Workshop Bayesian Corpus Studies

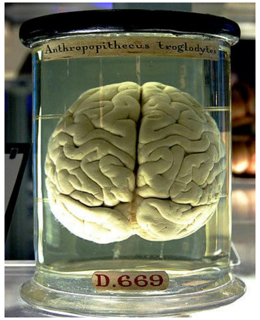
Christoph Finkensiep

Würzburg, Feb 2024

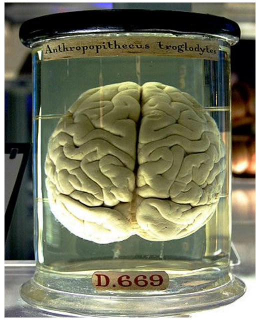
Session 1: Generative Modeling

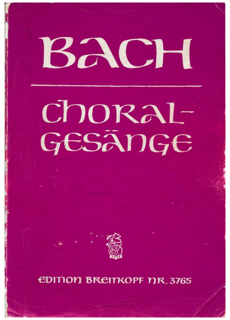
Motivation

Corpus Studies

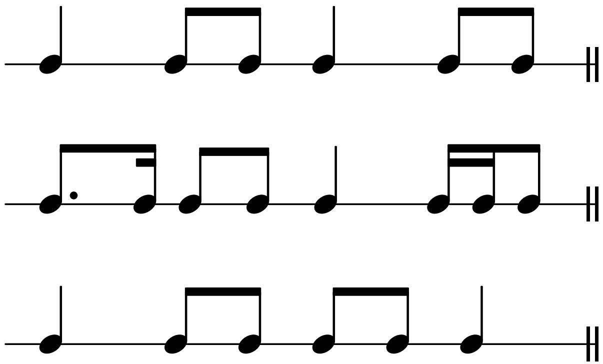


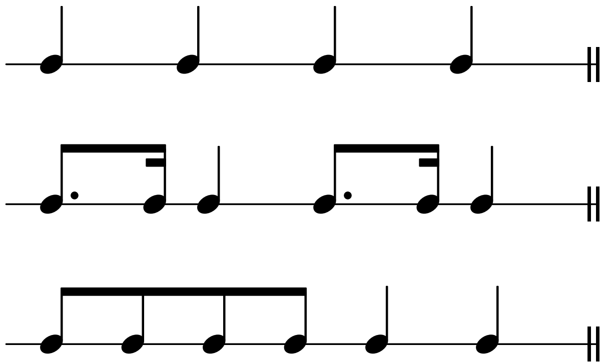
Corpus Studies





A Small Corpus Study





In which group are the patterns more regular?

Corpus Studies

The classical approach:

1. find / create a corpus
2. “operationalize” the quantity of interest
3. measure, do statistics

$$\begin{tikzcd}
\text{conclusions} \\
\text{observations} \arrow[u, "\textit{operationalize}"']
\end{tikzcd}$$

Corpus Studies

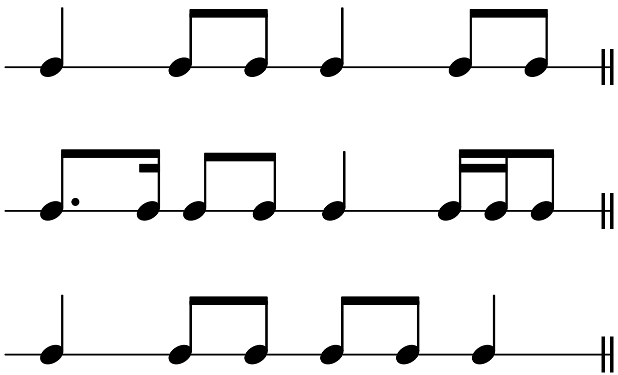
The classical approach:

1. find / create a corpus
2. "operationalize" the quantity of interest
3. measure, do statistics

conclusions

$\uparrow$ *operationalize*

observations



Corpus Studies

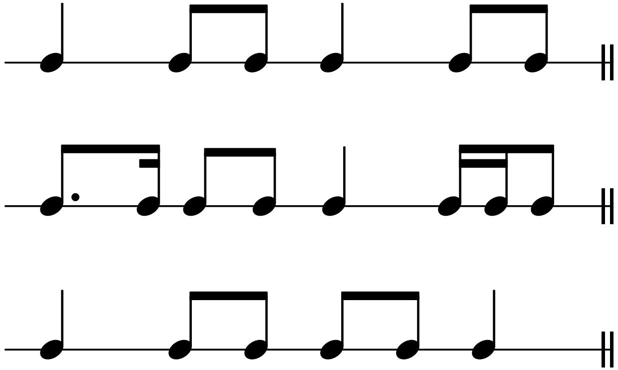
The classical approach:

1. find / create a corpus
2. $\boxed{\text{“operationalize” the quantity of interest}}$???
3. measure, do statistics

conclusions

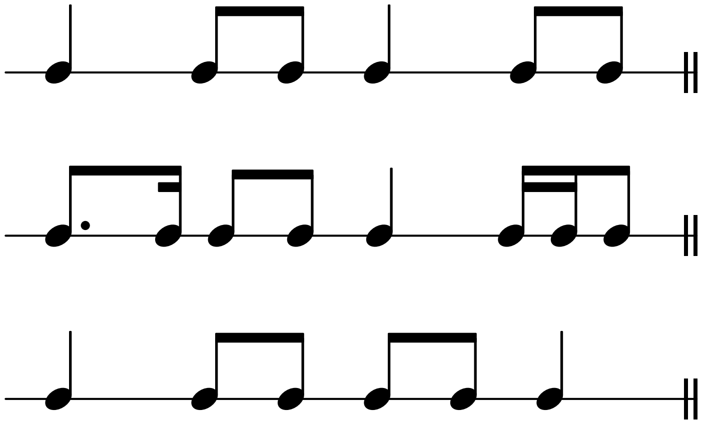
operationalize

observations



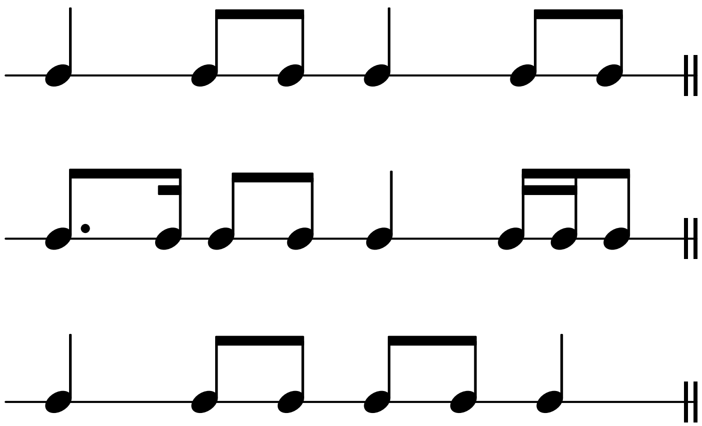
Problems with the Standard Approach

$$\mathrm{reg}(x) = \frac{1}{\#\ \text{unique beats in } x}$$



Problems with the Standard Approach

$$\mathrm{reg}(x) = \frac{1}{\#\ \text{unique beats in}\ x}$$

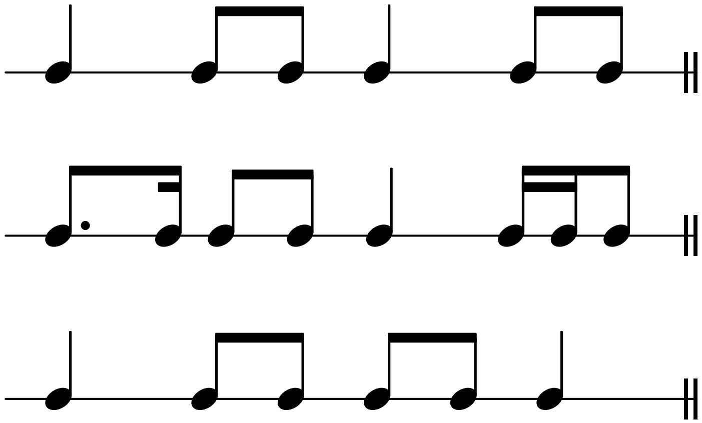


Measure:

- adequate?
- arbitrary, ad-hoc?
- overarching theory?

Problems with the Standard Approach

$$\mathrm{reg}(x) = \frac{1}{\#\ \text{unique beats in } x}$$



Measure:

- adequate?
- arbitrary, ad-hoc?
- overarching theory?

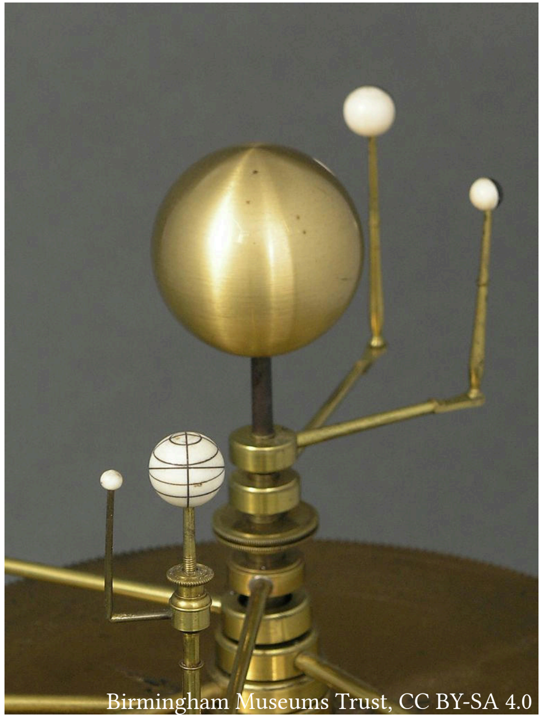
Stats:

- which statistic?
  - arithmetic mean: $\frac{1}{N}\sum_i x_i$
  - geometric mean: $\sqrt[N]{\prod_i x_i}$
- which test?

