Information Anatomy in Seismology

Mikhael T. Semaan, Ph. D.
EnviTrace LLC

Satellite's Eye

- Subsurface is tricky:
  - Coupled + complex physics (THM, fractures, materials…)
  - A nightmare to "see"
- Generically: prediction hard; understanding harder
- What to do?
  - Statistical modeling, incl.
  - AI/ML-based

https://img.freepik.com/premium-photo/detailed-crosssection-illustration-subsurface-layers-depicting-rocks-soil-industrial-influences-highlighting-geological-environmental-factors_95891-116721.jpg

Bird’s Eye

- Want to “do the best we can”
- Q: How good is that?
  - Information-theoretic eval. and bounds
  - Predictive relationships in multimodal data
- Why?
  - Probe limits of AI/ML
  - Lend interpretability + understanding

https://img.freepik.com/premium-photo/detailed-crosssection-illustration-subsurface-layers-depicting-rocks-soil-industrial-influences-highlighting-geological-environmental-factors_95891-116721.jpg

Human's Eye: Outline

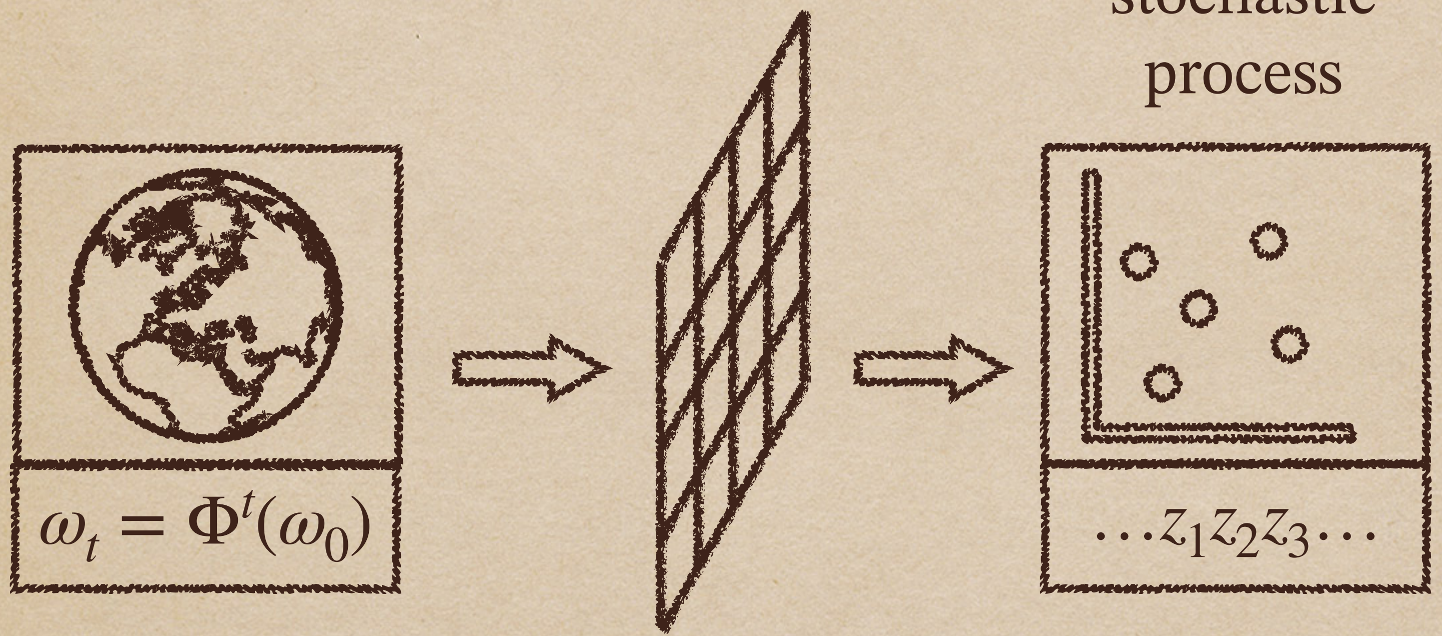
- From dynamical system to stochastic process
- Stochastic process basics
- "Information anatomy" and i-Diagrams
- Seismic catalog data
- Implementation
- Seismic catalog results
- Ideas for more
- Thanks + Q&A

Dynamics to Stochastics

full physics + partial observation = actual data

↑ (realized from)

stochastic process



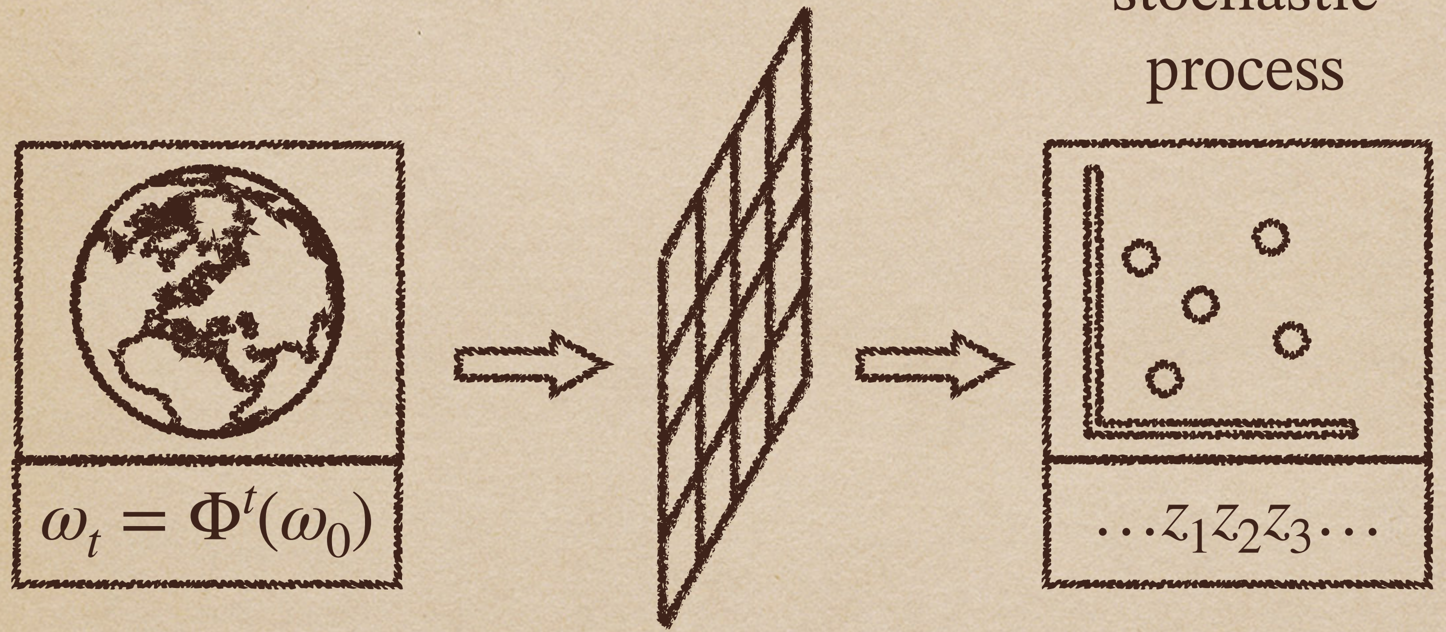
- Through one mechanism or another
(limited stations, finite precision instruments, analytical approximations, numerical stability limits, poor/missing data, scale of data and analysis…)

Dynamics to Stochastics

full physics + partial observation = actual data

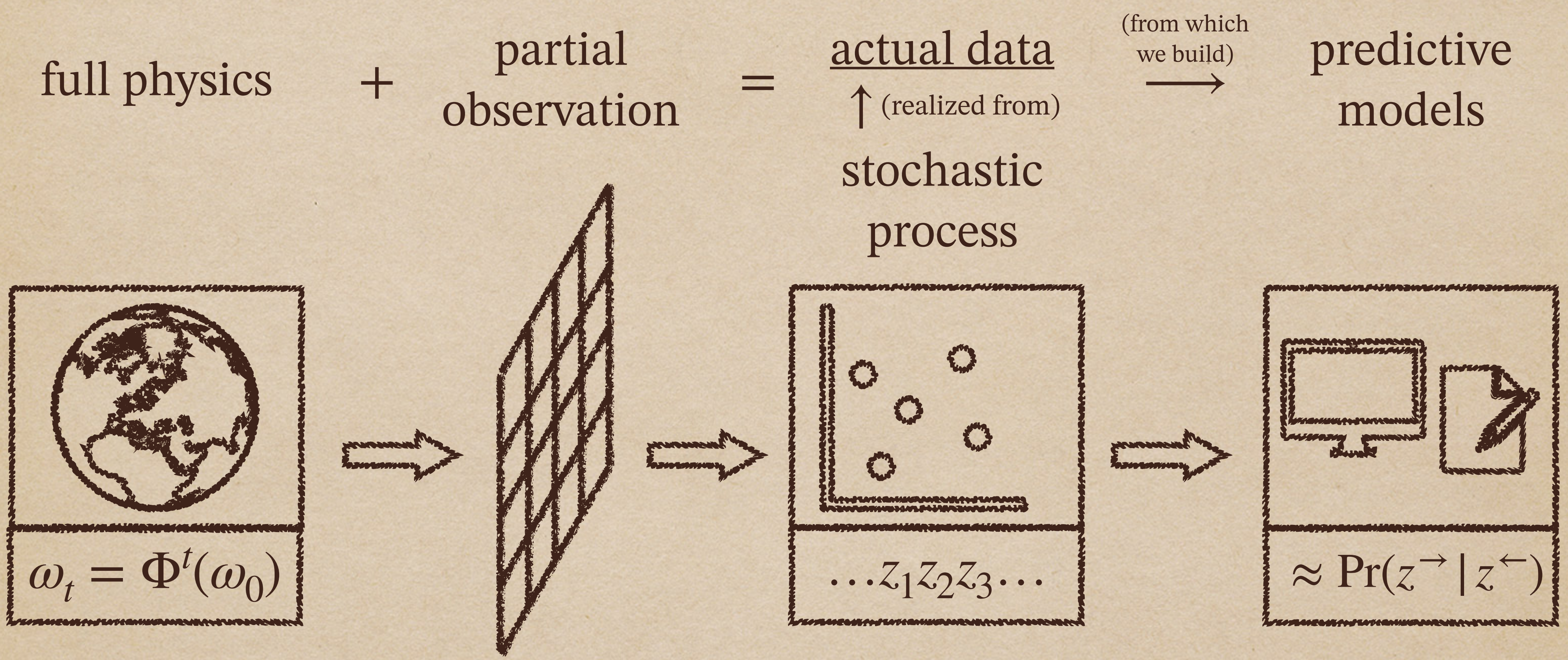
↑ (realized from)

stochastic process



- Through one mechanism or another—or combos thereof—
- In practice (even in theory!), we start with a stochastic process

