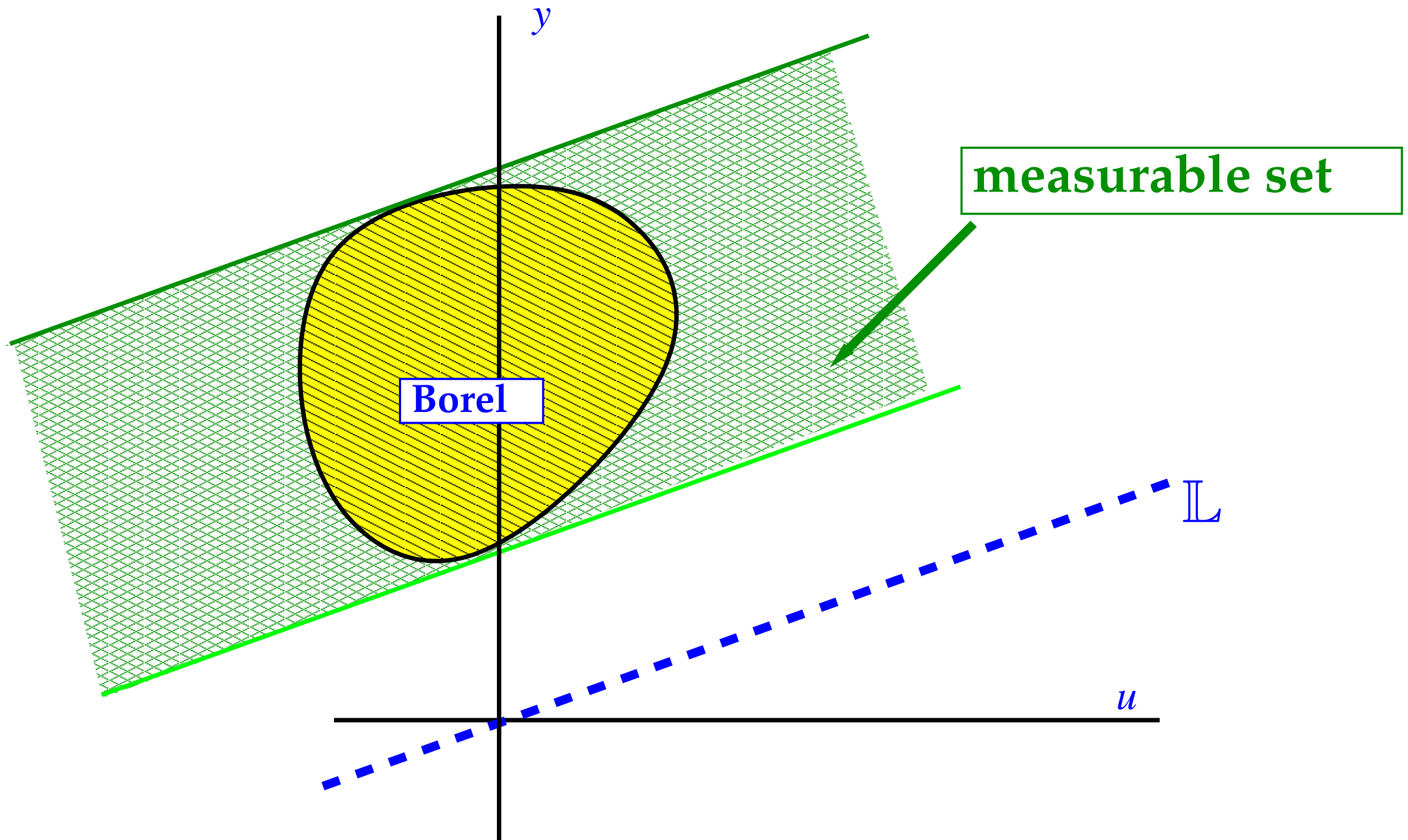
Special cases:

$$\mathbb{L} = \{0\} \quad \text{classical random vector}$$

$$P(\mathbb{L}) = 1 \quad \text{deterministic case}$$

In pictures

Sets for which the probability is defined:



Representation

A stochastic static linear system on \mathbb{R}^w admits a representation

$$Rw = \varepsilon$$

with R a real matrix and ε a classical real random vector.

Special cases:

$R = I \rightsquigarrow w = \varepsilon$ **classical random vector**

$\varepsilon = 0 \rightsquigarrow Rw = 0$ **deterministic system**

$\text{dimension}(\text{kernel}(R)) = \text{degrees of freedom.}$

Regression

- ▶ **Case $n = 2$. Def. says that $y - \alpha u$ is random but that u and y are NOT random variables (in the formal sense that the projections are not ‘measurable’ maps.)**
- ▶ **This is the intention of a regression model. There is no claim in such a model that u is random or deterministic, or that ε is dependent or independent of u or y .**

Examples

How do you weigh a cow?