Alternative: Work with Models

Alternative: Work with Models

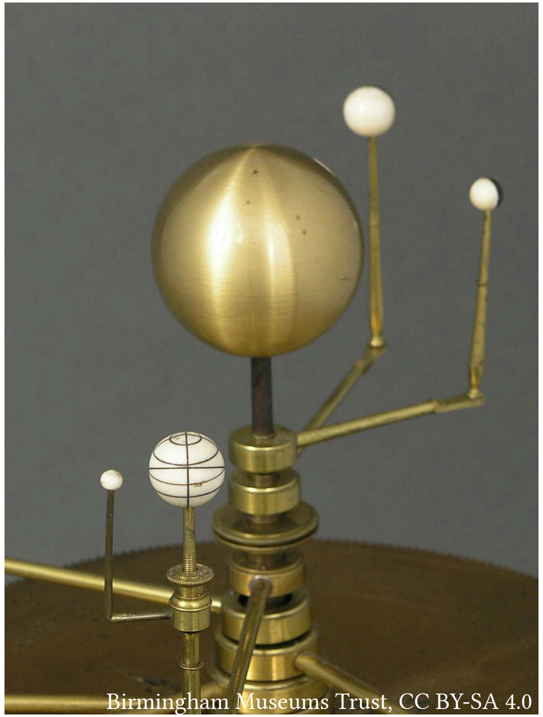
Models


Birmingham Museums Trust, CC BY-SA 4.0

Alternative: Work with Models

Models

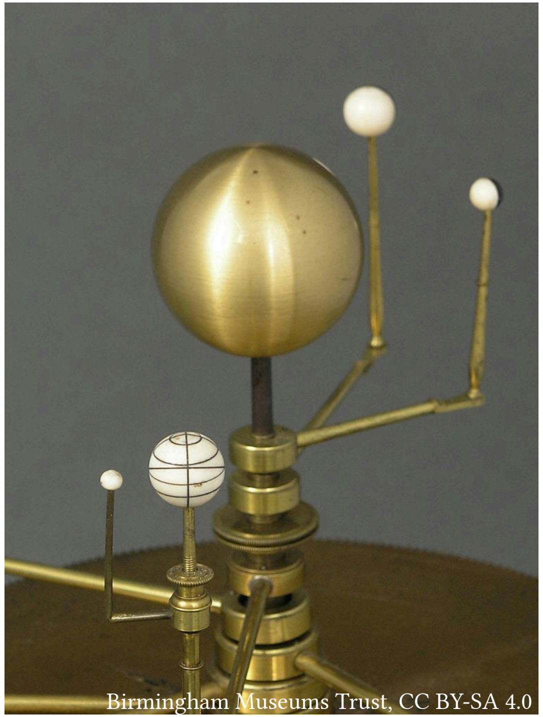
- describe a segment of the world (simplified)


Birmingham Museums Trust, CC BY-SA 4.0

Alternative: Work with Models

Models

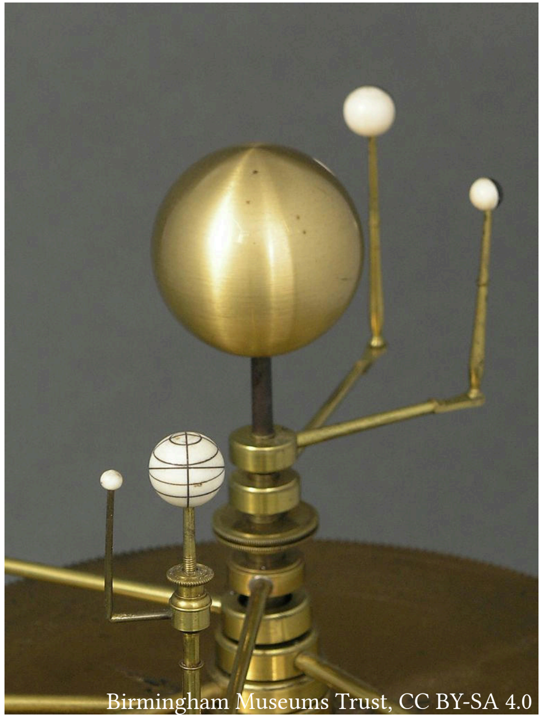
- describe a segment of the world (simplified)
- relevant **entities**, **properties** and **relations**



Alternative: Work with Models

Models

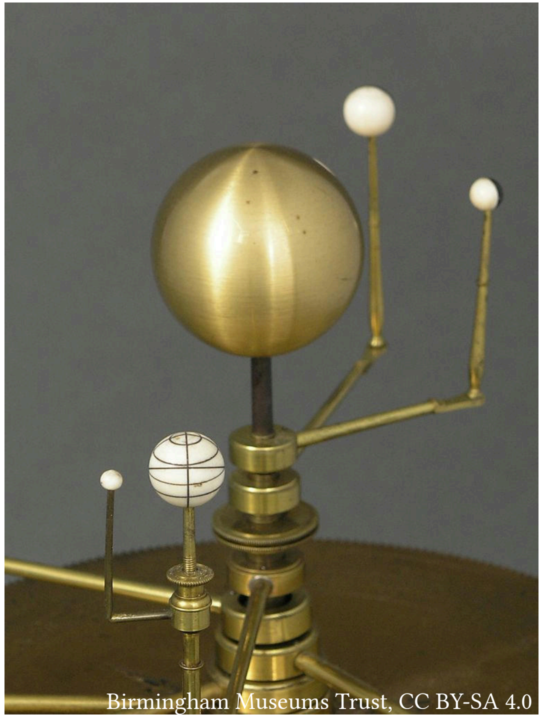
- describe a segment of the world (simplified)
- relevant **entities**, **properties** and **relations**
- explicit **assumptions**



Alternative: Work with Models

Models

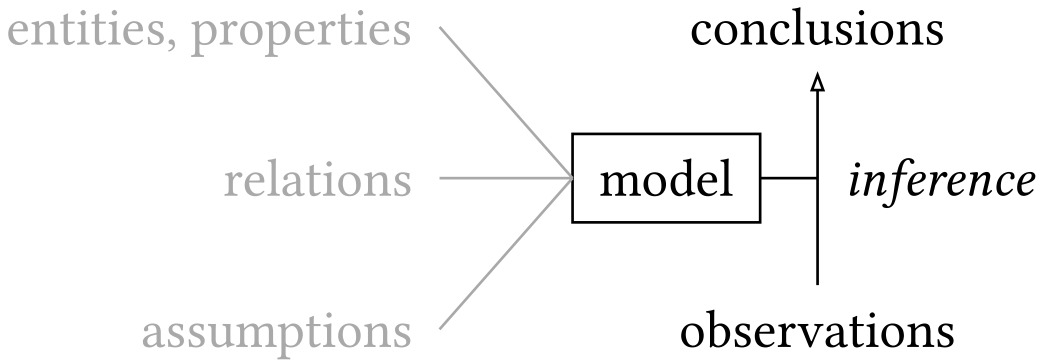
- describe a segment of the world (simplified)
- relevant **entities**, **properties** and **relations**
- explicit **assumptions**
- enable **simulation** and **inference**

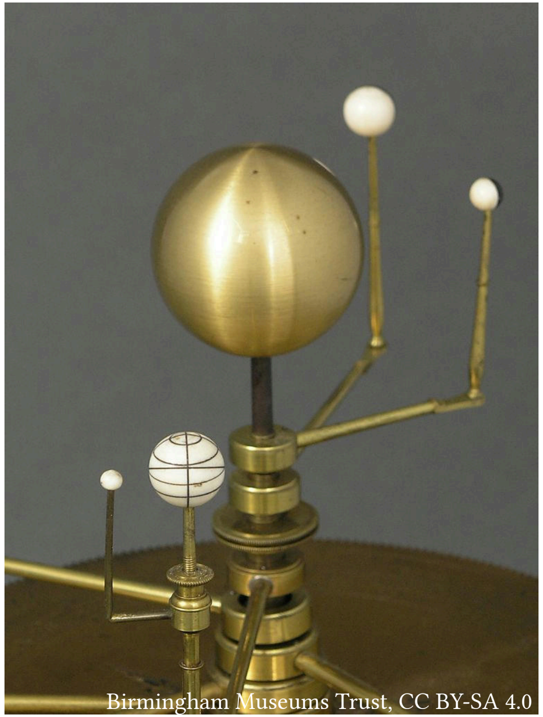


Alternative: Work with Models

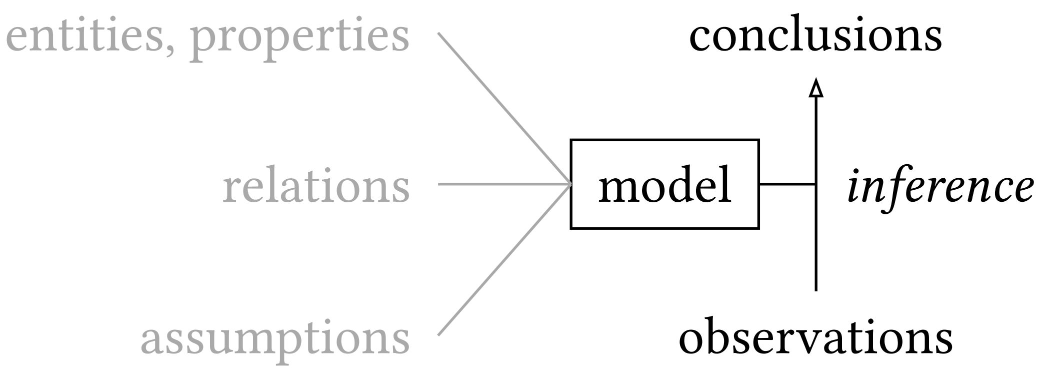
Models

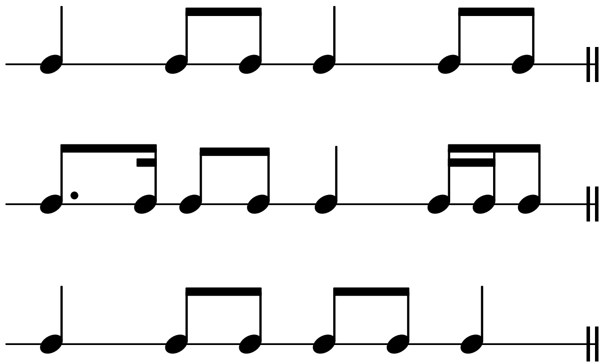
- describe a segment of the world (simplified)
- relevant **entities**, **properties** and **relations**
- explicit **assumptions**
- enable **simulation** and **inference**