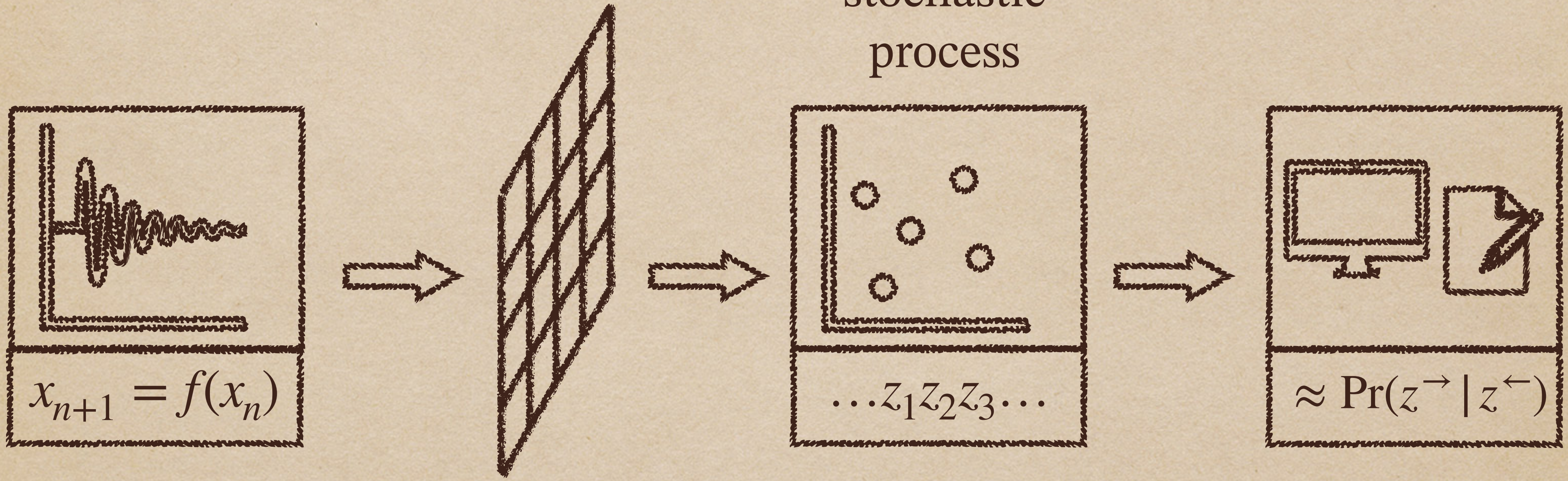
Dynamics to Stochastics



- Through one mechanism or another—or combos thereof—
- In practice (even in theory!), we start with a <u>stochastic process</u>
- Model build from there!

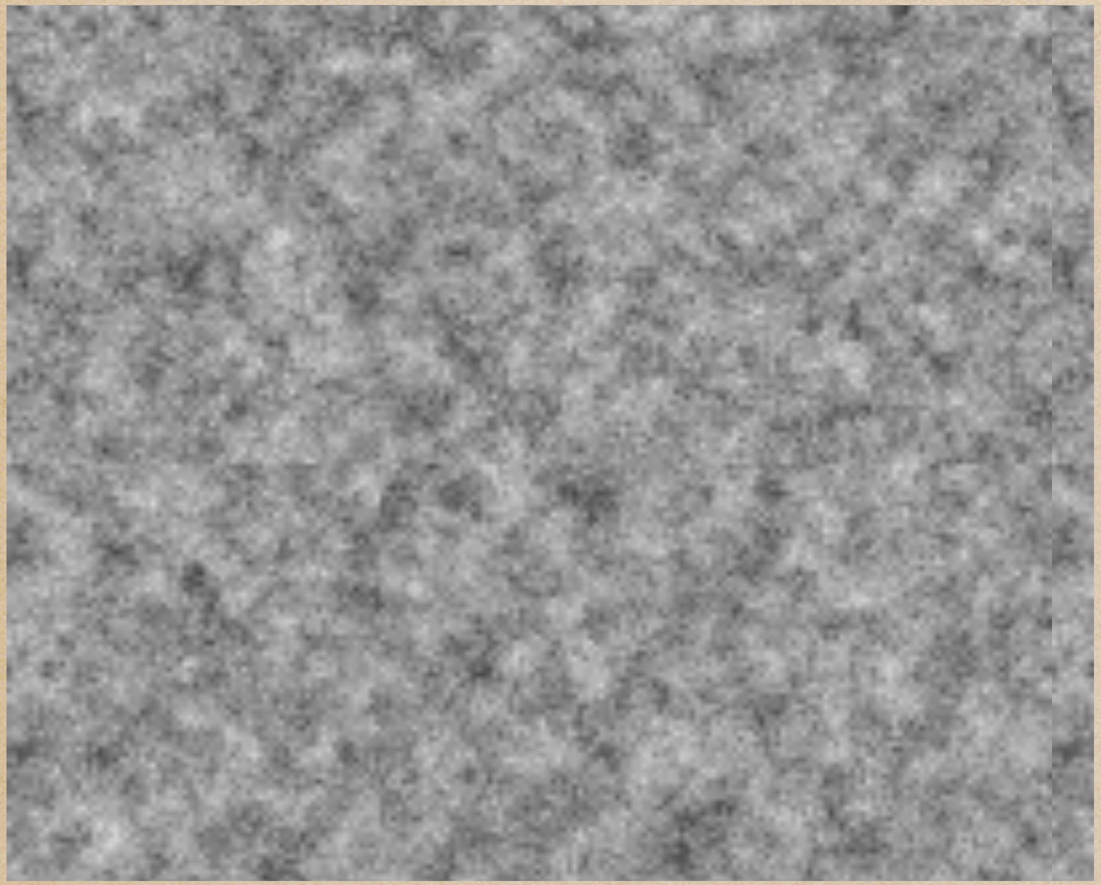
Dynamics to Stochastics

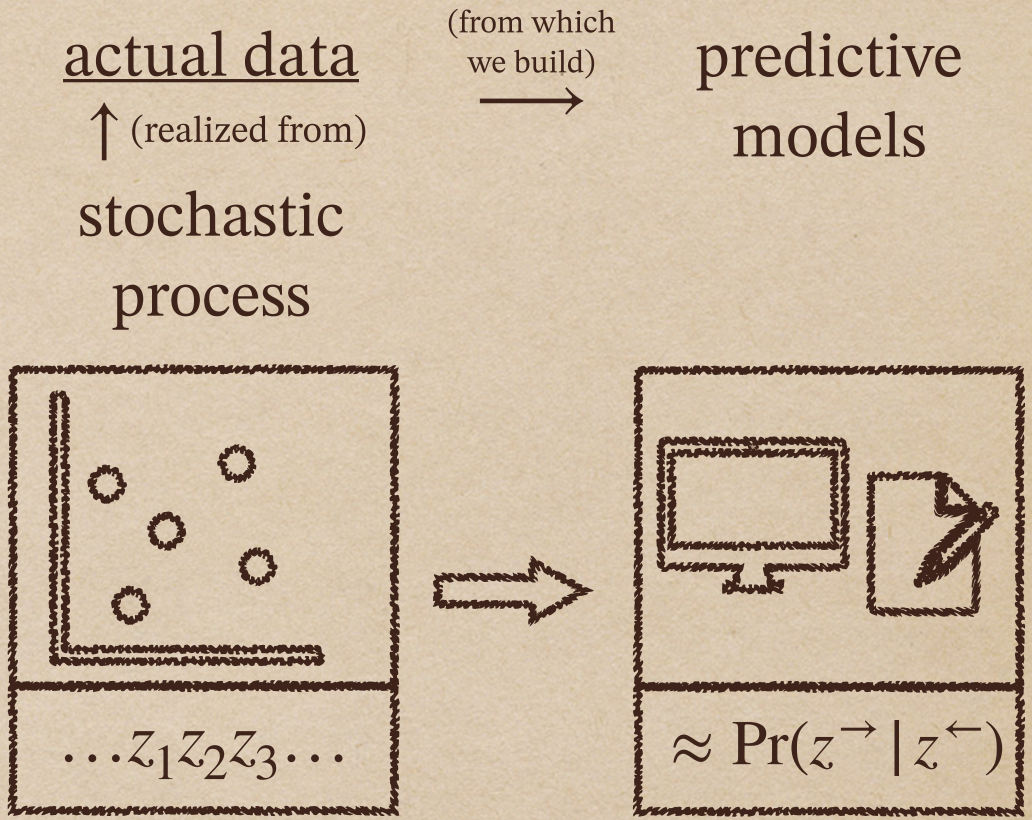
full physics + partial observation = actual data ↑ (realized from) stochastic process (from which we build) ⟶ predictive models



- Example: chaotic map—logistic, etc.
- Partitioning state space
  - → symbolic dynamics
  - → stationary, ergodic stochastic process
  - → nice invariant properties about underlying dynamical system

Dynamics to Stochastics





- "Worst case," we start with *only* data
- This is the root of scientific enterprise, incl. theory building
- "Purely" data-driven we can think of as *implicit modelling*
  - Precisely the case with e.g. Takens' and *Geometry* in dynamics

Stochastic Process

- $\mathcal{Z}$: system state space; examples:
  - # of seismic events above magnitude M
  - Exact positions/momenta of each volume element in a finite grid model of the subsurface
  - Etc. 😉

Stochastic Process

- Z_i : random variable (RV) for system state at time step i
- Our *stochastic process* is the joint RV:

$$\dots Z_{-3}\, Z_{-2}\, Z_{-1}\, Z_0 \quad Z_1 \quad Z_2 \quad Z_3 \dots$$

- From joint, can marginalize/condition as desired!