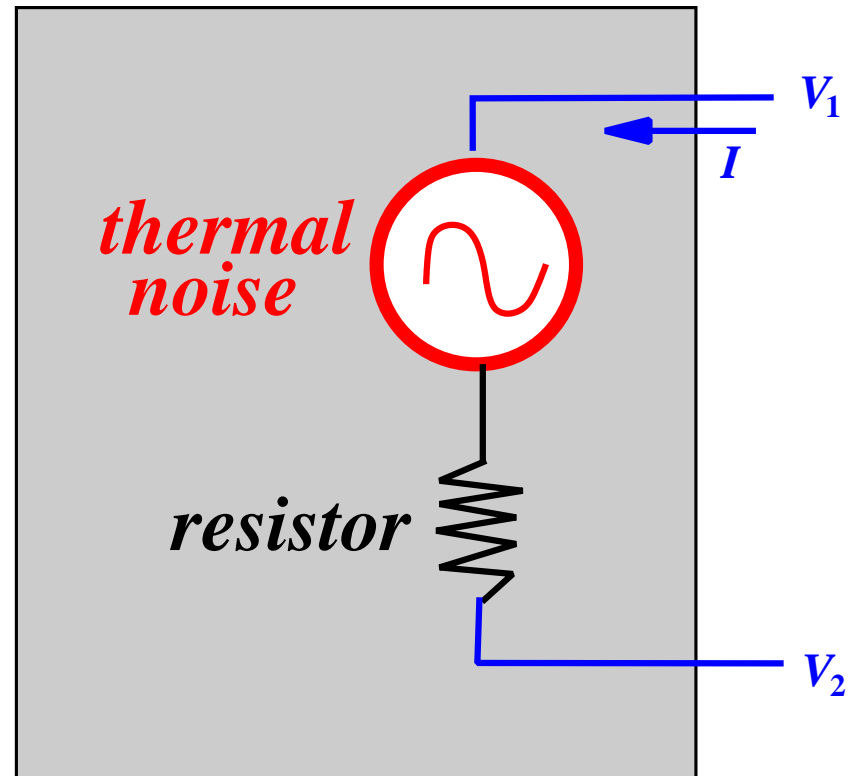


weight \propto circumference

is a random variable, not the weight or the circumference.