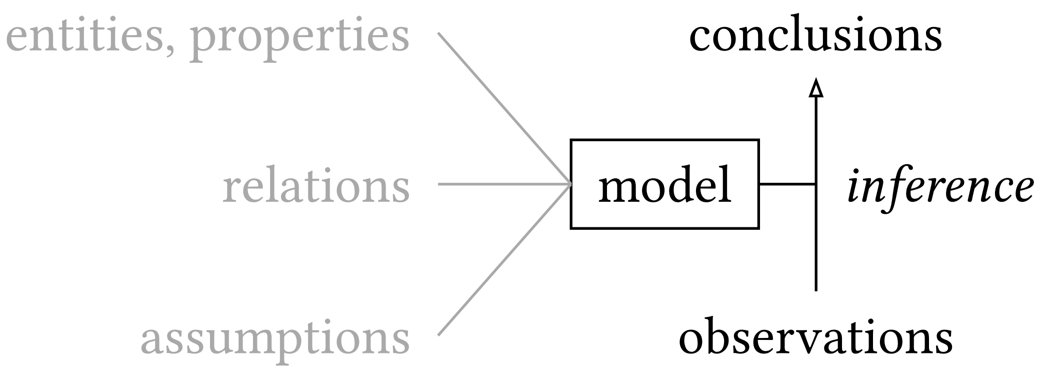


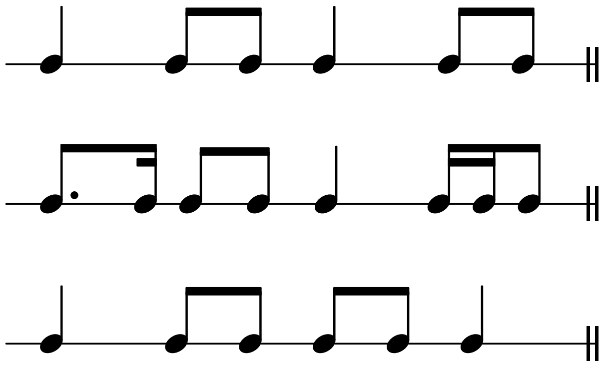
A Model of Rhythmic Regularity





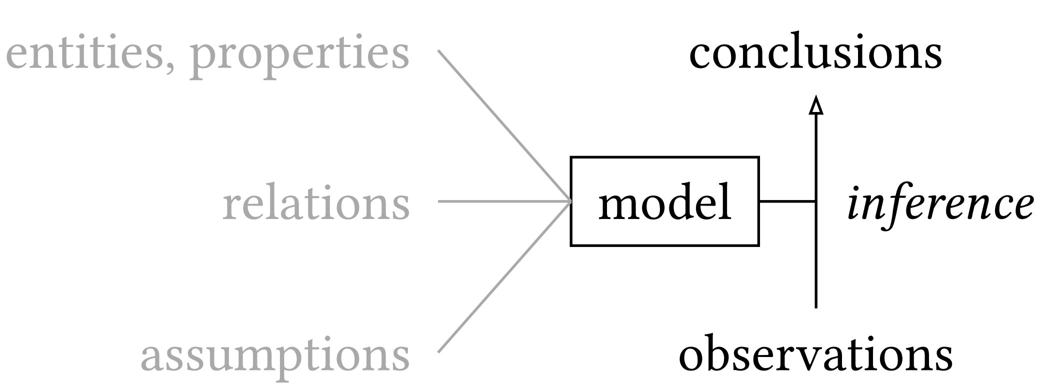
A Model of Rhythmic Regularity

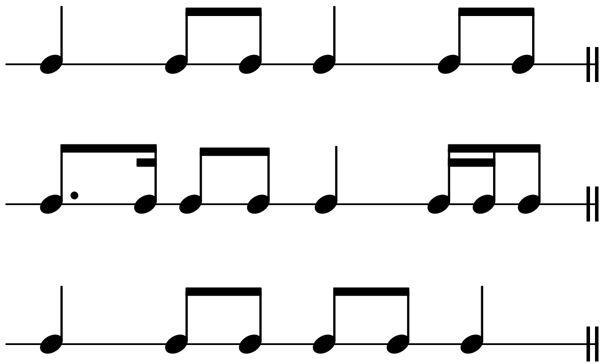




Entities and properties:

A Model of Rhythmic Regularity

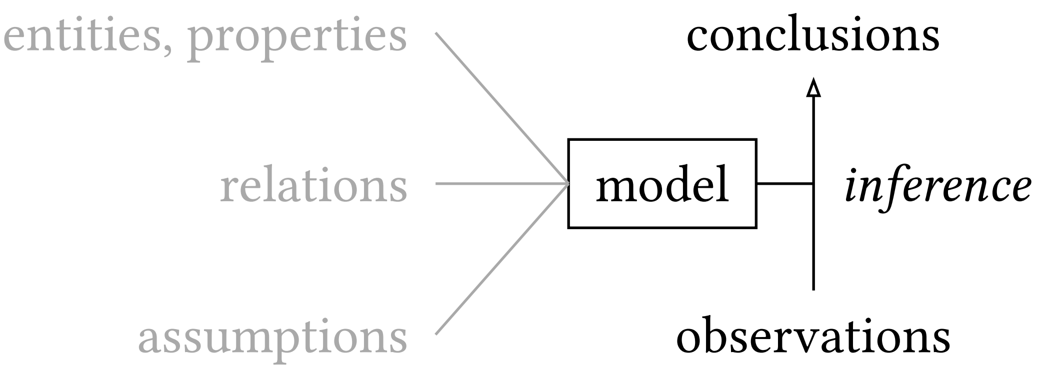


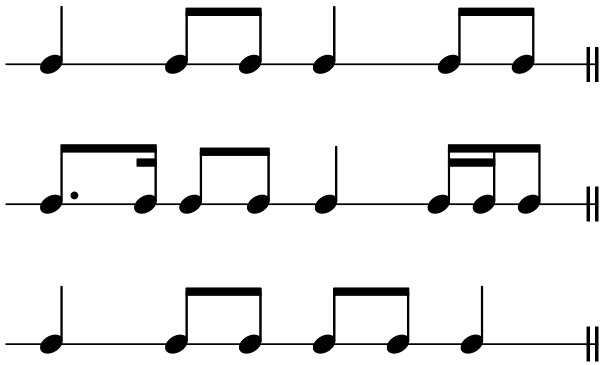


Entities and properties:

- pattern
- note
- group of patterns
  - regularity

A Model of Rhythmic Regularity



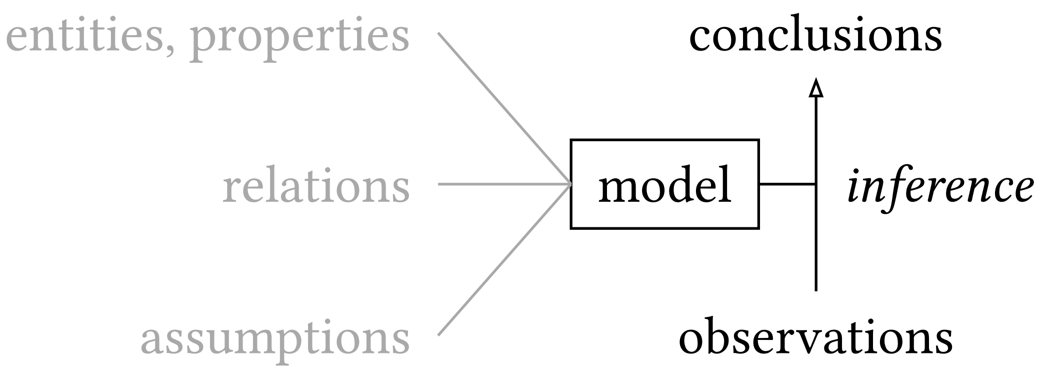


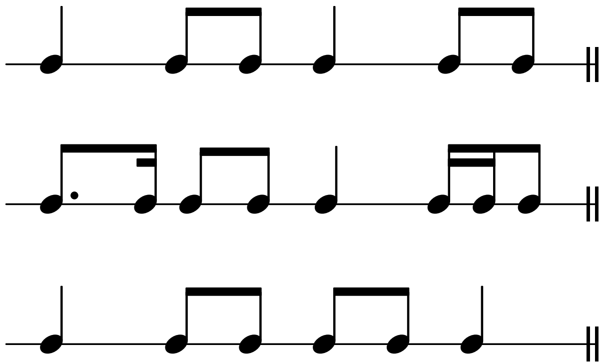
Entities and properties:

- pattern
- note
- group of patterns
  - regularity

Relations:

A Model of Rhythmic Regularity





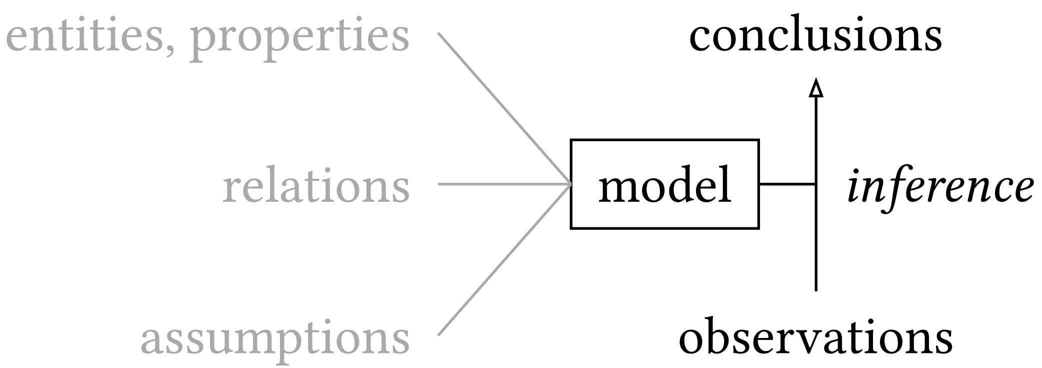
Entities and properties:

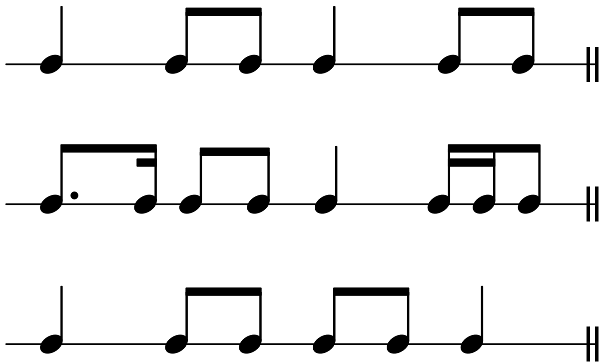
- pattern
- note
- group of patterns
  - regularity

Relations:

- patterns consist of notes
- groups have patterns
- regularity ↔ patterns?

A Model of Rhythmic Regularity





Entities and properties:

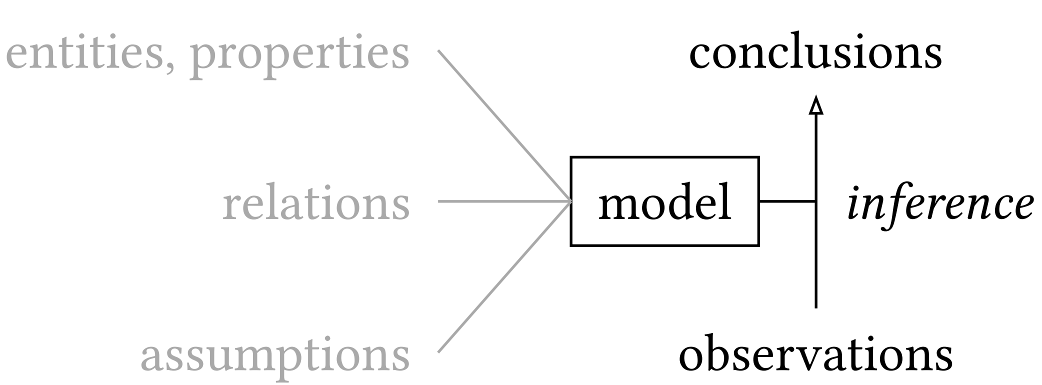
- pattern
- note
- group of patterns
  - regularity

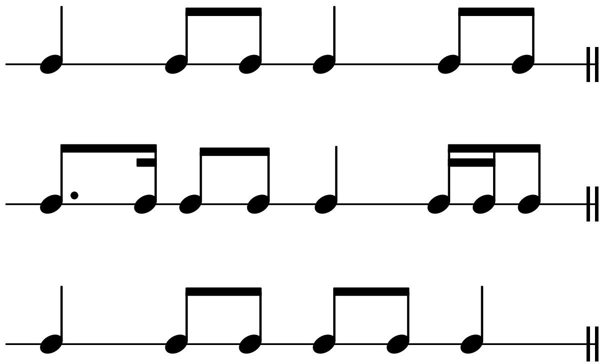
Relations:

- patterns consist of notes
- groups have patterns
- regularity $\leftrightarrow$ patterns?