Splitting Time

- Split joint RV into past and future

past | future

$$\ldots Z_{-3}\, Z_{-2}\, Z_{-1}\, Z_0 \;\big|\; Z_1\;\; Z_2\;\; Z_3 \ldots$$

Splitting Time

- Split joint RV into past and future
- Take Shannon entropies: “information in” past/future

past | future

$\ldots Z_{-3}\, Z_{-2}\, Z_{-1}\, Z_0 \;|\; Z_1 \;\; Z_2 \;\; Z_3 \ldots$

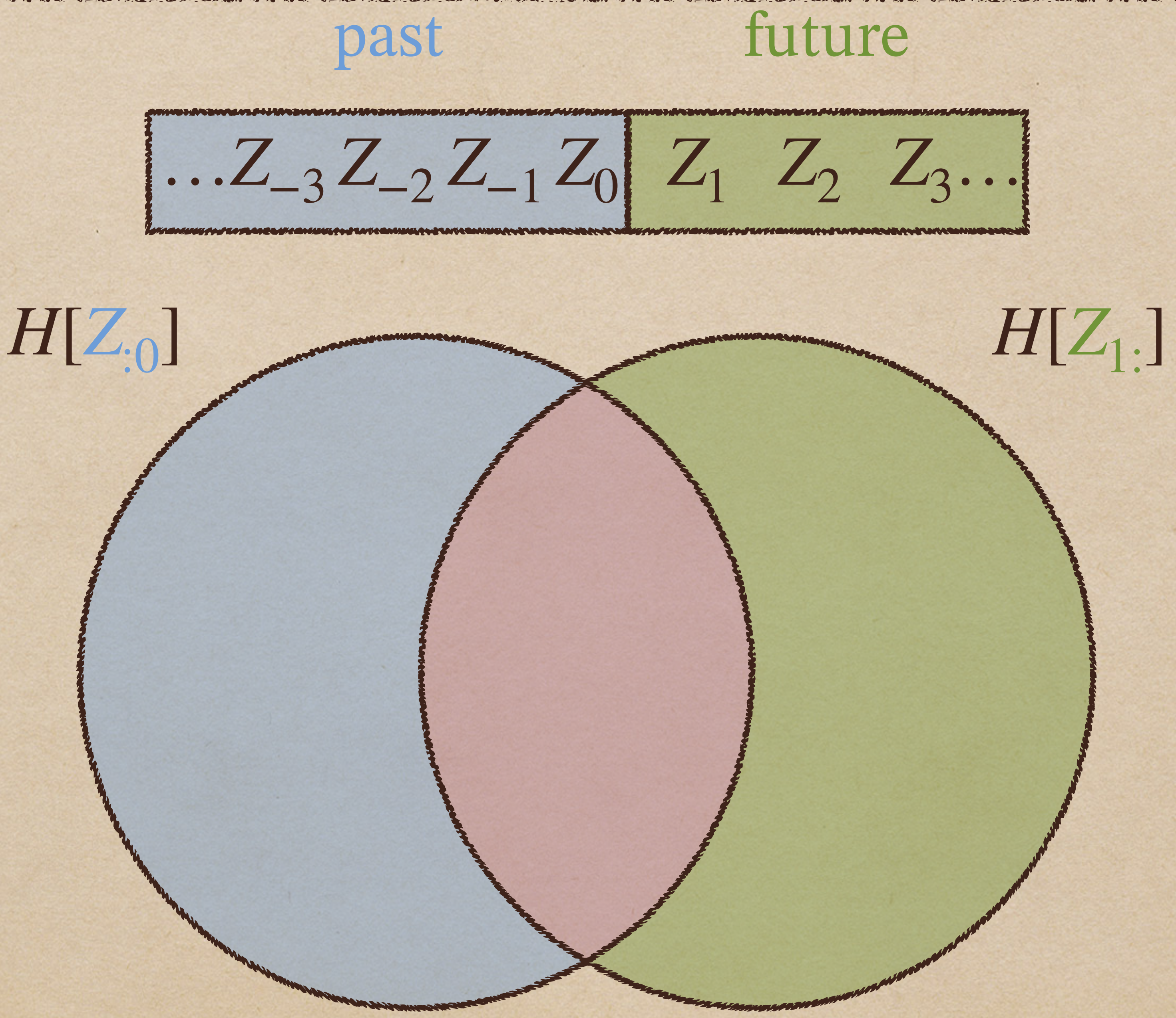
$H[Z_{:0}]$ | $H[Z_{1:}]$

- Reminder:

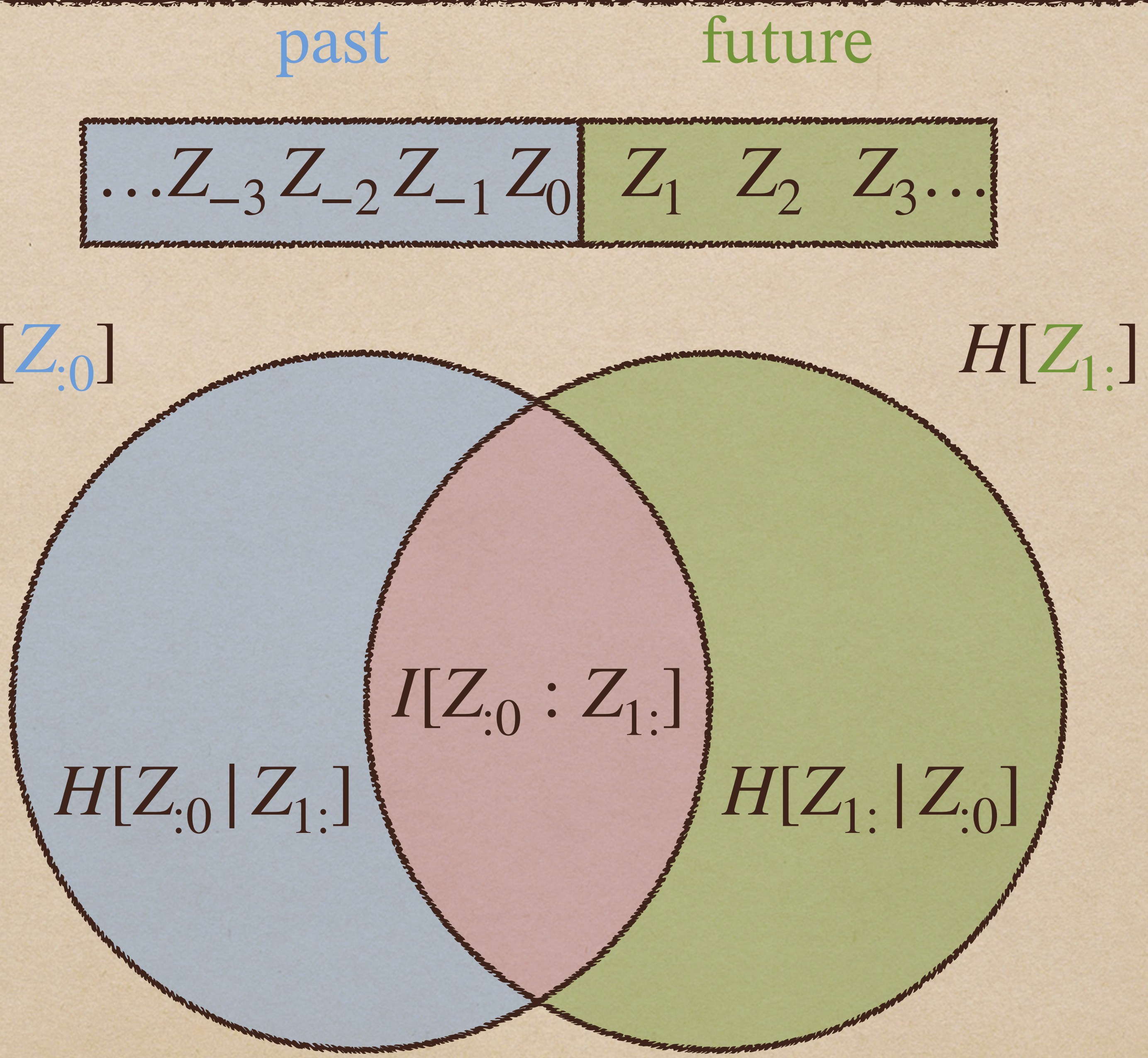
  $$H[X] = \mathbb{E}[-\log \Pr(X)]$$

- Literally, expected surprisal on observation of X
- <u>Surprising</u> outcome → very <u>informative</u>, and *vice versa*