Examples

Johnson-Nyquist resistor noise

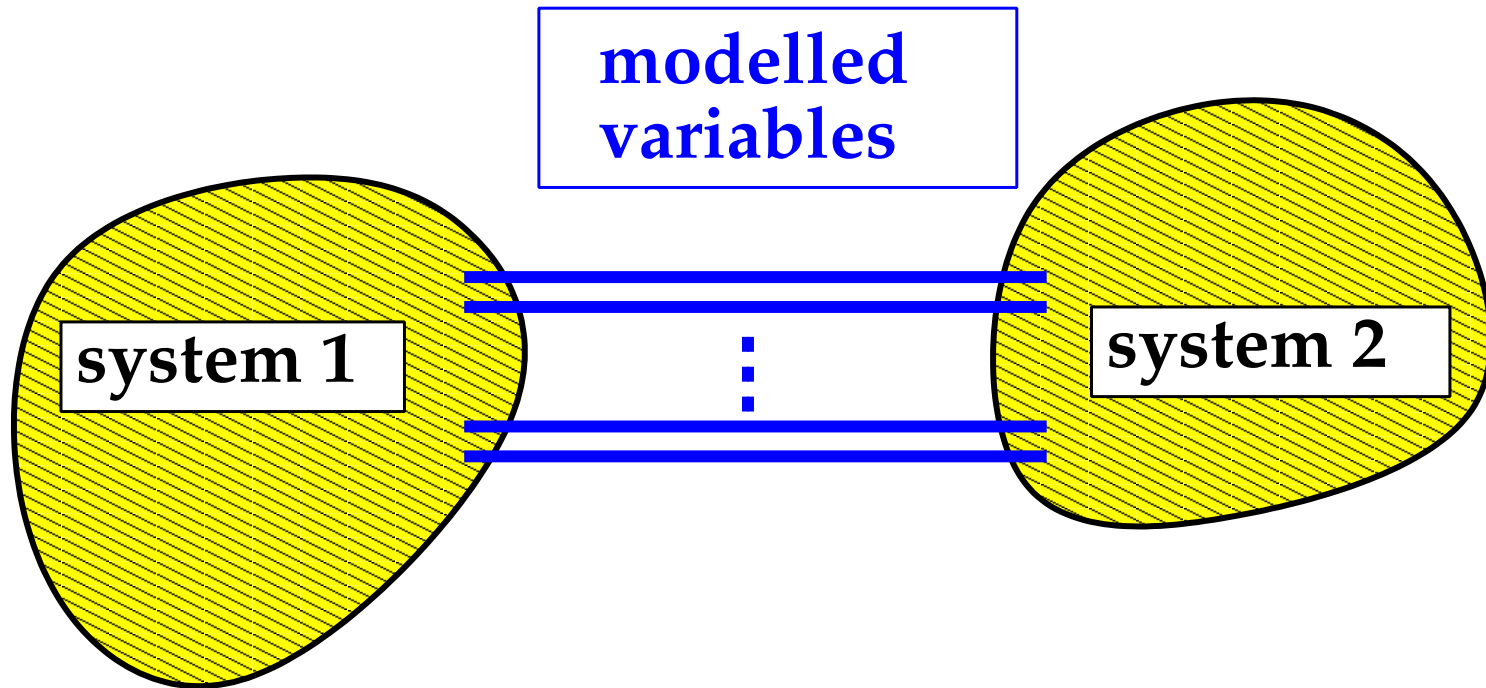


$$V_1 - V_2 - RI = V_{\text{noise}}$$

with V_{noise} a random variable

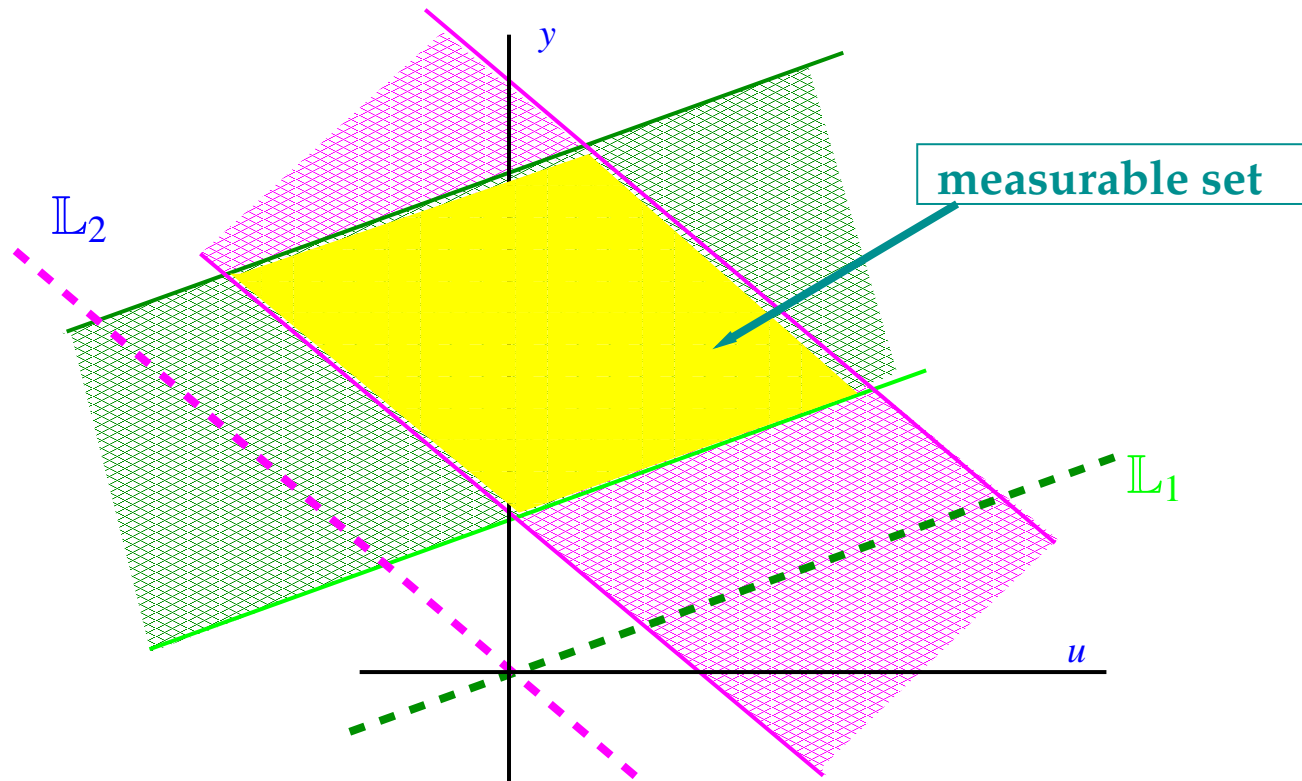
Interconnection

After **interconnection**, i.e., after modeling the environment, we obtain