Inference:

A Model of Rhythmic Regularity





Entities and properties:

- pattern
- note
- group of patterns
  - regularity

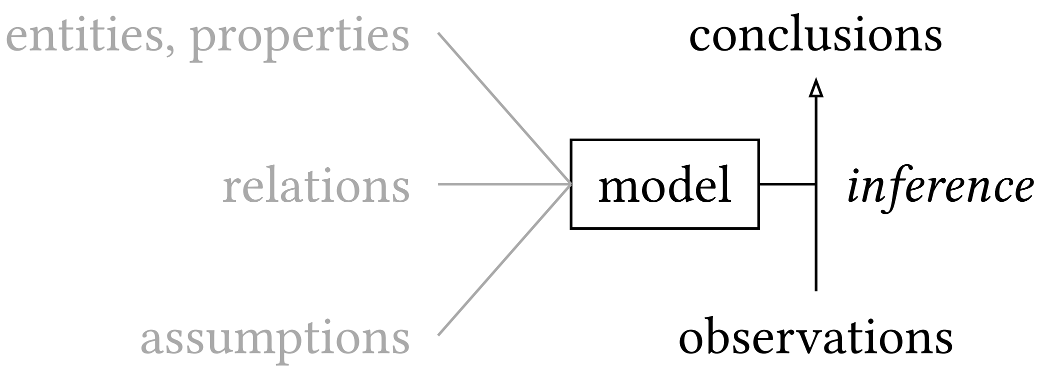
Relations:

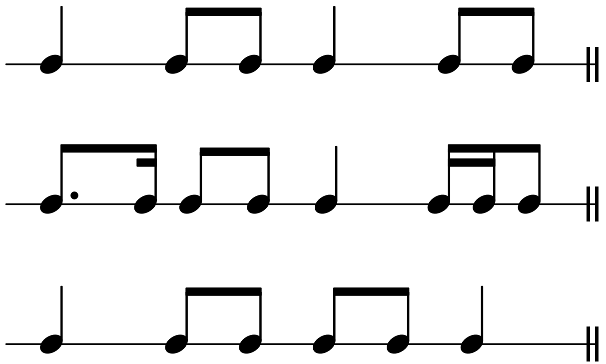
- patterns consist of notes
- groups have patterns
- regularity ↔ patterns?

Inference:

- patterns → regularity

A Model of Rhythmic Regularity





Entities and properties:

- pattern
- note
- group of patterns
  - regularity

Relations:

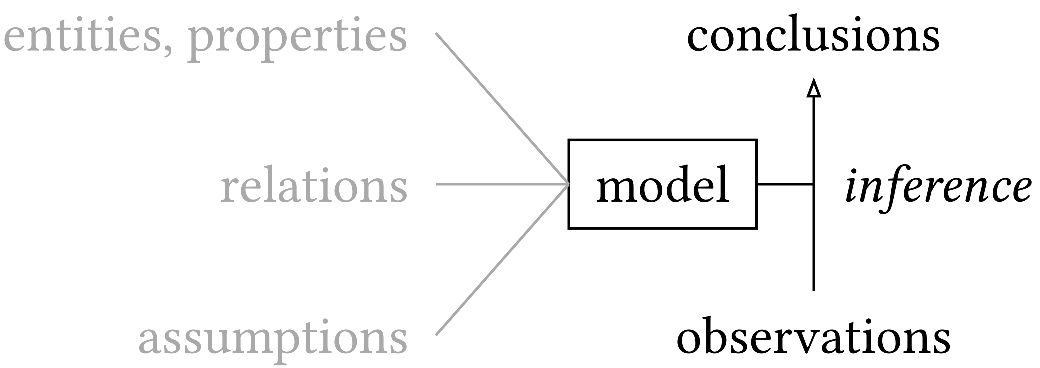
- patterns consist of notes
- groups have patterns
- regularity $\leftrightarrow$ patterns?

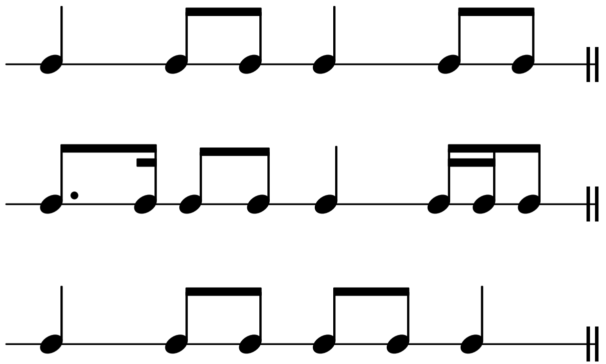
Inference:

- patterns $\rightarrow$ regularity

Simulation:

A Model of Rhythmic Regularity





Entities and properties:

- pattern
- note
- group of patterns
  - regularity

Relations:

- patterns consist of notes
- groups have patterns
- regularity ↔ patterns?

Inference:

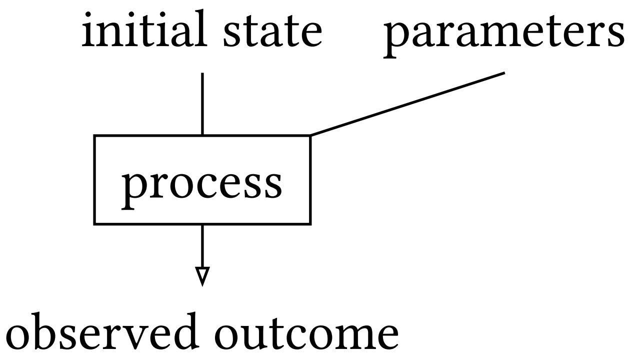
- patterns → regularity

Simulation:

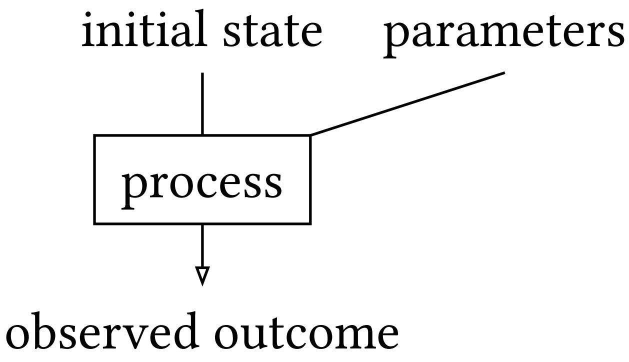
- generate new patterns

Generative Models

The Process Behind the Data



The Process Behind the Data



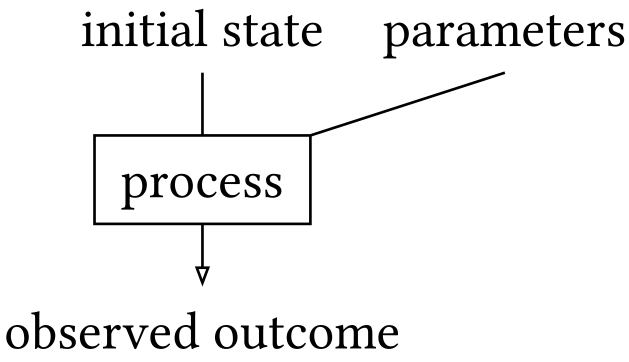
Simulation:

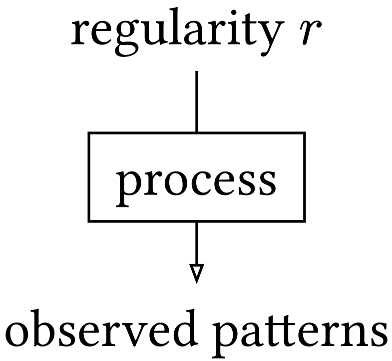
- run the process, change parameters

Inference:

- find plausible parameters

The Process Behind the Data





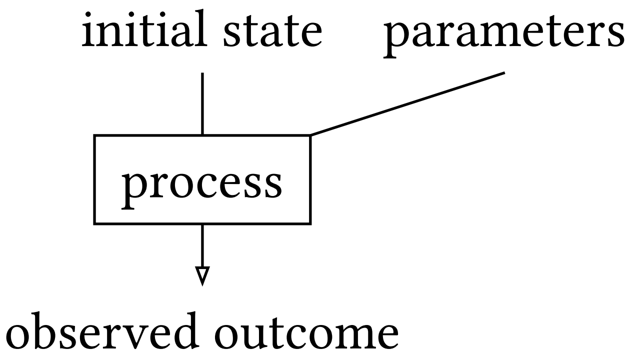
Simulation:

- run the process, change parameters

Inference:

- find plausible parameters

The Process Behind the Data

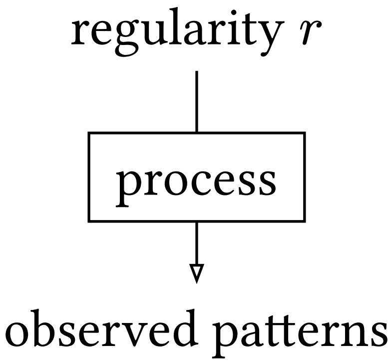


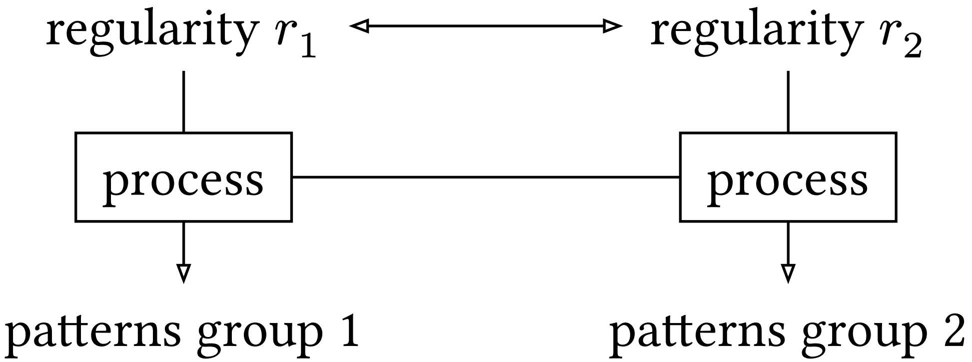
Simulation:

- run the process, change parameters

Inference:

- find plausible parameters

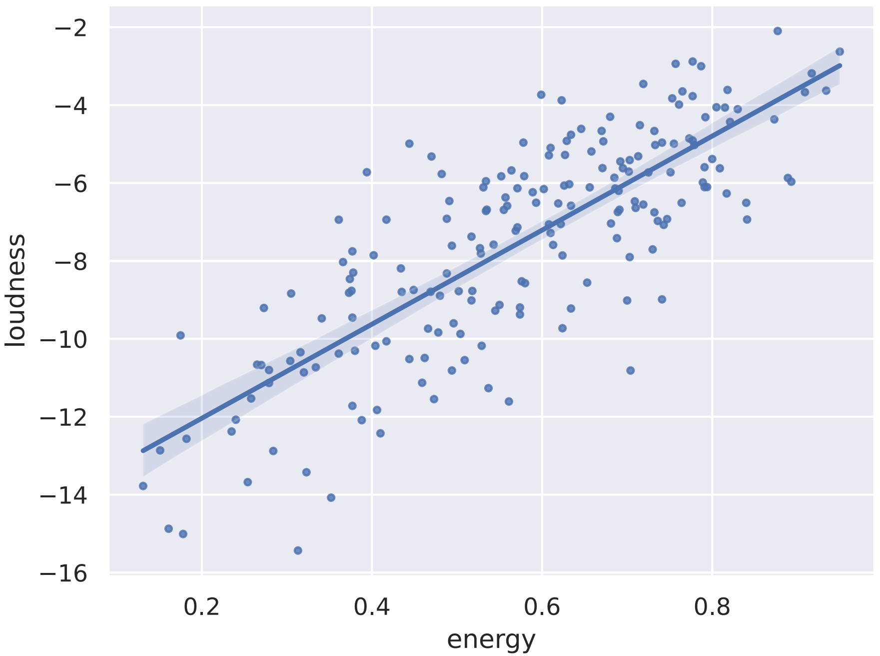




A Simple Example

Linear Regression: x is linearly related to y

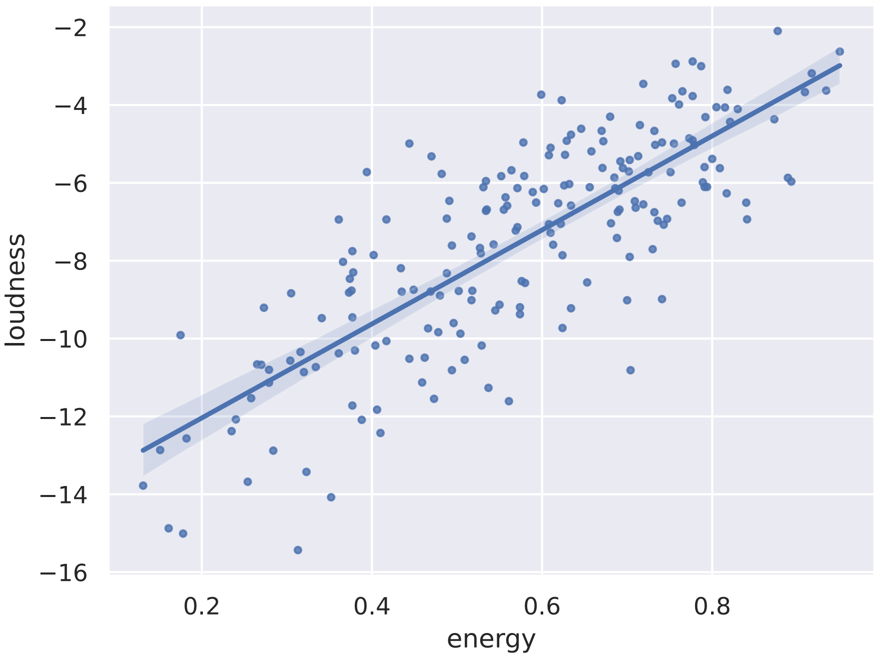
$$y \approx a \cdot x + b$$



A Simple Example

Linear Regression: x is linearly related to y

$$y \approx a \cdot x + b$$



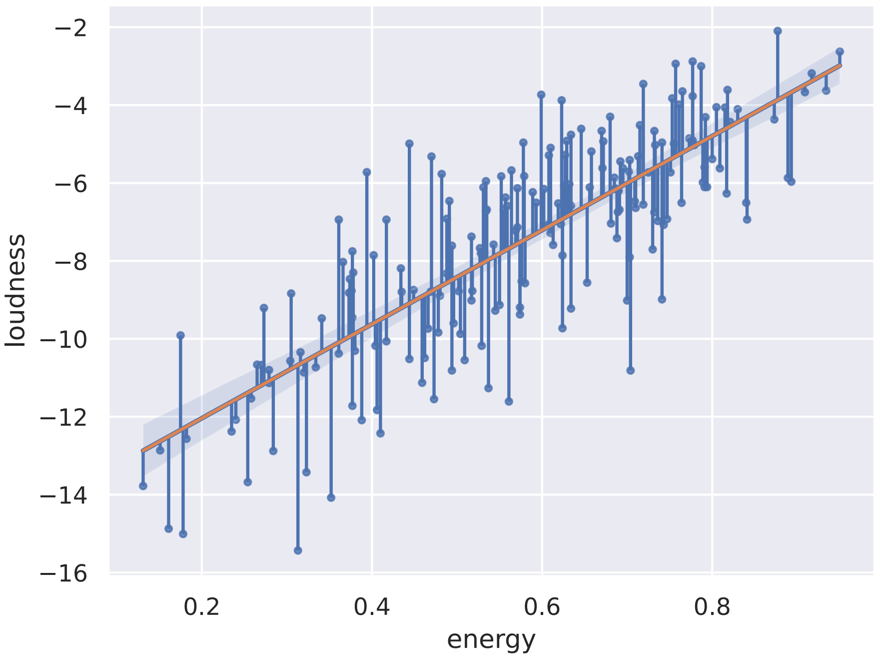
Observations: points (x and y)

Parameters: slope a, intercept b, variance σ^2

A Simple Example

Linear Regression: x is linearly related to y

$$y \approx a \cdot x + b$$



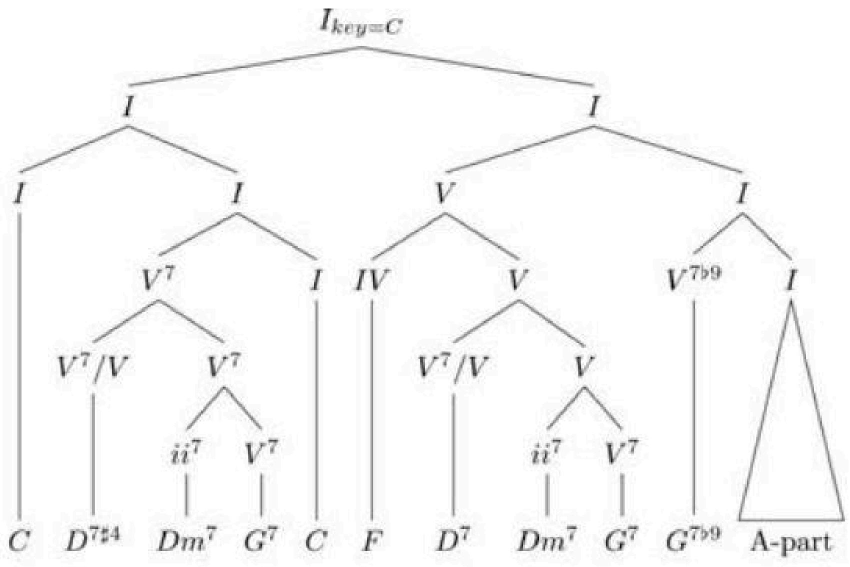
Observations: points (x and y)

Parameters: slope a, intercept b, variance σ^2

Generative Process:

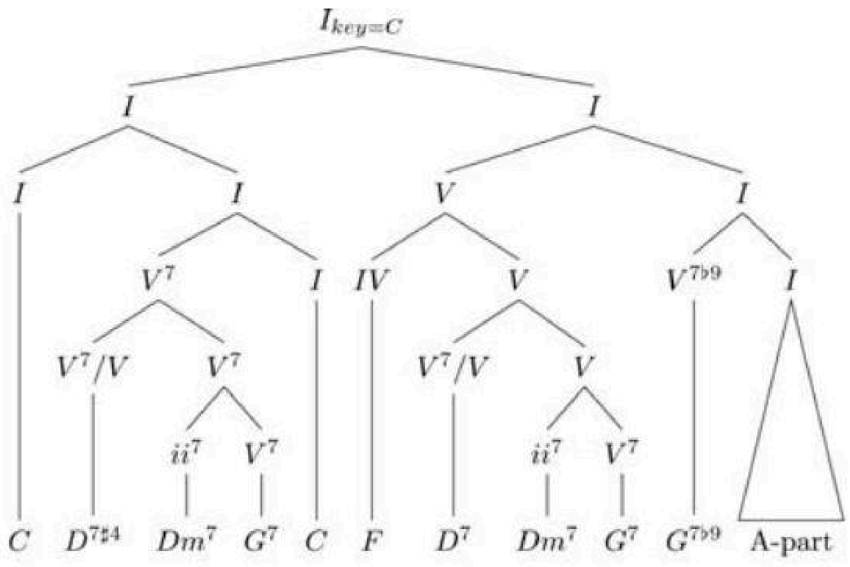
- choose parameters a, b, σ
- for each point i:
  - choose x_i
  - compute $f(x_i) = a \cdot x_i + b$
  - pick $y_i \sim \mathcal{N}(f(x_i), \sigma)$

A More Involved Example



$$I \longrightarrow I\ I$$
$$I \longrightarrow V\ I$$
$$\ldots$$

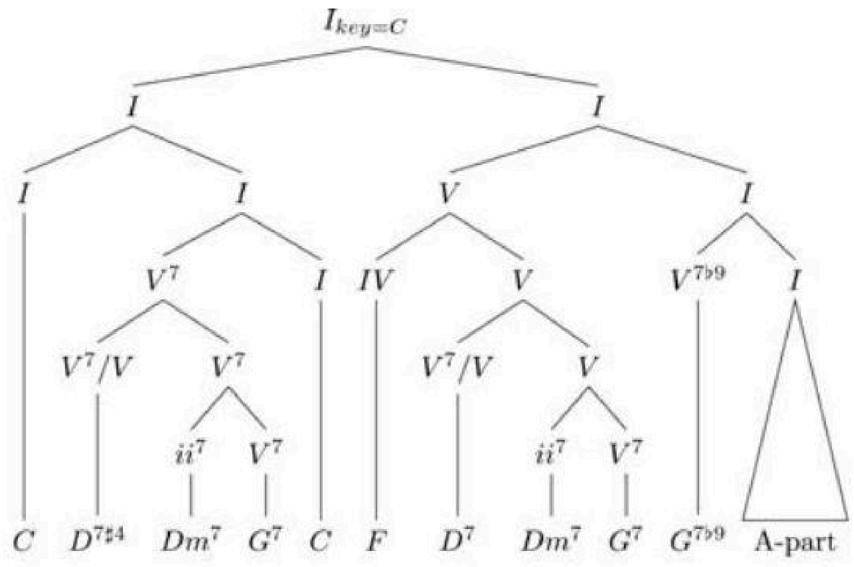
A More Involved Example



$$I \longrightarrow I\ I$$
$$I \longrightarrow V I$$
$$\ldots$$

Observations: chord sequences (pieces)