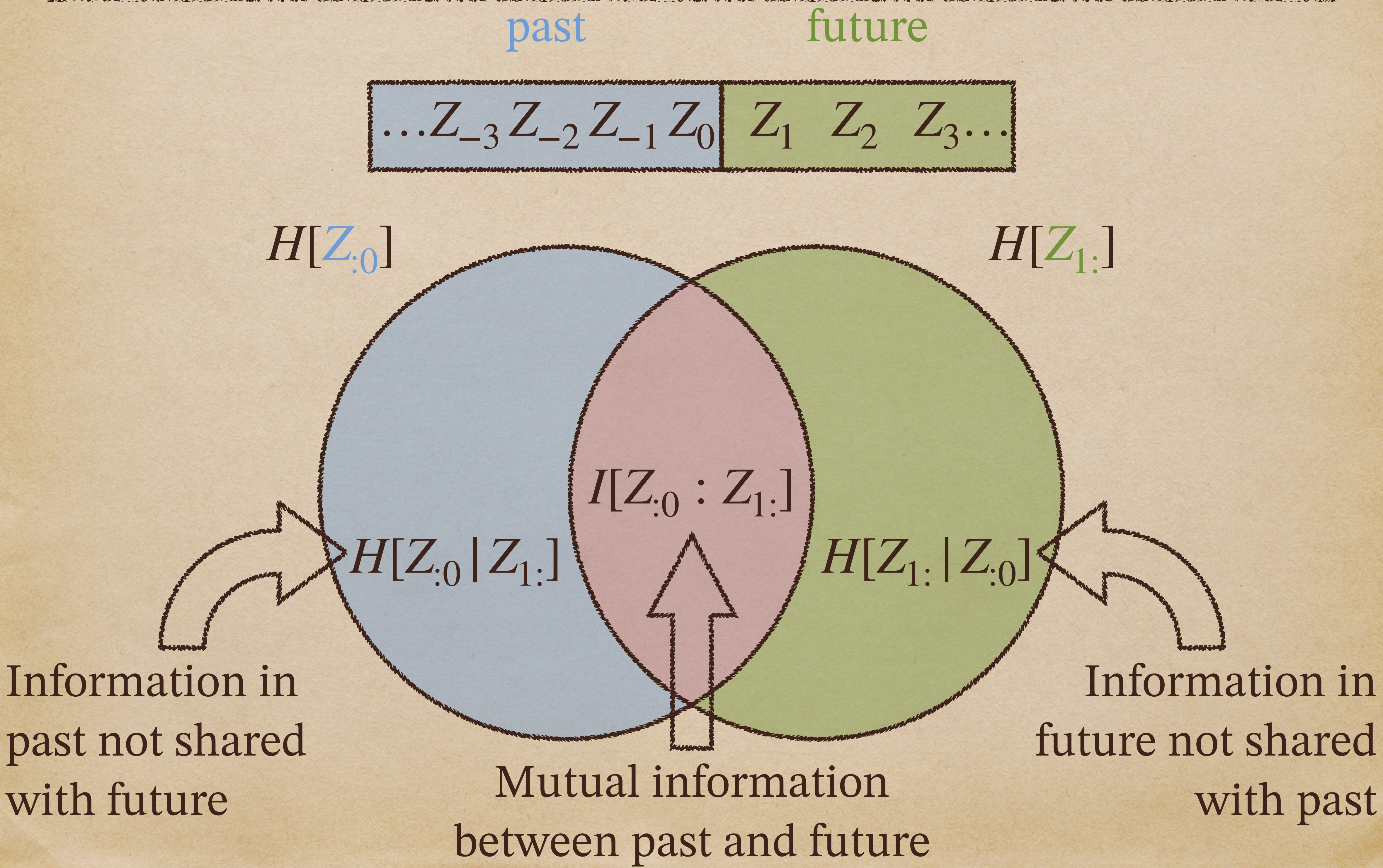
i-Diagram



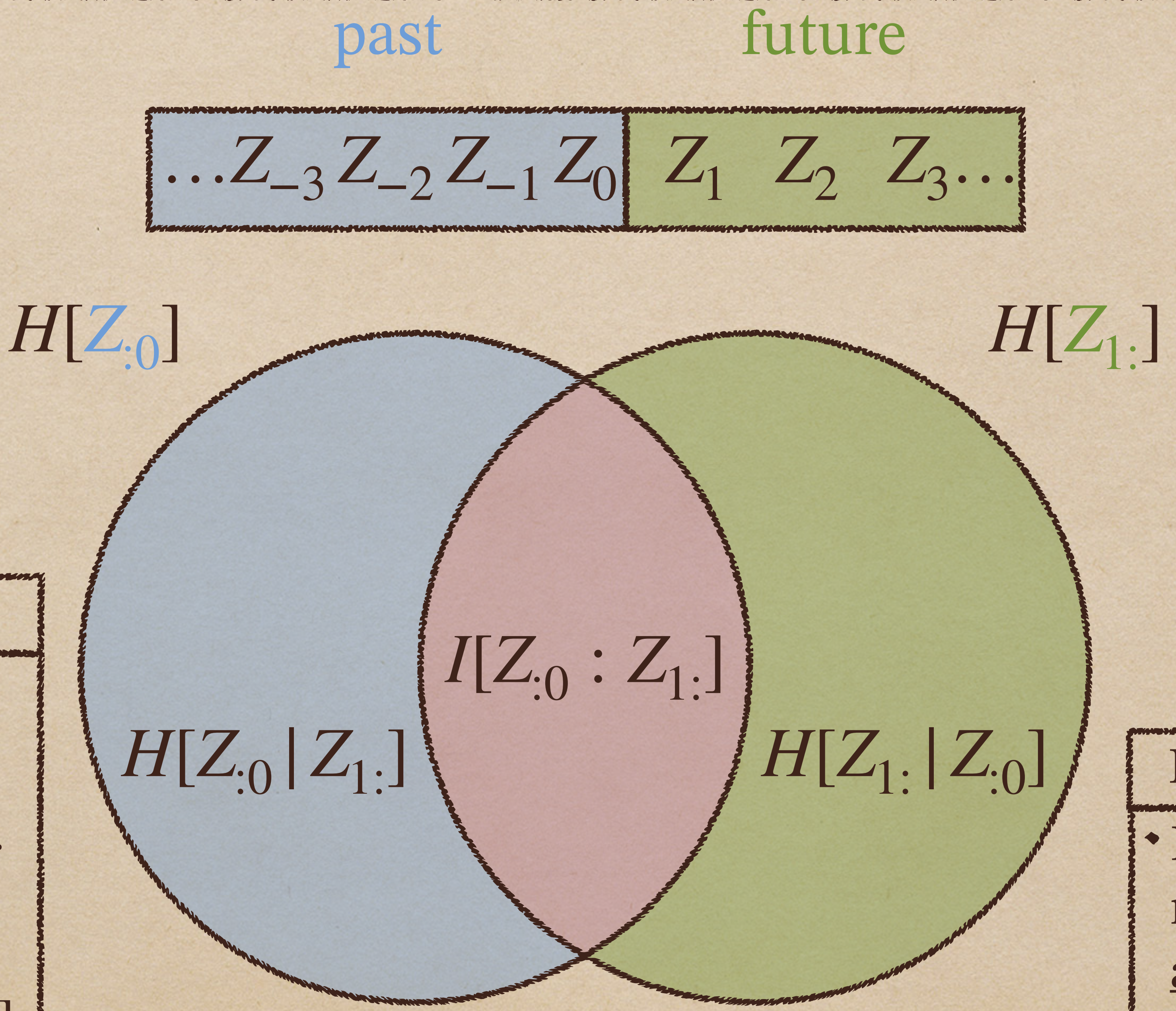
i-Diagram


past
future
$\ldots Z_{-3}\, Z_{-2}\, Z_{-1}\, Z_0$
$Z_1\ \ Z_2\ \ Z_3 \ldots$
$H[Z_{:0}]$
$H[Z_{1:}]$
$I[Z_{:0} : Z_{1:}]$
$H[Z_{:0} | Z_{1:}]$
$H[Z_{1:} | Z_{:0}]$

i-Diagram


past
future
$\ldots Z_{-3}\, Z_{-2}\, Z_{-1}\, Z_0$
$Z_1 \;\; Z_2 \;\; Z_3 \ldots$
$H[Z_{:0}]$
$H[Z_{1:}]$
$I[Z_{:0} : Z_{1:}]$
$H[Z_{:0} | Z_{1:}]$
$H[Z_{1:} | Z_{:0}]$
Information in past not shared with future
Mutual information between past and future
Information in future not shared with past

i-Diagram



Formula Key I

- Get joint dist. $\Pr(Z_{:0}, Z_{1:})$
- Marginalize for $\Pr(Z_{:0})$, $\Pr(Z_{1:})$
- Calc. $H[Z_{:0}, Z_{1:}]$, $H[Z_{:0}]$, $H[Z_{1:}]$

Formula Key II

- Relationships now visually additive to get conditionals, mutual info.

Our Time Series'

- Are, first of all, finite:

  $$z_1 z_2 \ldots z_N \equiv z_{1:N} \in \mathscr{Z}^N$$

- Suppose realizations of joint RV $Z_{1:N}$, but:
- Only one sample, so can't say much…
- Instead: build new RV $Z'_{1:L}$ for $L < N$, by delay embedding

Our Time Series'

- Instead: build new RV $Z'_{1:L}$ for $L < N$, by delay embedding

Our Time Series’

- Instead: build new RV $Z'_{1:L}$ for $L < N$, by delay embedding
- Example: $L = 3$, binary time series

$$z_{1:N} = 0\,1\,0\,1\,1\,0\,1\,1\,1\,0\,0\,1\,0\,0\,0\,0\,\ldots$$

$z'_1\ z'_2\ z'_3$	Count
0 0 0	0
0 0 1	0
0 1 0	0
0 1 1	0
1 0 0	0
1 0 1	0
1 1 0	0
1 1 1	0

Our Time Series'

- Instead: build new RV $Z'_{1:L}$ for $L < N$, by delay embedding
- Example: $L = 3$, binary time series

$z_{1:N} =$ 0 1 0 1 1 0 1 1 1 0 0 1 0 0 0 0 …

z'_1 z'_2 z'_3	Count
0 0 0	0
0 0 1	0
0 1 0	1
0 1 1	0
1 0 0	0
1 0 1	0
1 1 0	0
1 1 1	0

Our Time Series'

- Instead: build new RV $Z'_{1:L}$ for $L < N$, by <u>delay embedding</u>
- Example: $L = 3$, binary time series

$$z_{1:N} = 0\,\boxed{1\,0\,1}\,1\,0\,1\,1\,1\,0\,0\,1\,0\,0\,0\,0\,\ldots$$

$z'_1\ z'_2\ z'_3$	Count
0 0 0	0
0 0 1	0
0 1 0	1
0 1 1	0
1 0 0	0
1 0 1	1
1 1 0	0
1 1 1	0

Our Time Series'

- Instead: build new RV $Z'_{1:L}$ for $L < N$, by delay embedding
- Example: $L = 3$, binary time series

$$z_{1:N} = 0\ 1\ \boxed{0\ 1\ 1}\ 0\ 1\ 1\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ \ldots$$

$z'_1\ z'_2\ z'_3$	Count
0 0 0	0
0 0 1	0
0 1 0	1
0 1 1	1
1 0 0	0
1 0 1	1
1 1 0	0
1 1 1	0

Our Time Series'

- Instead: build new RV $Z'_{1:L}$ for $L < N$, by delay embedding
- Example: $L = 3$, binary time series

$$z_{1:N} = 0\,1\,0\,1\,1\,0\,1\,1\,1\,0\,0\,1\,0\,0\,0\,0 \ldots$$

$z'_1\, z'_2\, z'_3$	Count
0 0 0	2
0 0 1	1
0 1 0	2
0 1 1	2
1 0 0	1
1 0 1	2
1 1 0	2
1 1 1	1

Our Time Series'

- Instead: build new RV $Z'_{1:L}$ for $L < N$, by delay embedding
- Example: $L = 3$, binary time series

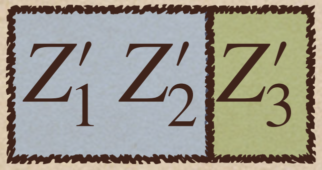
$$z_{1:N} = 0\,1\,0\,1\,1\,0\,1\,1\,1\,0\,0\,1\,0\,\boxed{0\,0\,0}\ldots$$

$z'_1\ z'_2\ z'_3$	$\Pr'_{1:3}$
0 0 0	2/13
0 0 1	1/13
0 1 0	2/13
0 1 1	2/13
1 0 0	1/13
1 0 1	2/13
1 1 0	2/13
1 1 1	1/13

- After, normalize to build $\Pr'_{1:3}$
- RV $Z'_{1:3} \sim \Pr'_{1:3}$

Our Time Series'