Interconnection

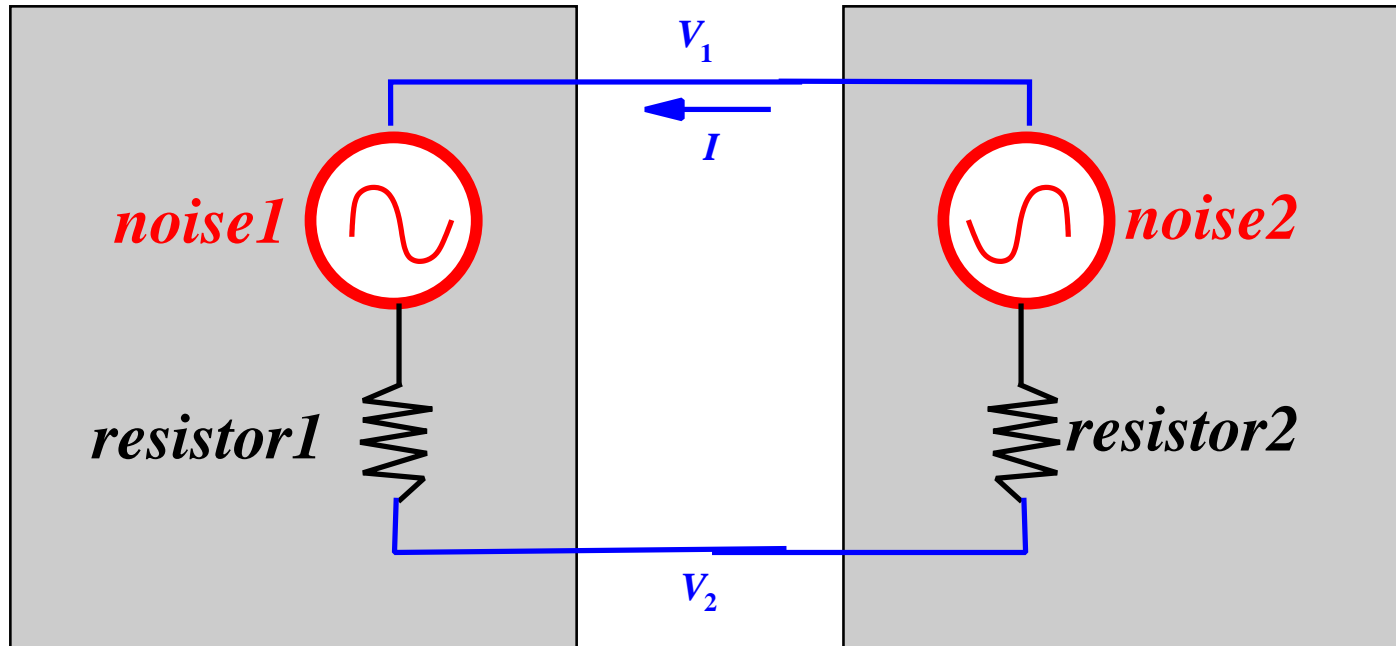
Leading to the σ -algebra generated by the intersections, and the product measure:



Special case: $\mathbb{L}_2 = \{u = 0\}$,

u is then a random variable independent of \mathcal{E}_1 .