A More Involved Example

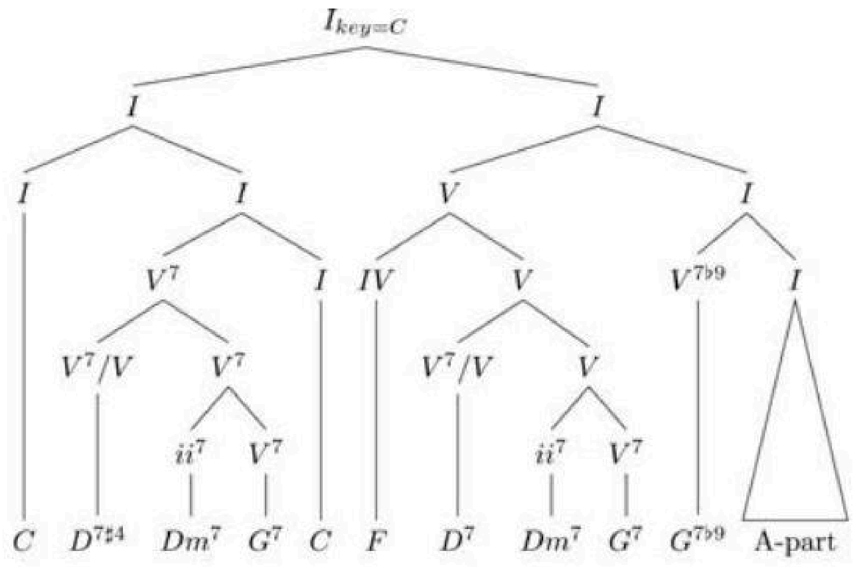


$$I \longrightarrow I\ I$$
$$I \longrightarrow V\,I$$
$$\ldots$$

Observations: chord sequences (pieces)

Parameters: Rules and probabilities

A More Involved Example



$$I \longrightarrow I\ I$$

$$I \longrightarrow V I$$

$\dots$

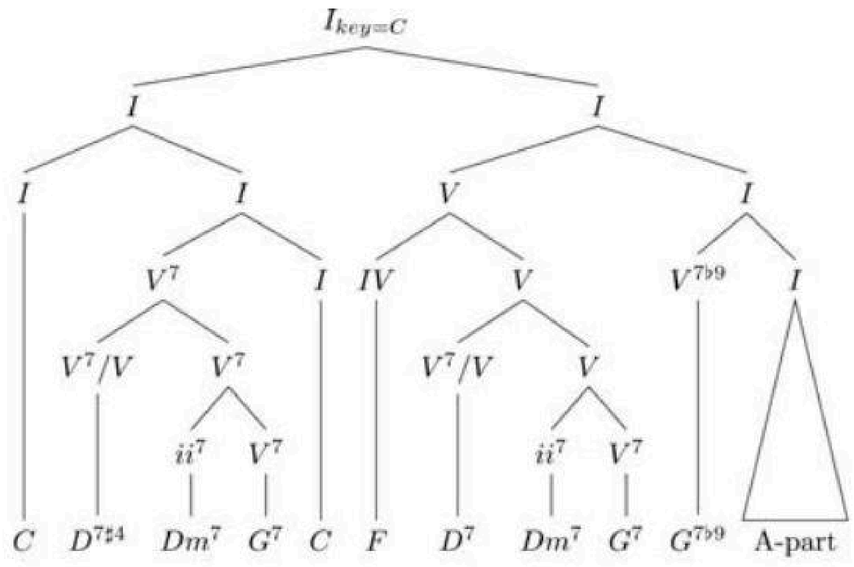
Observations: chord sequences (pieces)

Parameters: Rules and probabilities

Process:

- choose grammar rules R
- choose rule probabilities p_R
- for each piece i:
  - sample a derivation d_i
  - observe the resulting chord sequence

A More Involved Example



$$I \longrightarrow I\ I$$
$$I \longrightarrow V\, I$$
$$\cdots$$

Observations: chord sequences (pieces)

Parameters: Rules and probabilities

Process:

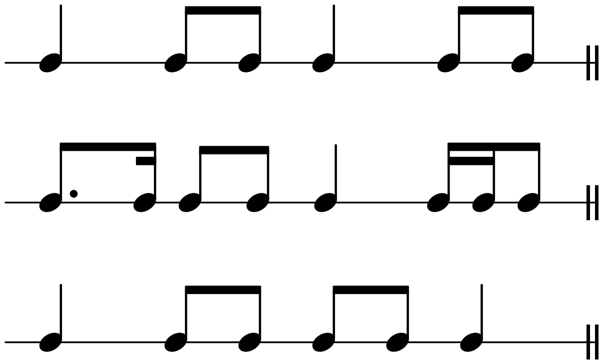
- choose grammar rules R
- choose rule probabilities p_R
- for each piece i:
  - sample a derivation d_i
  - observe the resulting chord sequence

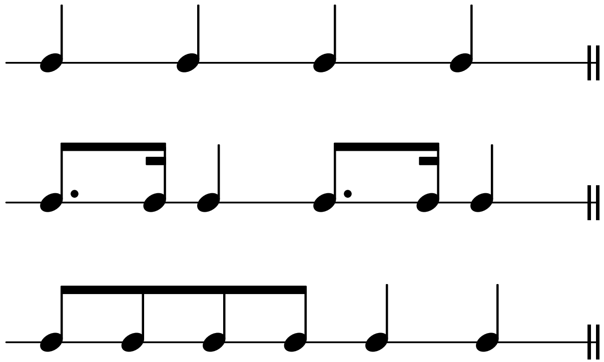
Inference:

- most plausible derivation d_i for each piece
- most plausible rules and probabilities

A *Generative* Model of Rhythmic Regularity

Generate this!





`abab` `aaaa`

`cbad` `caca`

`abba` `bbaa`

Simplifying assumptions: always 4/4, we don't look inside "beats".

Possible Solutions

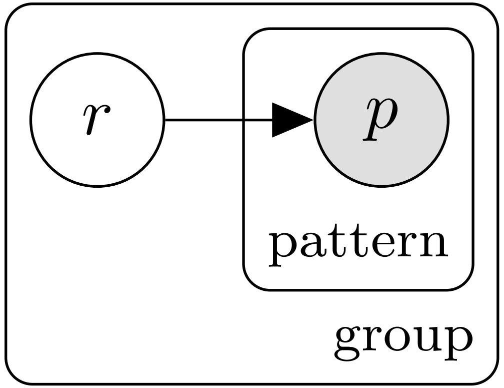
Generate corpus:

- for each group g:
  - choose regularity r_g:
  - for each pattern i:
    - sample pattern p_{gi} using r_g

Possible Solutions

Generate corpus:

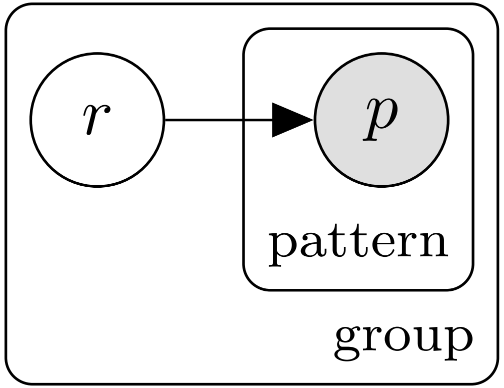
- for each group g:
  - choose regularity r_g:
  - for each pattern i:
    - sample pattern p_{gi} using r_g



Possible Solutions

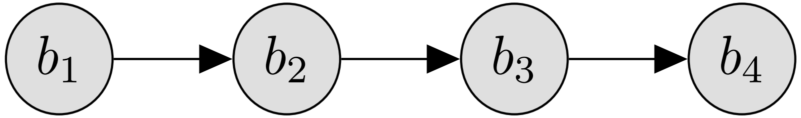
Generate corpus:

- for each group g:
  - choose regularity r_g:
  - for each pattern i:
    - sample pattern p_{gi} using r_g



Generate pattern (based on predecessor):

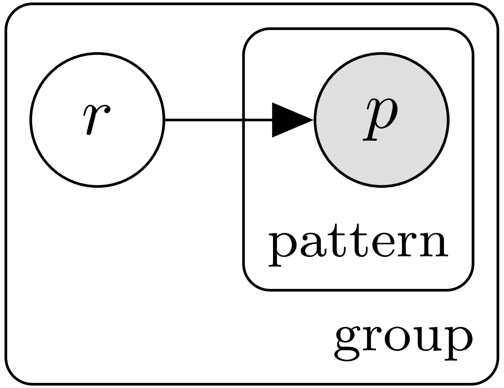
- choose b_1 randomly
- for each following beat i:
  - flip coin (r_g):
    - heads: $b_i = b_{i-1}$
    - tails: choose new beat for b_i



Possible Solutions

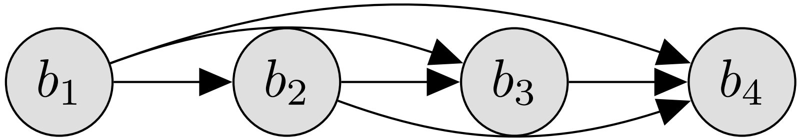
Generate corpus:

- for each group g:
  - choose regularity r_g:
  - for each pattern i:
    - sample pattern p_{gi} using r_g



Generate pattern (based on position):

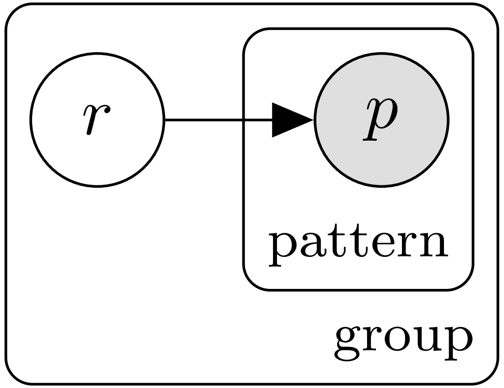
- choose b_1 randomly
- flip coin (r_g):
  - heads: $b_2 = b_1$
  - tails: new beat for b_2
- flip coin (r_g):
  - heads: repeat first half
  - tails:
    - repeat or new b_3 ...
    - repeat or new b_4 ...