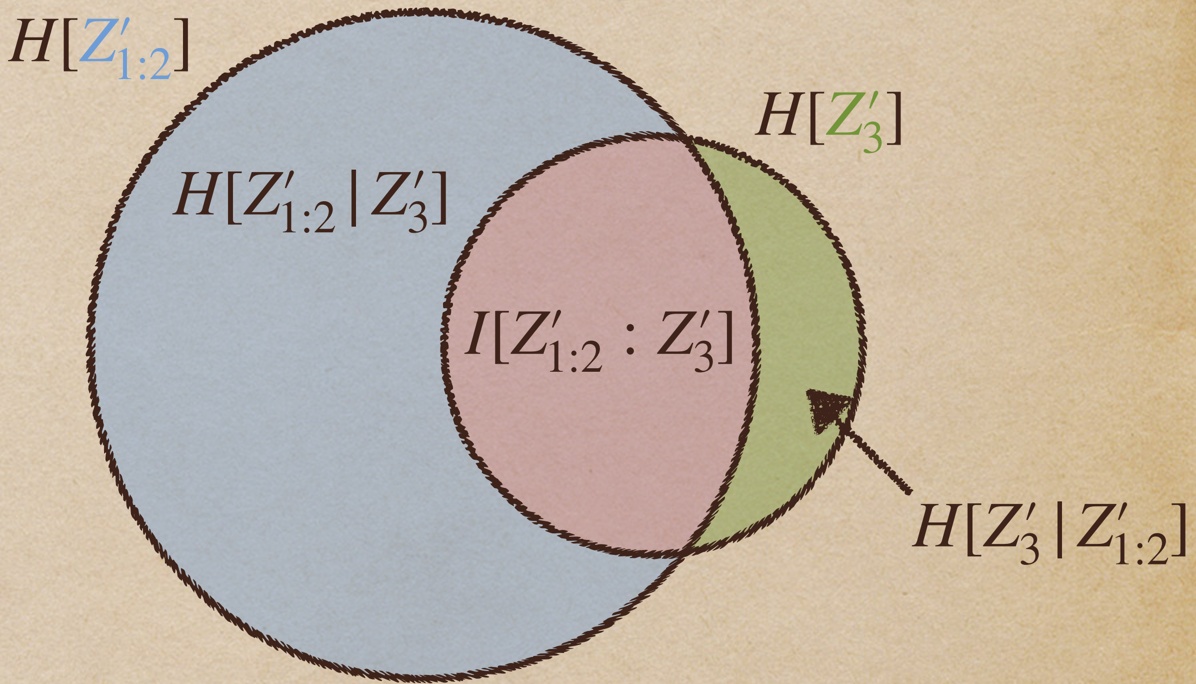
- Instead: build new RV $Z'_{1:L}$ for $L < N$, by delay embedding
- Example: $L = 3$, binary time series
- Now we have:



Our Time Series'

- Instead: build new RV $Z'_{1:L}$ for $L < N$, by <u>delay embedding</u>
- Example: $L = 3$, binary time series
- Now we have:

$Z'_1 \; Z'_2 \; Z'_3$



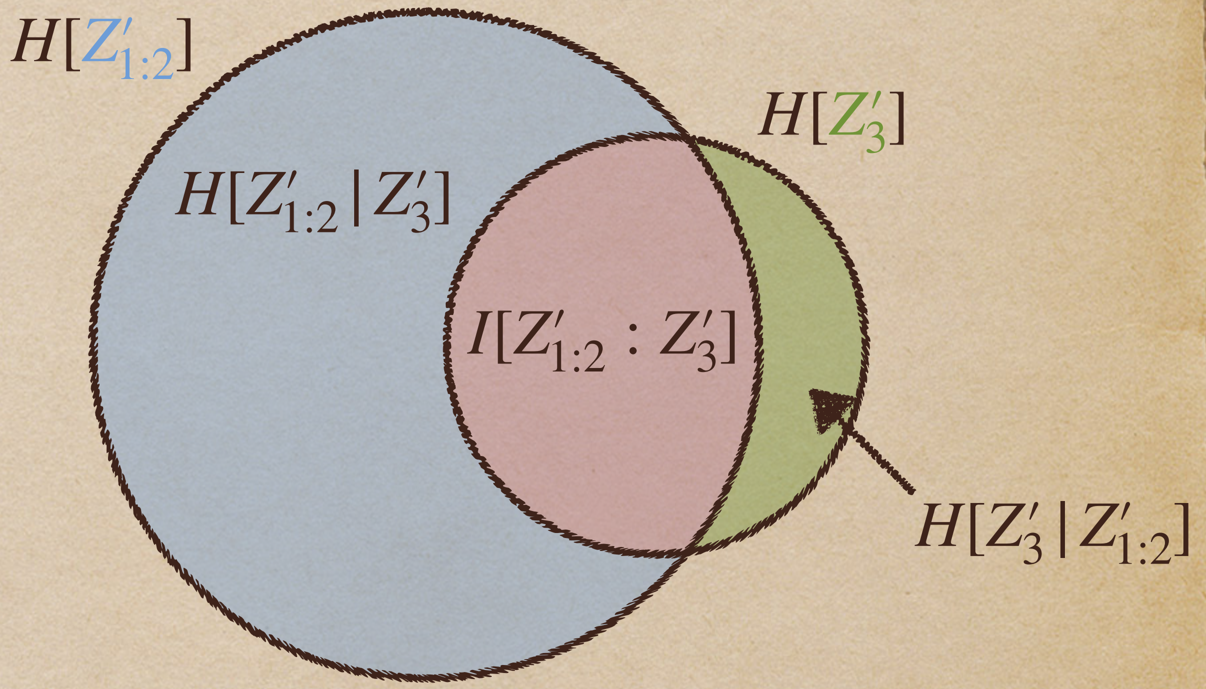
- For our made-up example:

Info. Measure	Value (bit)
$H[Z'_{1:3}]$	2.700
$H[Z'_{1:2}]$	2.511
$H[Z'_3]$	0.996
$H[Z'_3 \| Z'_{1:2}]$	0.189
$H[Z'_{1:2} \| Z_3]$	1.705
$I[Z'_{1:2} : Z'_3]$	0.806

Our Time Series'

- Instead: build new RV $Z'_{1:L}$ for $L < N$, by delay embedding
- Example: $L = 3$, binary time series
- Now we have:

$Z'_1 \; Z'_2 \; Z'_3$



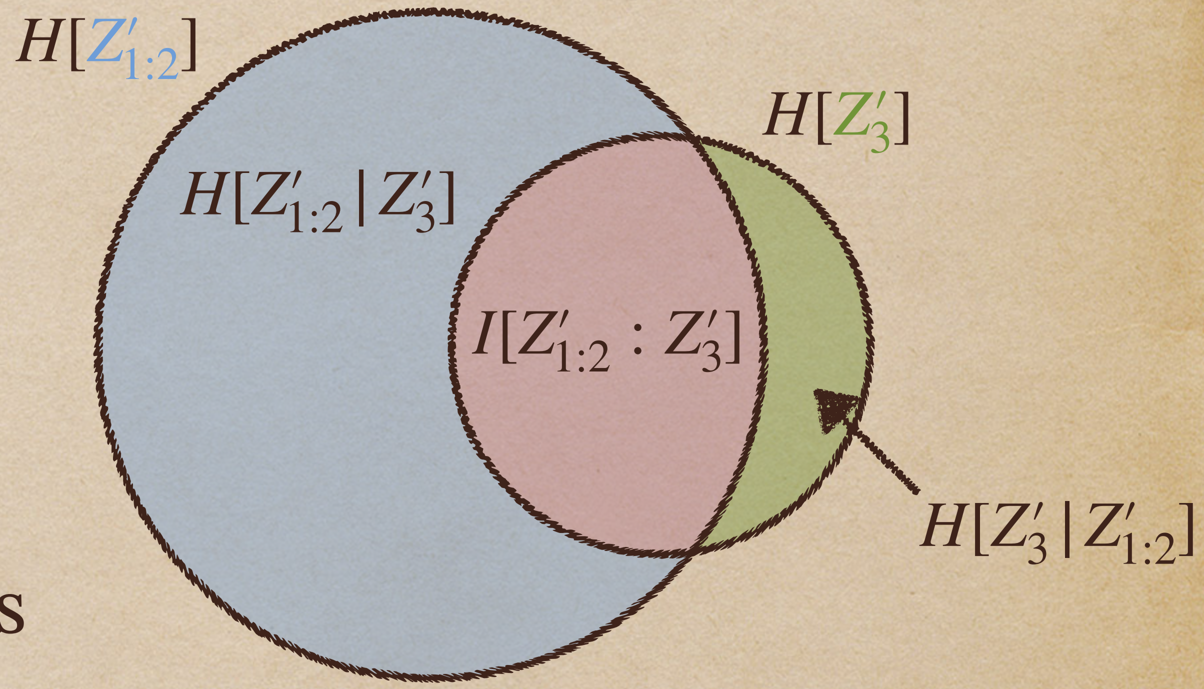
- For our made-up example:

Info. Measure	Value (bit)
$H[Z'_{1:3}]$	2.700
$H[Z'_{1:2}]$	2.511
$H[Z'_3]$	0.996
$H[Z'_3 \| Z'_{1:2}]$	0.189
$H[Z'_{1:2} \| Z'_3]$	1.705
$I[Z'_{1:2} : Z'_3]$	0.806

- If we had assumed IID, ~coin flip
- But conditioned on past for temporal structure, quite predictable!

Our Time Series'

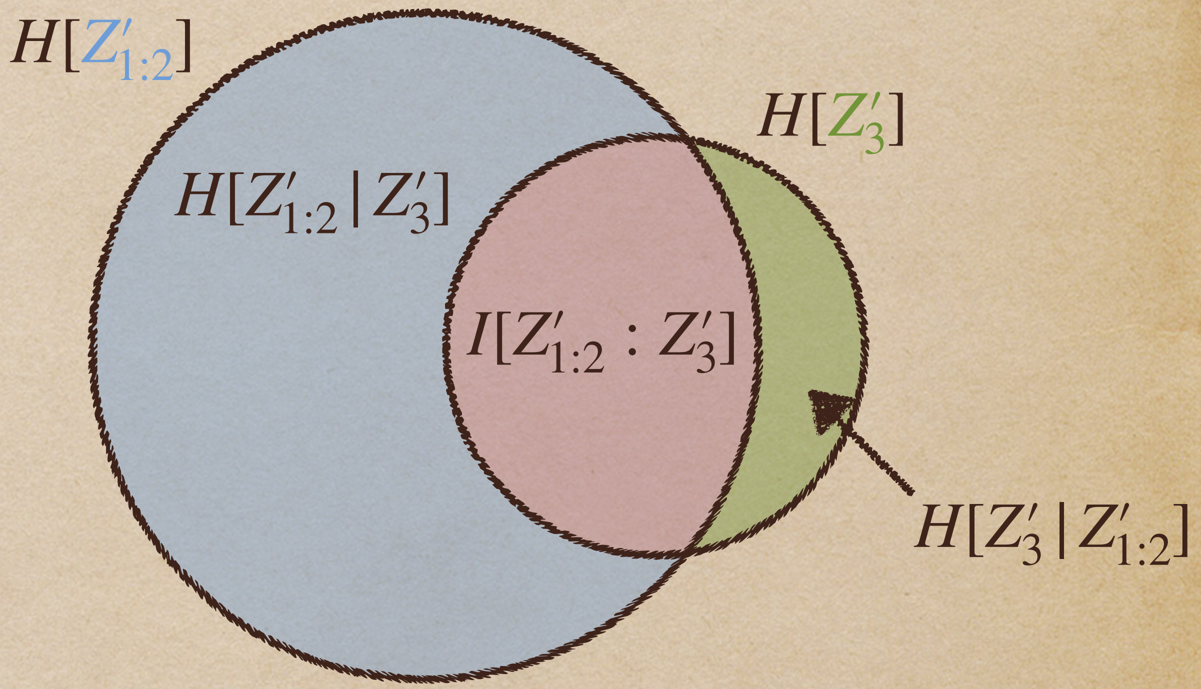
- Instead: build new RV $Z'_{1:L}$ for $L < N$, by <u>delay embedding</u>
- Example: $L = 3$, binary time series
- Now we have:



- Generally: lots of choices!
  - Bi/trivariate (+, sort of)
  - Choices of past/future lengths
  - Extension to multi-series w/ joint state spaces

Our Time Series'

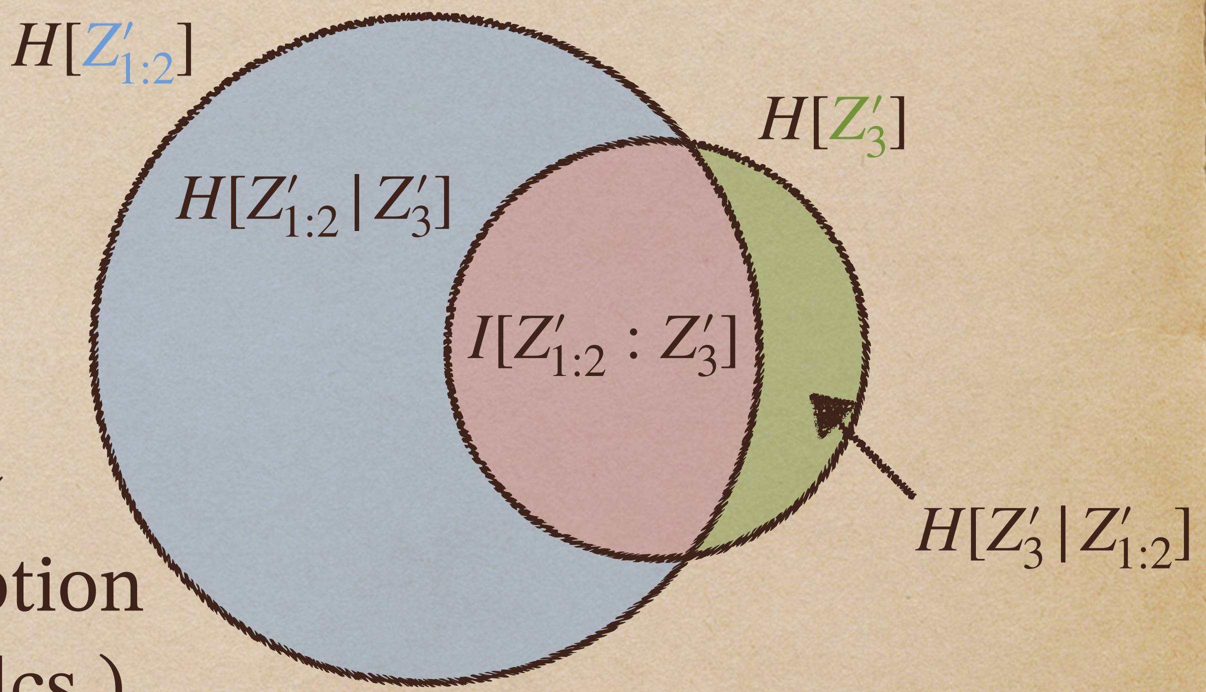
- Instead: build new RV $Z'_{1:L}$ for $L < N$, by <u>delay embedding</u>
- Example: $L = 3$, binary time series
- Now we have:



- Caveats
  - Strict convergence and links from $Z'_{1:L} \rightarrow Z_{i:i+L-1}$ rely on stationarity, ergodicity
  - Embedding space explodes w/increasing L (so need tons of data)

Our Time Series'

- Instead: build new RV $Z'_{1:L}$ for $L < N$, by <u>delay embedding</u>
- Example: $L = 3$, binary time series
- Now we have:



- But still <u>interpretable</u> probe of predictive relationships in data
  - Much richer than IID assumption (e.g. in "standard" entropy calcs.)
  - If starting with joint dist., can probe full conditional/mutual relationships
  - Free from <u>distributional</u> assumptions

Actual Data: Seismic Catalog(s)

- Start with tabular data of seismic events
- Each event has a <u>location</u> (space/time) and <u>magnitude</u>

ID	Timestamp	Lon. (deg)	Lat. (deg)	Dep. (m)	Mag.
1	2025-12-12T10:04:...	1			
2					
3					
4					
5					
⋮					

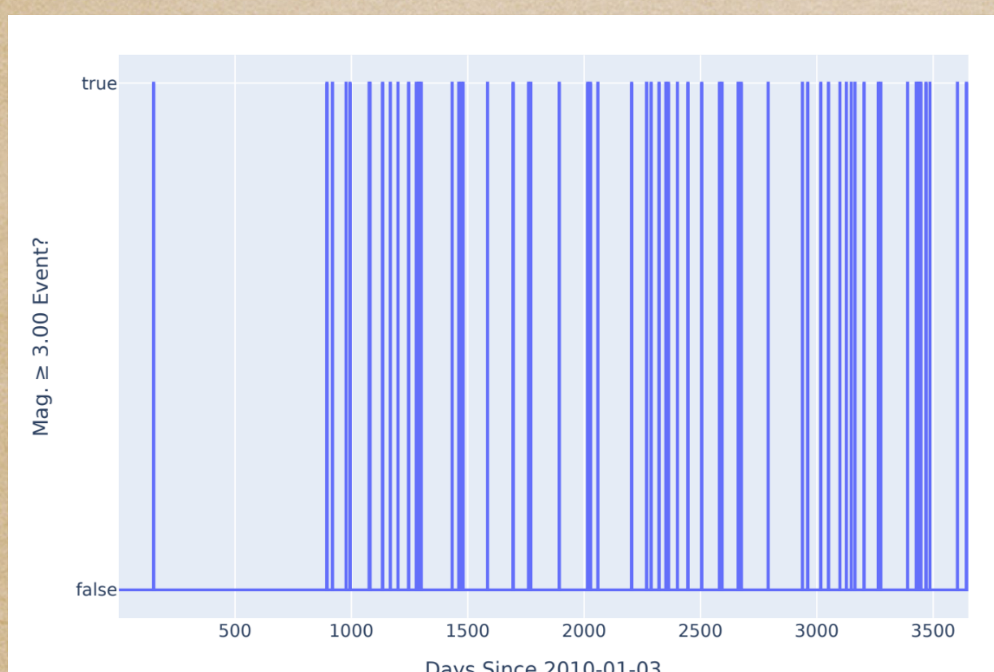
Seismic Catalog(s) → Time Series

ID	Timestamp	Lon. (deg)	Lat. (deg)	Dep. (m)	Mag.
1	2025-12-12T10:04:…	1			
2					
3					
4					
5					
⋮					

Seismic Catalog(s)→Time Series

ID	Timestamp	Lon. (deg)	Lat. (deg)	Dep. (m)	Mag.
1	2025-12-12T10:04:…	1			
2					
3					
4					
5					
⋮					

- Bin times (days, weeks…)
- Select features
  - # Events
  - Mag. Threshold
  - …(choices!)



- Now: can carry out information-anatomic analyses!

The Workflow

Catalog Data

- Example: from Geysers study area ~2010-2019
- Can integrate any w/“standard” data

→

Time Series’

- Bin time for series
- Bin value space as relevant
- Group by location as relevant

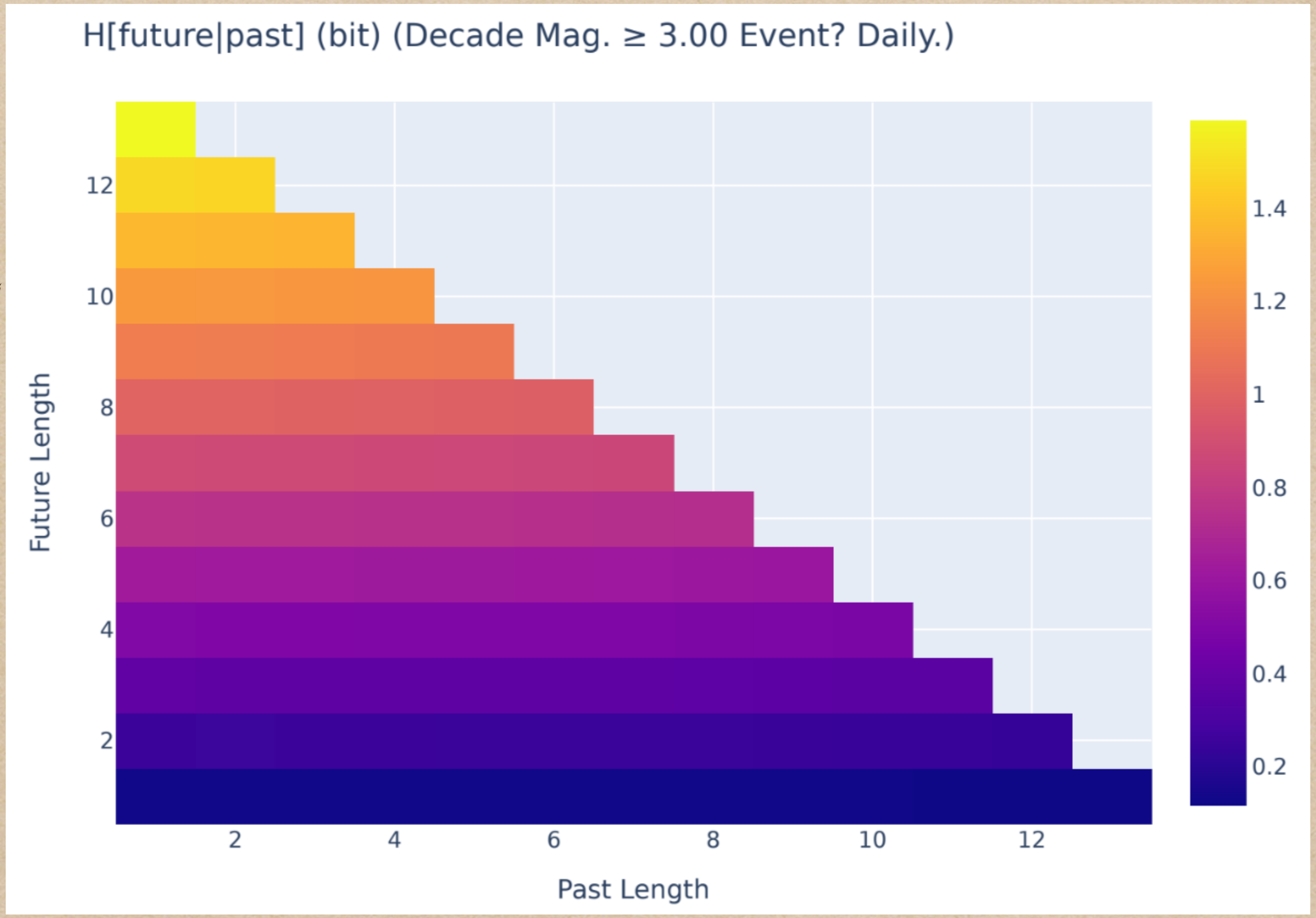
→

Info. Anatomy

- Delay embed TS w/length L
- Build joint probability w/ `ComplexityMeasures.jl`
- Marginalize/condition as desired
- Calc. entropies

Result Highlight I: Cond. Ent.