Example

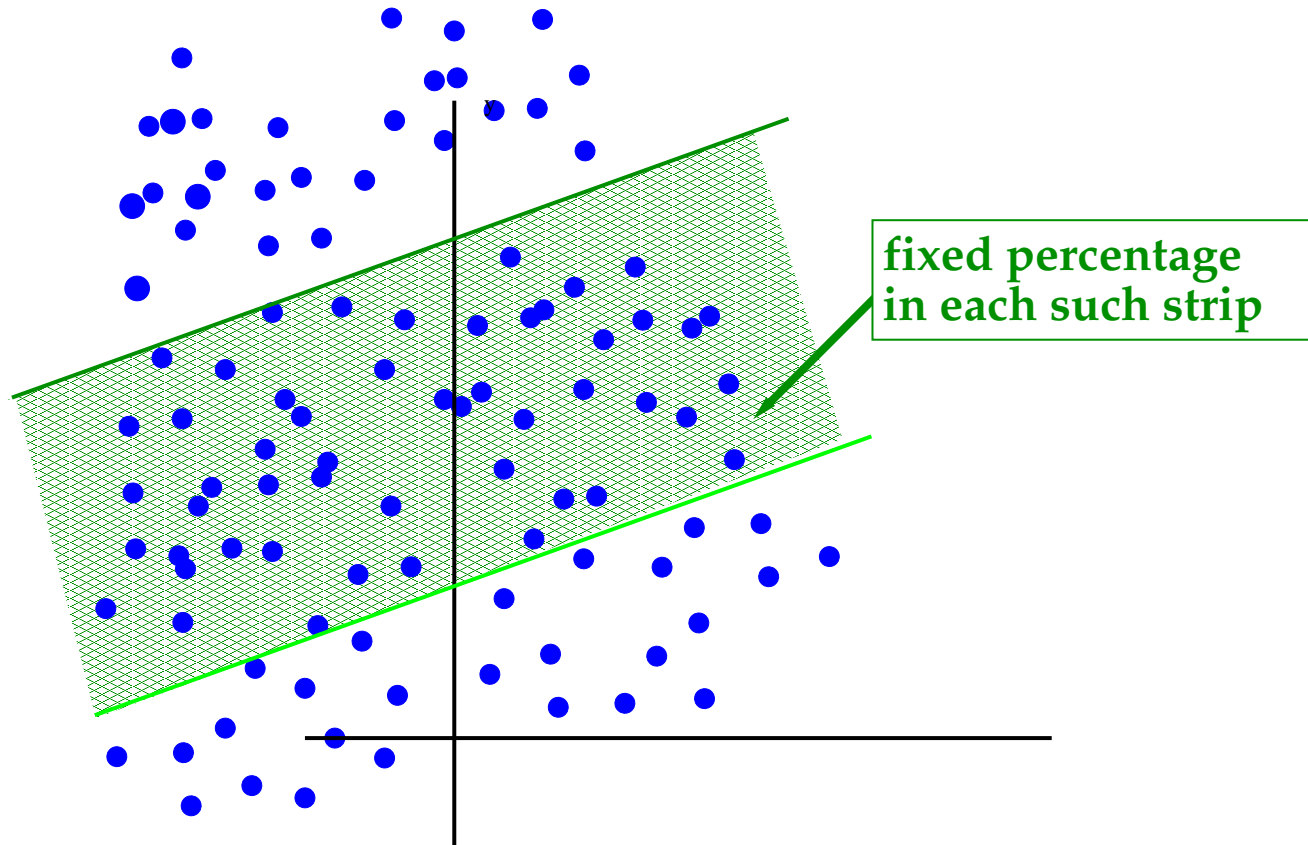


\rightsquigarrow

$$I = \frac{\varepsilon_1 + \varepsilon_2}{R_1 + R_2} \quad V_1 - V_2 = \frac{(R_1 + R_2)\varepsilon_1 + R_1\varepsilon_2}{R_1 + R_2}$$

Regression

Regardless of the experimental conditions (i.e., of the interconnection)



STOCHASTIC BEHAVIORS

DYNAMIC CASE

Stochastic linear time-invariant system

A **stochastic linear time-invariant dynamical system** is given by a stationary random process ε and a polynomial matrix $R \in \mathbb{R}^{\bullet \times w}[\xi]$.

The behavior consists of all $w : \mathbb{Z} \rightarrow \mathbb{R}^w$ such that

$$R(\sigma)w = \varepsilon$$

Stochastic linear time-invariant system

A **stochastic linear time-invariant dynamical system** is given by a stationary random process ε and a polynomial matrix $R \in \mathbb{R}^{\bullet \times w}[\xi]$.

The behavior consists of all $w : \mathbb{Z} \rightarrow \mathbb{R}^w$ such that

$$R(\sigma)w = \varepsilon$$

In particular, there exists \mathcal{M} , an $\mathbb{R}[\xi, \xi^{-1}]$ -submodule of $\mathbb{R}[\xi, \xi^{-1}]^w$, such that

$$[[f \in \mathcal{M}]] \Rightarrow [[f^\top (\sigma, \sigma^{-1}) w \text{ is a stationary process}]]$$

Stochastic linear time-invariant system

In particular, there exists \mathcal{M} , an $\mathbb{R}[\xi, \xi^{-1}]$ -submodule of $\mathbb{R}[\xi, \xi^{-1}]^w$, such that

$[[f \in \mathcal{M}] \Rightarrow [[f^\top (\sigma, \sigma^{-1}) w$ is a stationary process]]

In fact, $\mathcal{M} =$ the module generated by the transposes of the rows of R .

If $f^\top = hR$, then $f (\sigma, \sigma^{-1}) w = h (\sigma, \sigma^{-1}) \varepsilon$.

Stochastic linear time-invariant system

In particular, there exists \mathcal{M} , an $\mathbb{R}[\xi, \xi^{-1}]$ -submodule of $\mathbb{R}[\xi, \xi^{-1}]^w$, such that

$\llbracket f \in \mathcal{M} \rrbracket \Rightarrow \llbracket f^\top (\sigma, \sigma^{-1}) w \text{ is a stationary process} \rrbracket$

In fact, $\mathcal{M} =$ the module generated by the transposes of the rows of R .