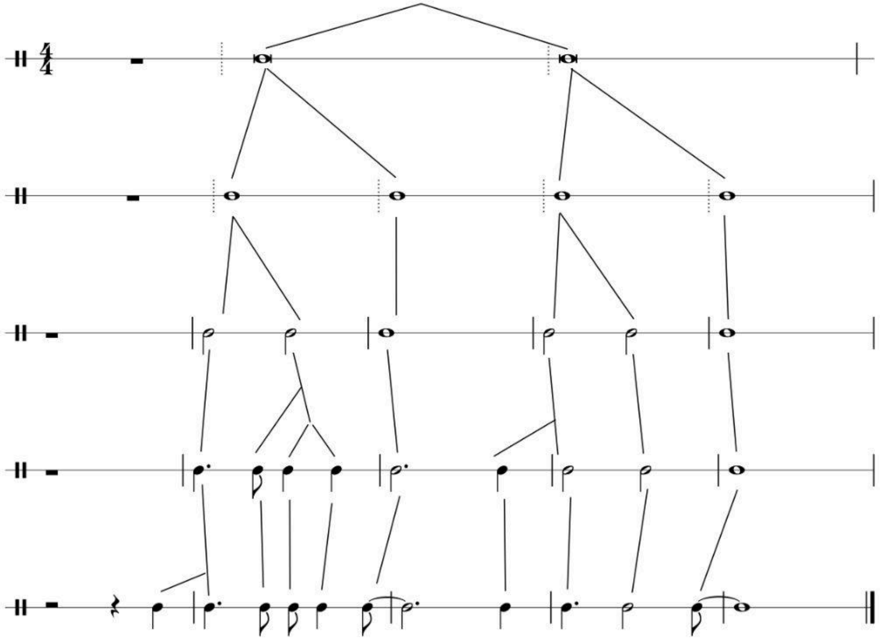
Possible Solutions

Generate corpus:

- for each group g:
  - choose regularity r_g:
  - for each pattern i:
    - sample pattern p_{gi} using r_g



A more detailed model:

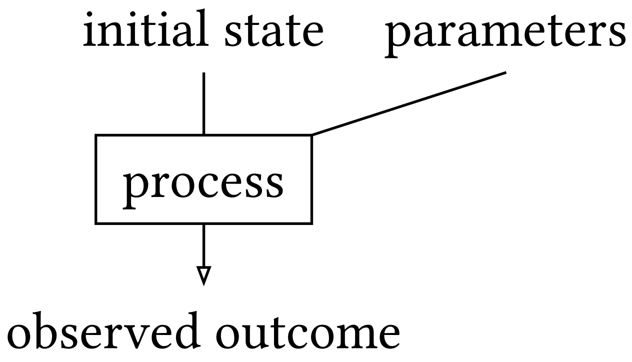


generate full rhythm, flip coins to repeat

Advantages of Generative Models

Very Explicit:

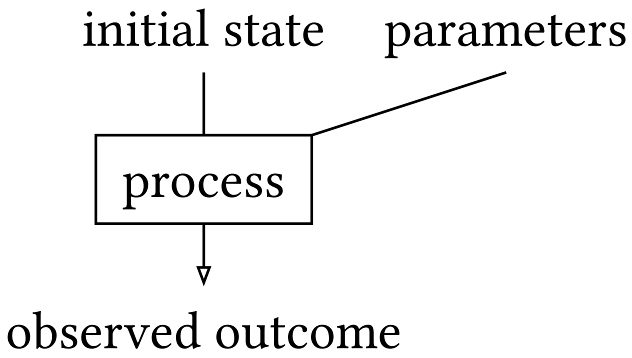
- write down the entire generative process
  - this is how you think/pretend it works
  - links entities (observations, parameters)



Advantages of Generative Models

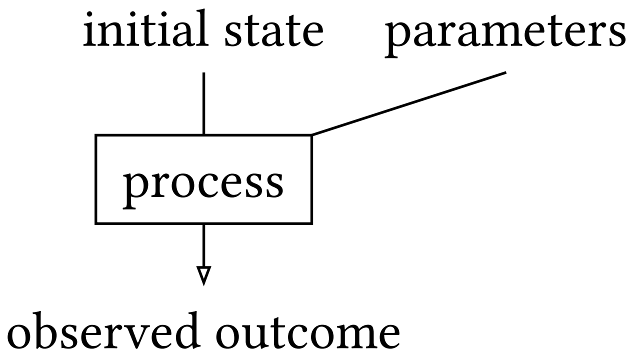
Very Explicit:

- write down the entire generative process
  - this is how you think/pretend it works
  - links entities (observations, parameters)
- can be discussed



Advantages of Generative Models

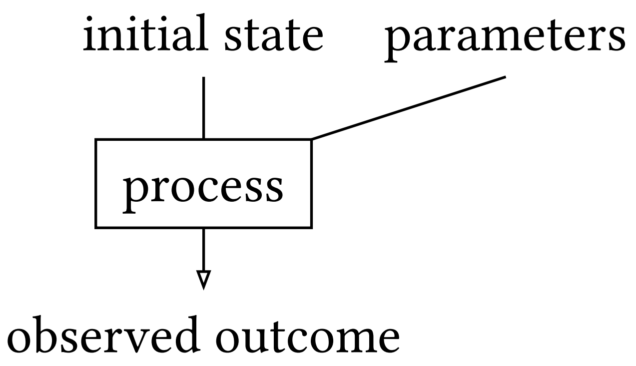
Very Explicit:
- write down the entire generative process
  - this is how you think/pretend it works
  - links entities (observations, parameters)
- can be discussed
- can be used to understand a phenomenon better
  - What is regularity? Why is music regular?



Advantages of Generative Models

Very Explicit:

- write down the entire generative process
  - this is how you think/pretend it works
  - links entities (observations, parameters)
- can be discussed
- can be used to understand a phenomenon better
  - What is regularity? Why is music regular?



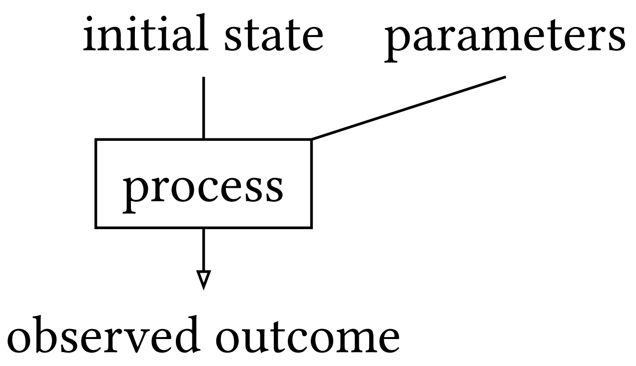
Simulation:

- run the generative process
- manipulate process / parameters

Advantages of Generative Models

Very Explicit:

- write down the entire generative process
  - this is how you think/pretend it works
  - links entities (observations, parameters)
- can be discussed
- can be used to understand a phenomenon better
  - What is regularity? Why is music regular?



Simulation:

- run the generative process
- manipulate process / parameters

Inference: ???

Inference

Quantifying Uncertainty

"Random variable" X: uncertain quantity or property

Quantifying Uncertainty

“Random variable” X: uncertain quantity or property

- future event (coin flip)

Quantifying Uncertainty

“Random variable” X: uncertain quantity or property

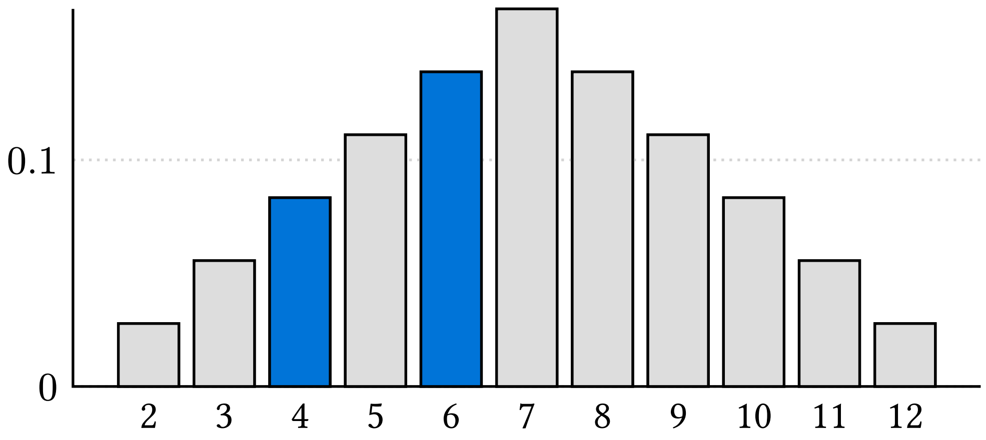
- future event (coin flip)
- unobservable (regularity of a corpus)

Quantifying Uncertainty

"Random variable" X: uncertain quantity or property

- future event (coin flip)
- unobservable (regularity of a corpus)

discrete: mass function $p(x)$

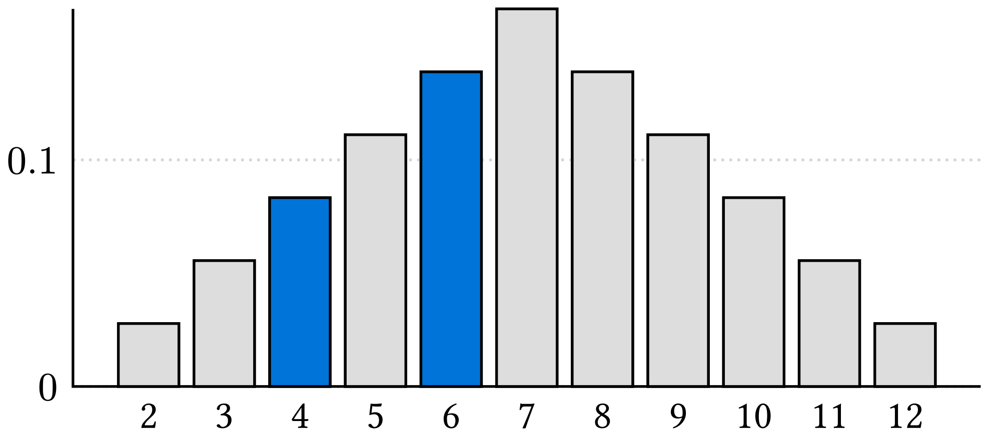


Quantifying Uncertainty

"Random variable" X: uncertain quantity or property

- future event (coin flip)
- unobservable (regularity of a corpus)

discrete: mass function $p(x)$



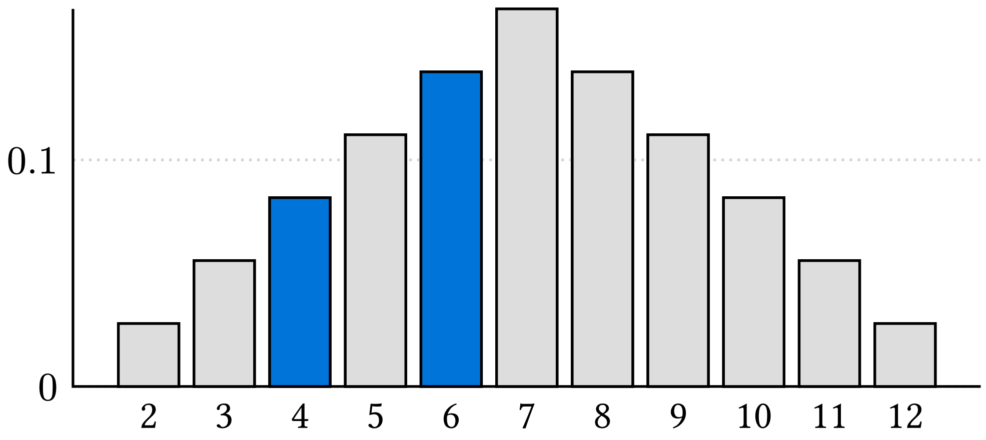
$$P(X \in \{4, 6\}) = p(4) + p(6)$$

Quantifying Uncertainty

"Random variable" X: uncertain quantity or property

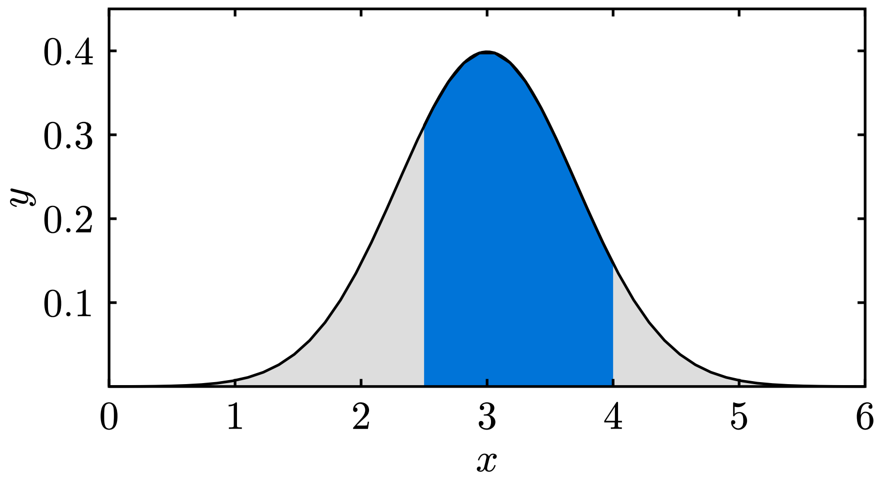
- future event (coin flip)
- unobservable (regularity of a corpus)