If we try to forecast this long



Based on this much history

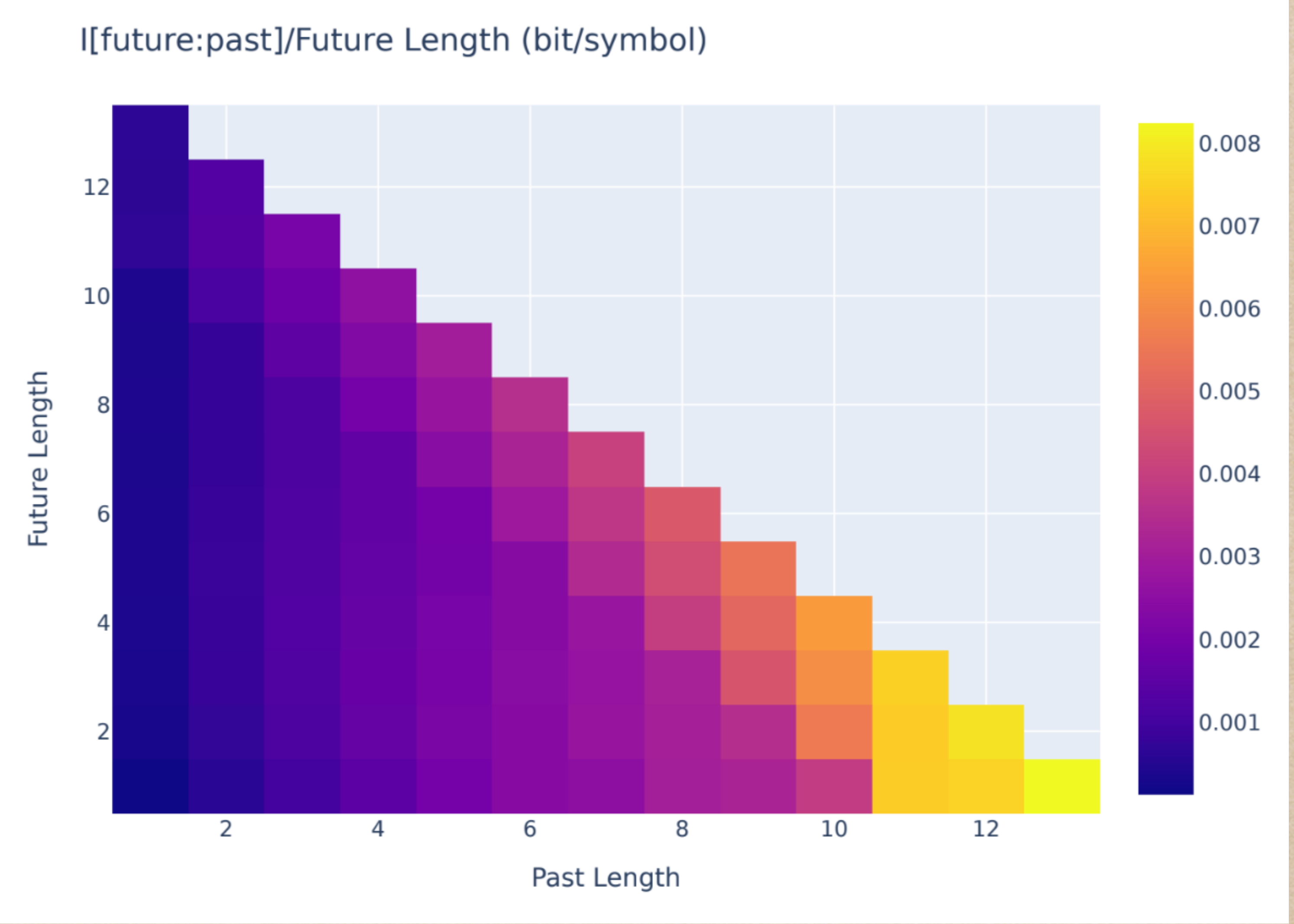
We have this much average forecast uncertainty (bits)

<u>Meaning?</u>

- Longer forecasting "harder" **and**
- Does **not** improve much w/long histories
- Knowing this **saves compute**

Result Highlight I: Cond. Ent.

If we try to forecast this long



<u>Meaning?</u>

- Entire history relevant, but
- *I* extremely low

As before,

- Knowing this **saves compute**

Based on this much history

Past/future **share** this much info. (bits) per time step

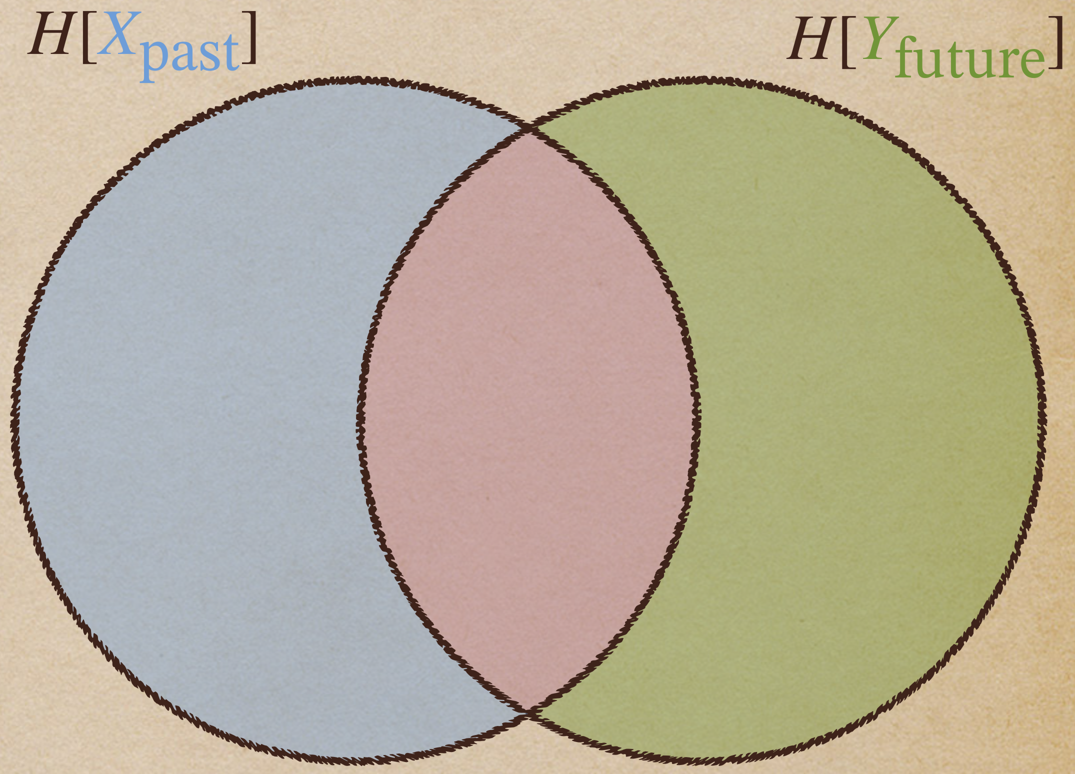
Relevance and Conclusions

- Addresses and complements ML-based projection
- No distributional assumptions
- Directly characterize conditional uncertainties
- Can be used as data exploration step
  to avoid unnecessary compute
- Can be used as validation step on trained model,
  comparing to “raw” data
- Gives physics-based insight <u>**at the level of observed data**</u>

Upcoming Developments

- Dependence of large seismic events on fluid injection activity
- Info. Gained by incorporating microseismic events, even if below mag. of completeness
  - *Without* expensive ML train/run until needed
- Arbitrary bivariate/trivariate correlation (not only future/past)

Multivariate series:
seismic, injection, external, …

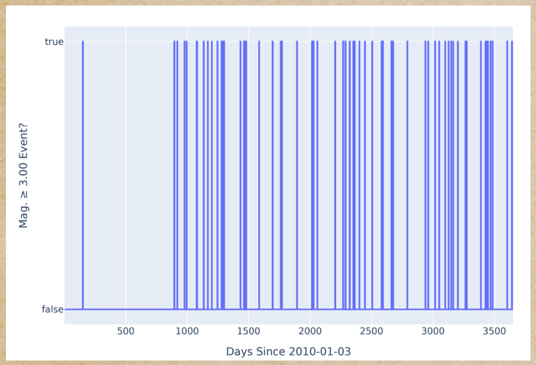


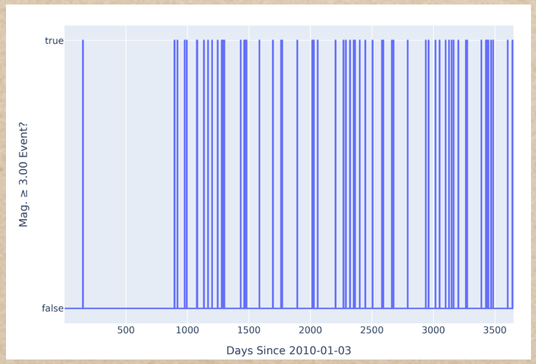
Between arbitrary variable blocks!

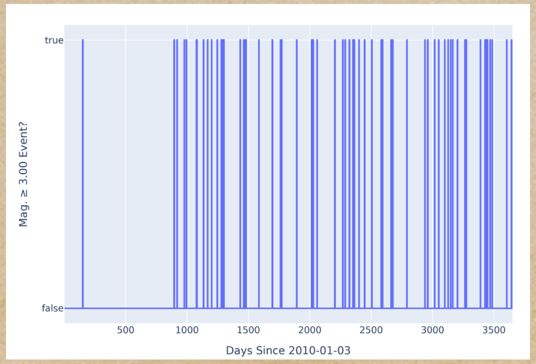
Seismic Catalog(s) → Time Series'

ID	Timestamp	Lon. (deg)	Lat. (deg)	Dep. (m)	Mag.
1	2025-12-12T10:04:...	1			
2					
3					
4					
5					
⋮					

- Bin times (days, weeks…)
- **Group by locations**
- Select features
  - # Events
  - Mag. Threshold
  - …(choices!)