If $f^\top = hR$, then $f (\sigma, \sigma^{-1}) w = h (\sigma, \sigma^{-1}) \varepsilon$.

To be worked out:

**Representation questions, their uniqueness,
system identification issues, ...**

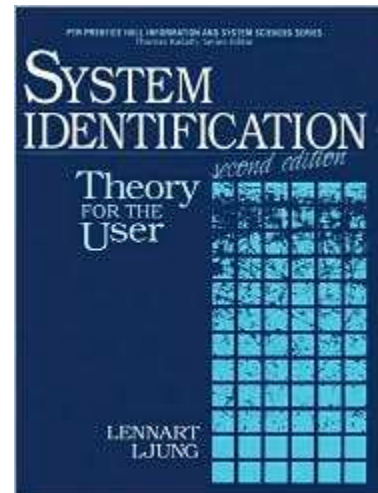
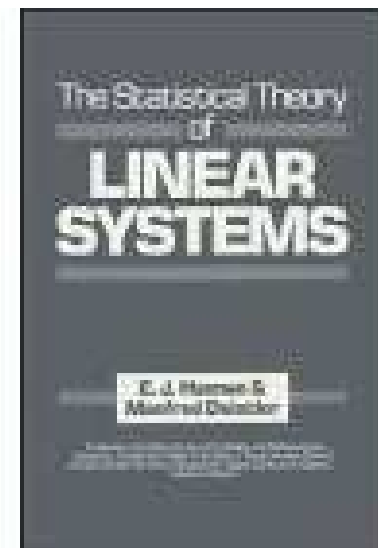
MODELING DISTURBANCES

AS STOCHASTIC PROCESSES

Stochastics in ARMAX systems

$$A(\sigma)y = X(\sigma)u + M(\sigma)\varepsilon$$

The mathematics behind ARMAX systems are among the most elegant, appealing, and subtle in system theory.



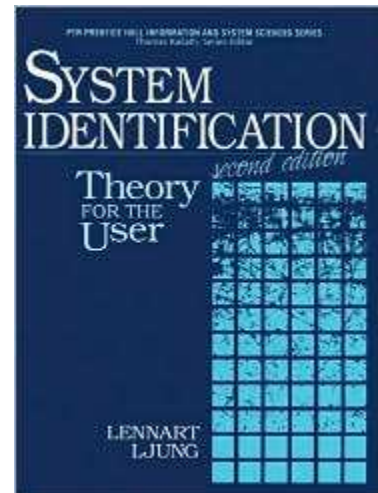
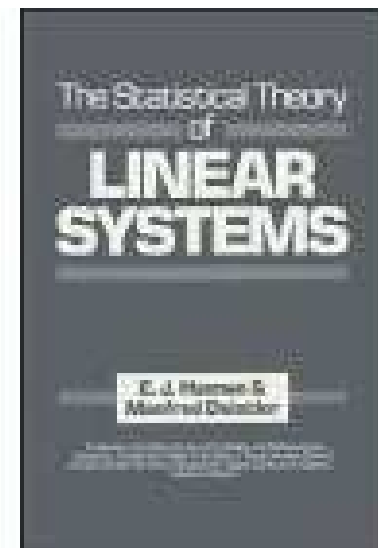
But what about the modeling aspect?

Stochastics in ARMAX systems

$$A(\sigma)y = X(\sigma)u + M(\sigma)\varepsilon$$

What is the rationale of assuming that the disturbances ε are stochastic processes?

Should one interpret probability in the sense of **relative frequency**? or in the sense of **degree of belief**?



Degree of belief

If probability in ARMAX system identification is to be interpreted in the sense of degree of belief, then

- ▶ what is the sense of worrying about consistency and asymptotic efficiency in SYSID?**

Degree of belief

If probability in ARMAX system identification is to be interpreted in the sense of degree of belief, then

- ▶ **what is the sense of worrying about consistency and asymptotic efficiency in SYSID?**
- ▶ **why should **we** care about **their** degree of belief?**

Degree of belief

If probability in ARMAX system identification is to be interpreted in the sense of degree of belief, then

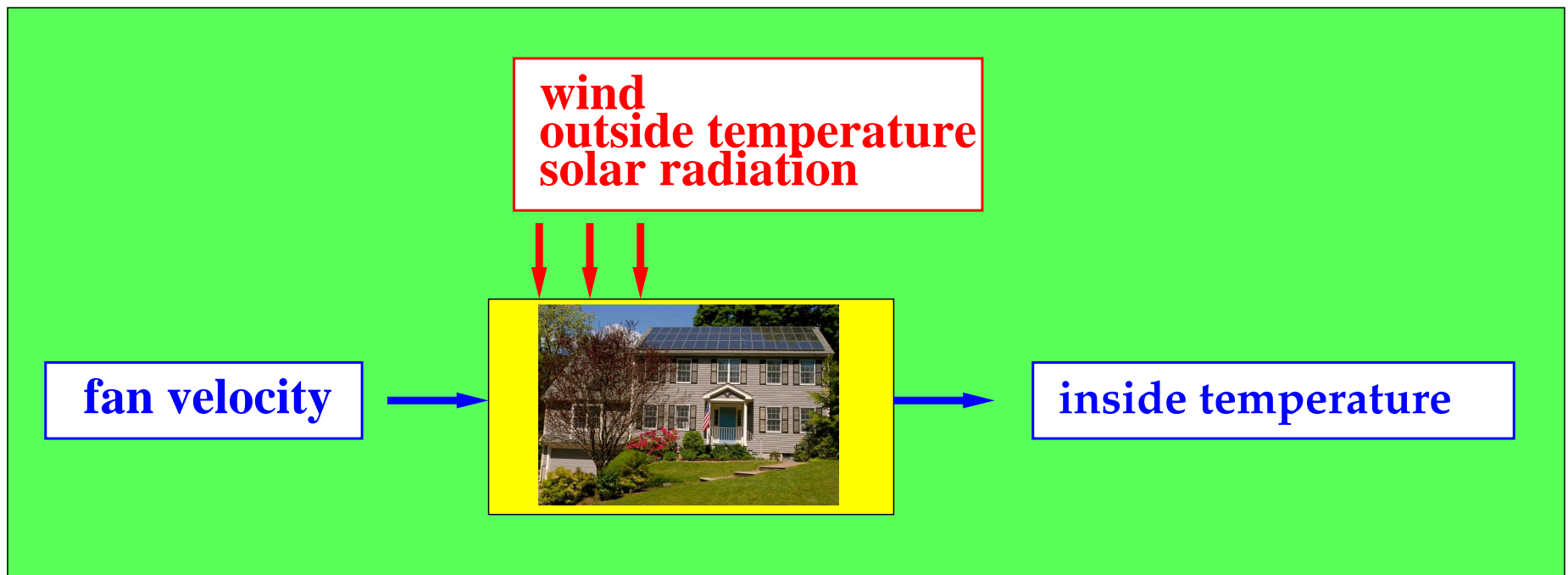
- ▶ what is the sense of worrying about consistency and asymptotic efficiency in SYSID?**
- ▶ why should **we** care about **their** degree of belief?**
- ▶ why not simply stick to least squares, and be much more parsimonious in expressing beliefs?**

Relative frequency

When there is a clear existing real ensemble, relative frequencies are clear and real. Is this the case in time-series and uncertain dynamical systems?

Relative frequency

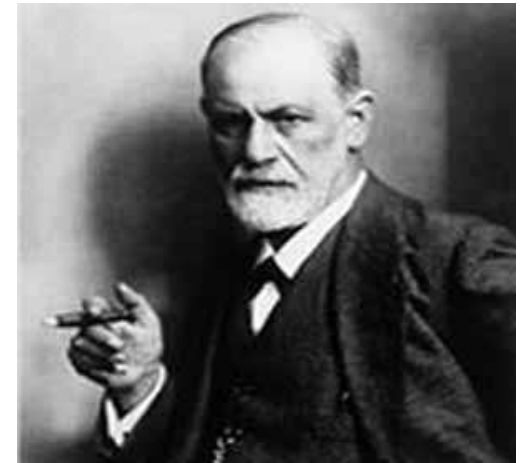
When there is a clear existing real ensemble, relative frequencies are clear and real. Is this the case in time-series and uncertain dynamical systems?



Are these ‘disturbances’ stochastic processes, even approximatey? If so, why?

Uncertainty

The universal use of probability as a panacea for modeling uncertainty in systems and control (and elsewhere) is for me a constant source of discomfort, for a feeling of **Das Unbehagen in der Kultur**



Is probability real?



What is the probability of heads ?

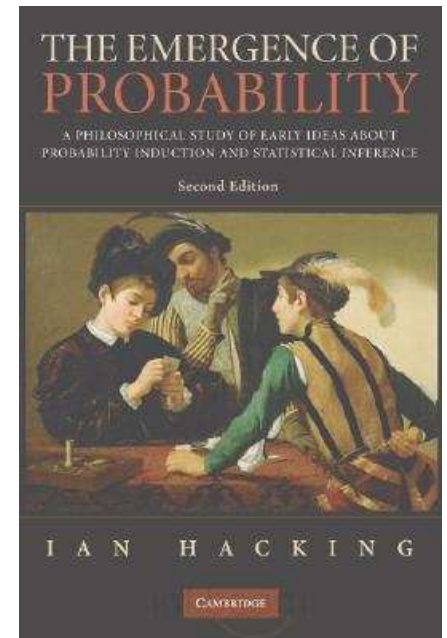
**Many seem to believe that the randomness is
in the coin !**

Is probability real?