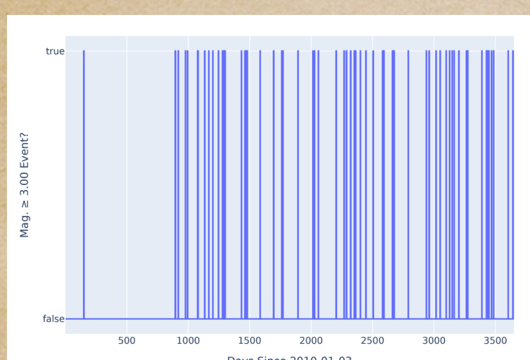


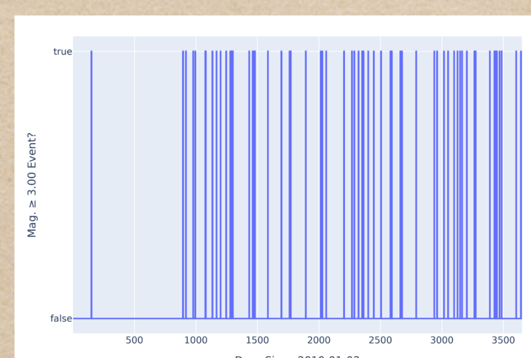
- Now: can probe **arbitrary space/time relationships!**

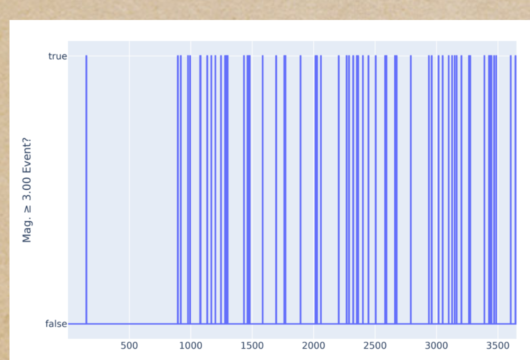
Seismic Catalog(s) → Time Series'

ID	Timestamp	Lon. (deg)	Lat. (deg)	Dep. (m)	Mag.
1	2025-12-12T10:04:...	1			
2					
3					
4					
5					
⋮					

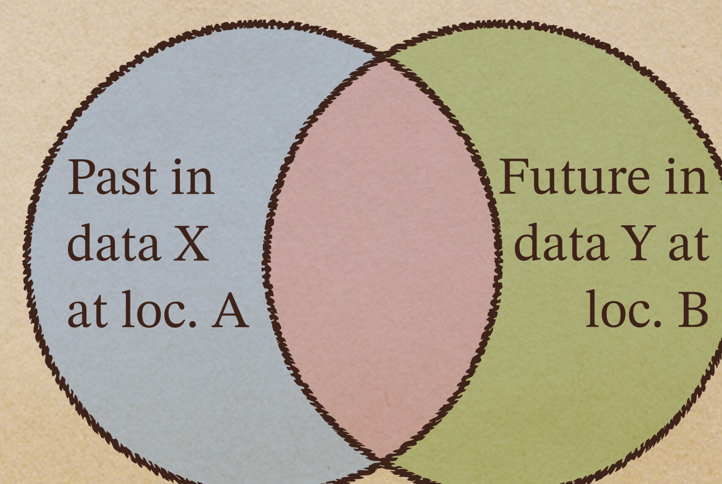
- Bin times (days, weeks...)
- **Group by locations**
- Select features
  - # Events
  - Mag. Threshold
  - ...(choices!)







- Ex:

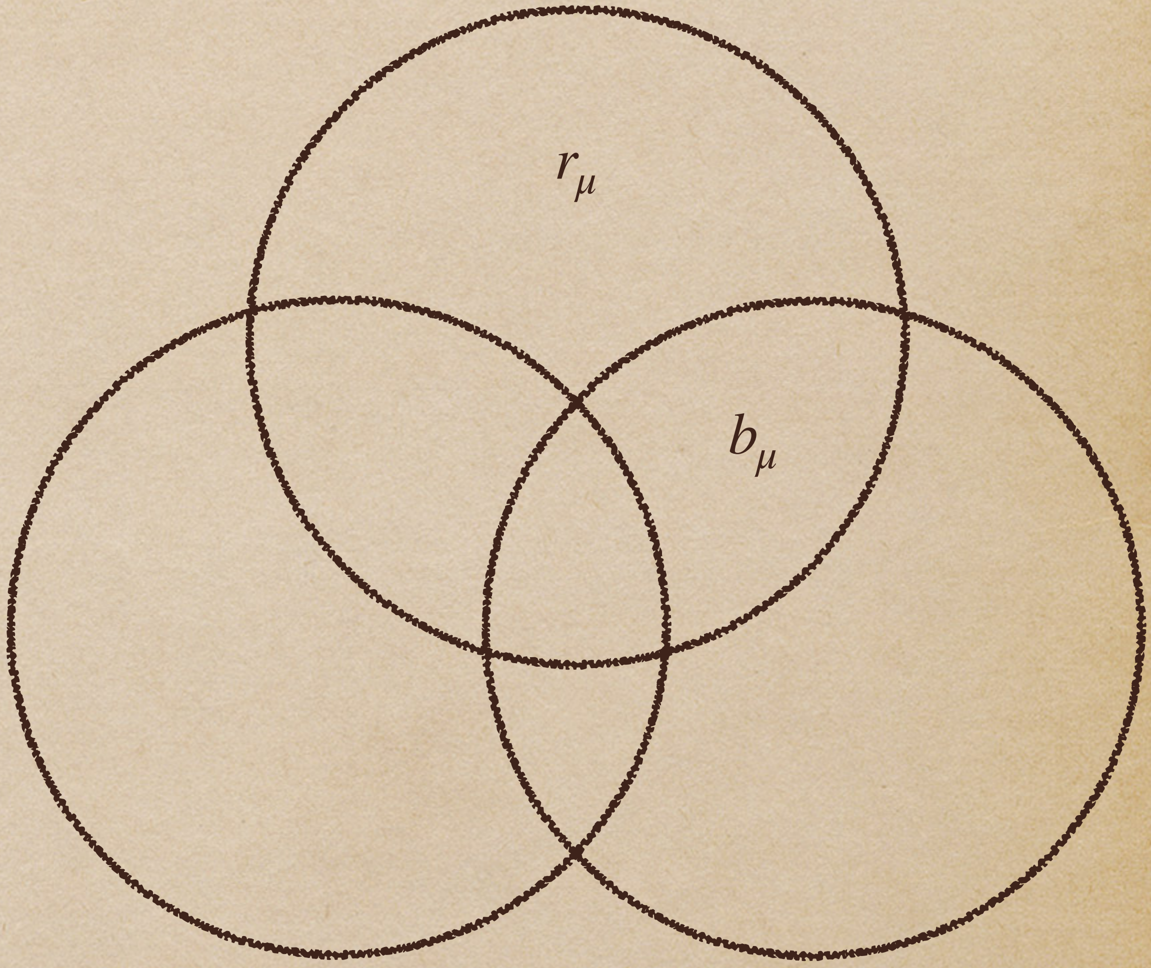


EVERYTHING PAST HERE NOT FOR TALK

Inspo: Anatomy of a Bit (will likely spruce up and put brief story as ~~backup in case they're bored with bivariate)~~

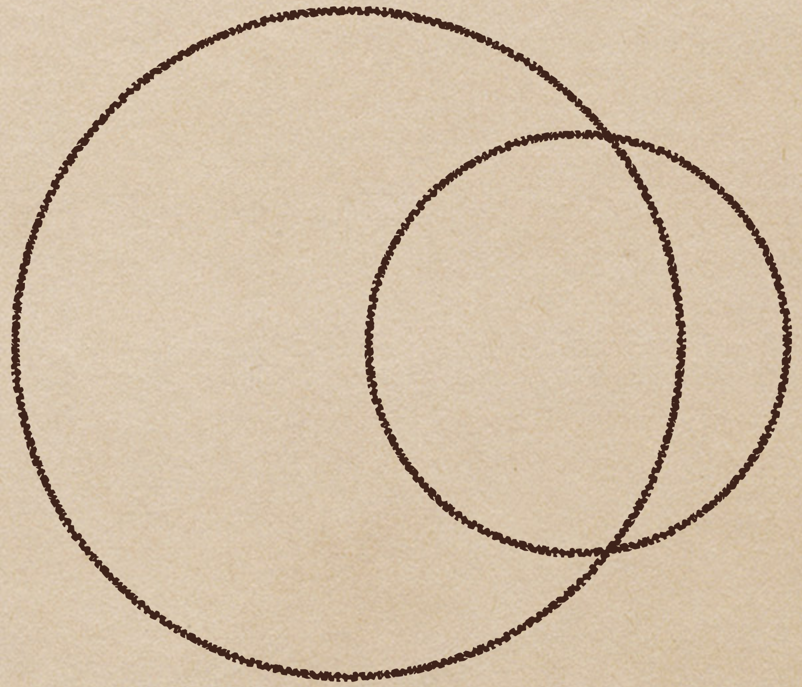
- I-Diagram visualization
- "Information atoms"

Atom	Defn.	Meaning
r_μ	$H[Z_0 \mid Z_{:-1}, Z_{1:}]$	Info. in present, **not** shared w/past or future
b_μ	$I[Z_0 : Z_{1:} \mid Z_{:-1}]$	Info. in present, **is** shared w/future
h_μ	$r_\mu + b_\mu$	New info. in present (Shannon entropy rate)

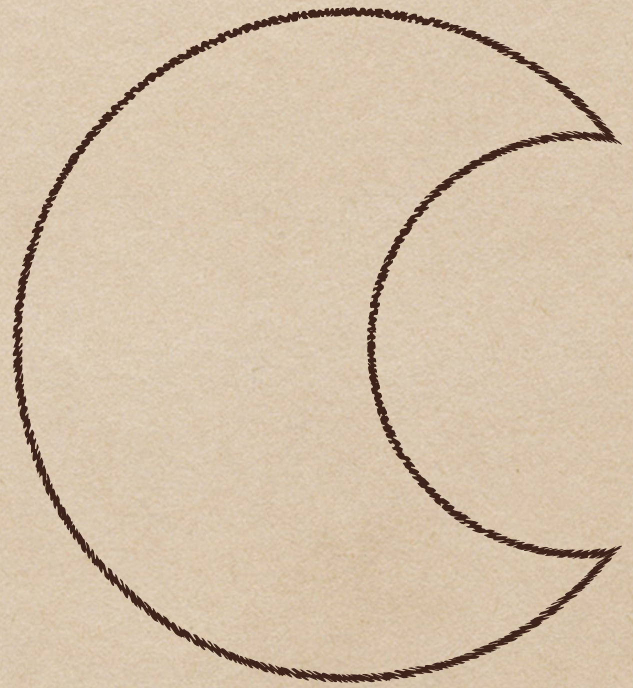


- Details in James *et al.*
- Main ideas: taxonomize <u>temporal</u> and <u>spatial</u> (and <u>both</u>) time series structure

Slide Support Shapes



Slide Support Shapes



Slide Support Shapes

Slide Support Shapes