What is the probability of heads ?

*“The propensity to give heads is as much a property of the coin as its mass, and the stable long run frequency found on repeated trials is an **objective fact of nature** independent of anyone’s knowledge of it”* I. Hacking, p. 14.

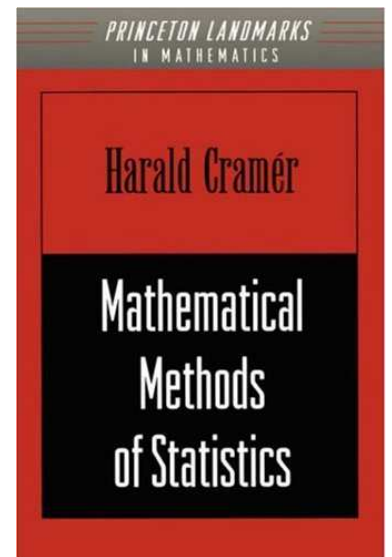


Is probability real?



What is the probability of heads ?

“*The numbers p_r [the probability of the outcome r] should in fact be regarded as **physical constants** of the particular die that we are using*” H. Cramer, p. 154



Physics or stochastics?

Persi Diaconis builds a coin tosser



and discovered that if the coin is tossed exactly the same way, it falls on the same side 100% of the time.

The press appears indignified:

The Not So Random Coin Toss

 Listen by David Kestenbaum



[Larger Image of the Machine](#)

Susan Holmes
Statistician Persi
Diaconis' mechanical
coin flipper.

All Things Considered, February 24, 2004 ·

Flipping a coin may not be the fairest way to settle disputes. About a decade ago, statistician Persi Diaconis started to wonder if the outcome of a coin flip really is just a matter of chance. He had Harvard University engineers build him a mechanical coin flipper. Diaconis, now at Stanford University, found that if a coin is launched exactly the same way, it lands exactly the same way.

The randomness in a coin toss, it appears, is introduced by sloppy humans. Each human-generated flip has a different height and speed, and is caught at a different angle, giving different outcomes.

Physics or stochastics?

The press appears indignified:

The Not So Random Coin Toss

 Listen by David Kestenbaum



All Things Considered, February 24, 2004 ·

‘sloppy humans’?????

Flipping a coin may not be the fairest way to settle disputes. About a decade ago, statistician Persi Diaconis started to wonder if the outcome of a coin flip really is just a

You must mean sloppy professors who indoctrinate students and journalists to believe that the randomness could conceivably be in the coin...

[Larger Image of the](#)

[Full Image](#)

Susan Holmes

Statistics

Diaconis' mechanical coin flipper.

random event. Diaconis, now at Harvard University, found that if a coin is attached exactly the same way, it lands exactly the same way.

The randomness in a coin toss, it appears, is introduced by sloppy humans. Each human-generated flip has a different height and speed, and is caught at a different angle, giving different outcomes.

Physics or stochastics?

The scientists come to the following conclusions:

We conclude that coin tossing is ‘physics’, not ‘random’

P. Diaconis, S. Holmes and R. Montgomery, Dynamical bias in the coin toss,

SIAM Review, 2007, page 211.

I could have told them that without the benefit of a machine...

Physics or stochastics?

The scientists come to the following conclusions:

If we have this much trouble analyzing a common coin toss, the reader can imagine the difficulty we have with interpreting typical stochastic assumptions in econometric analysis

Agreed, from the bottom of my heart!

CONCLUSIONS

Conclusions

▶ **An open stochastic system is best defined in terms of unusual σ -algebra.**

\rightsquigarrow a crisper definition, which does not require input/output separation, and avoids the discussion of statistical dependence of input and noise.

Conclusions

- ▶ **I am uncomfortable with the use of probability as a panacea for uncertainty.**

Conclusions

- ▶ **I am uncomfortable with the use of probability as a panacea for uncertainty.**
- ▶ **I find it difficult to fathom the origin of the conviction that uncertainty is intrinsic in some systems, e.g., coins and dice, and wiggly time-series. Comes (in part) from misunderstanding ‘closed’ versus ‘open’ systems.**

Possible exception: QM.

Copies of the lecture frames will be available from/at

Jan.Willems@esat.kuleuven.be

<http://www.esat.kuleuven.be/~jwillems>

Copies of the lecture frames will be available from/at

Jan.Willems@esat.kuleuven.be

<http://www.esat.kuleuven.be/~jwillems>

Thank you

Thank you

Thank you

Thank you

Thank you

Thank you

Thank you

Thank you



Manfred, enjoy your retirement!