discrete: mass function $p(x)$



$$P(X \in \{4, 6\}) = p(4) + p(6)$$

continuous: density function $p(x)$

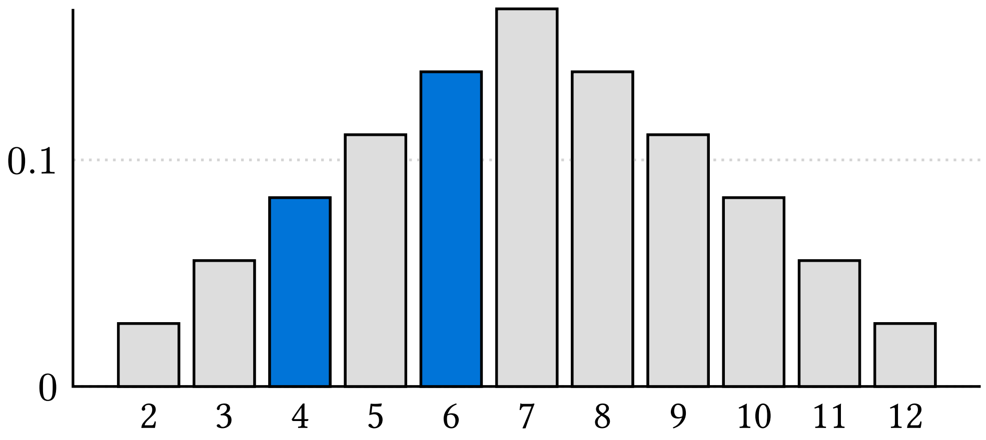


Quantifying Uncertainty

"Random variable" X: uncertain quantity or property

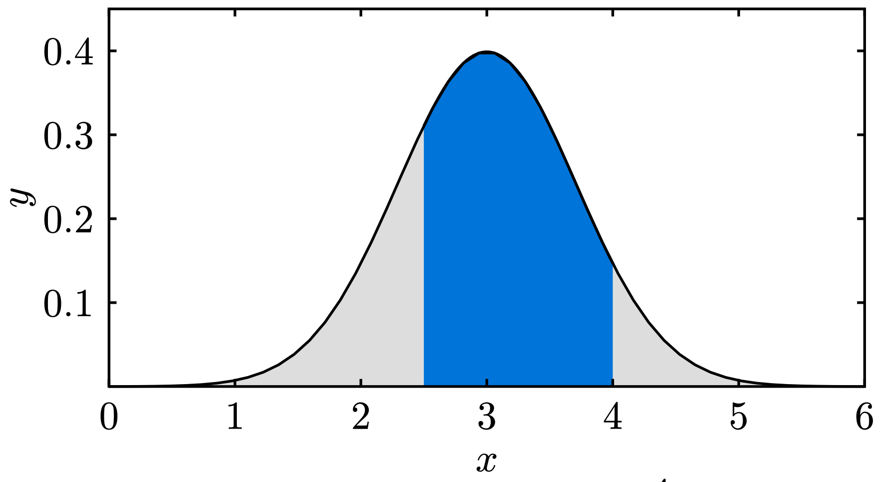
- future event (coin flip)
- unobservable (regularity of a corpus)

discrete: mass function $p(x)$



$$P(X \in \{4, 6\}) = p(4) + p(6)$$

continuous: density function $p(x)$



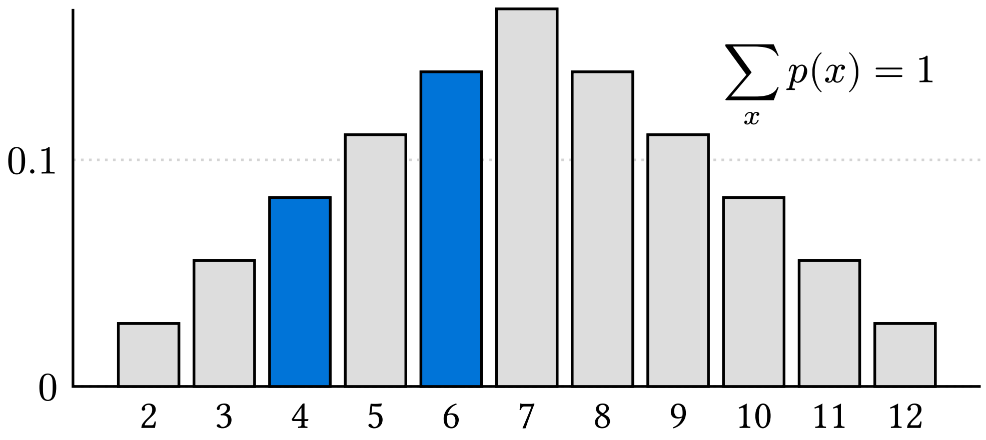
$$P(2.5 \leq X \leq 4) = \int_{2.5}^{4} p(x)\,\mathrm{d}x$$

Quantifying Uncertainty

"Random variable" X: uncertain quantity or property

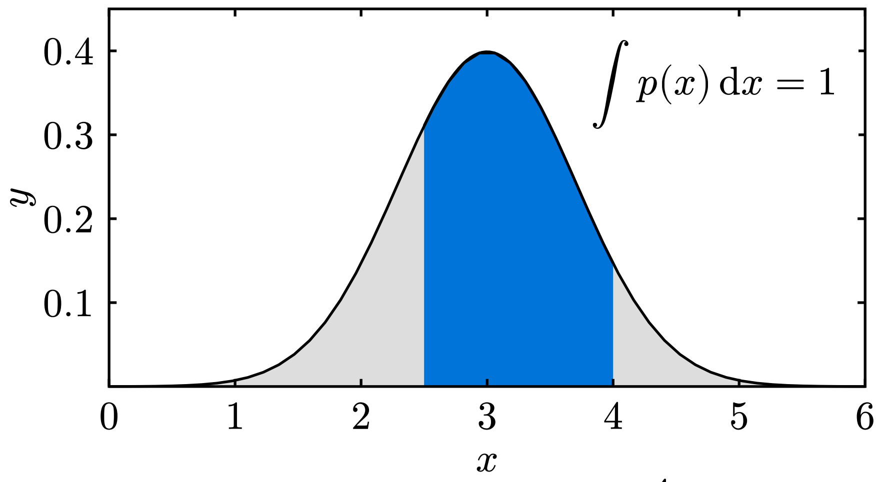
- future event (coin flip)
- unobservable (regularity of a corpus)

discrete: mass function $p(x)$



$$P(X \in \{4, 6\}) = p(4) + p(6)$$

continuous: density function $p(x)$



$$P(2.5 \leq X \leq 4) = \int_{2.5}^{4} p(x)\,\mathrm{d}x$$

Distributions over Several Variables

Several variables: **joint** distribution

$$p(x, y, z)$$

Distributions over Several Variables

Several variables: **joint** distribution

$$p(x, y, z)$$

"don't care": **marginal** distribution

$$\begin{aligned} p(x, y) &\quad \text{ignore } z \\ p(x) &\quad \text{ignore } y \text{ and } z \end{aligned}$$

Distributions over Several Variables

Several variables: **joint** distribution

$$p(x, y, z)$$

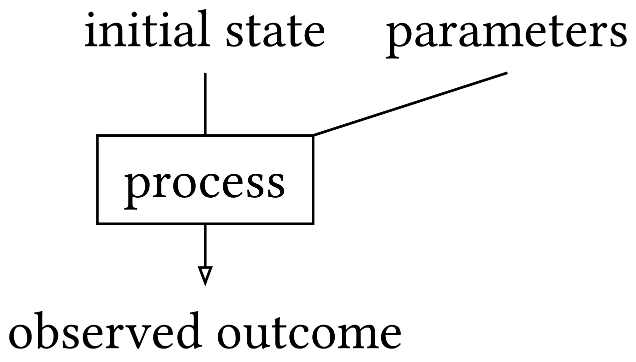
"don't care": **marginal** distribution

$$\begin{aligned} p(x, y) &\quad \text{ignore } z \\ p(x) &\quad \text{ignore } y \text{ and } z \end{aligned}$$

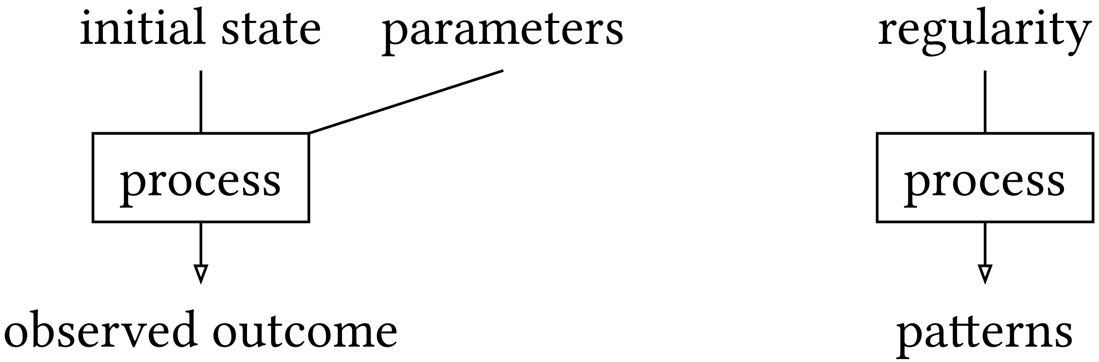
"observation": **conditional** distribution

$$\begin{aligned} p(x \mid y, z) &\quad x \text{ given } y \text{ and } z \\ p(x, z \mid y) &\quad x \text{ and } z \text{ given } y \\ p(y \mid x) &\quad y \text{ given } x \text{ (ignoring } z) \end{aligned}$$

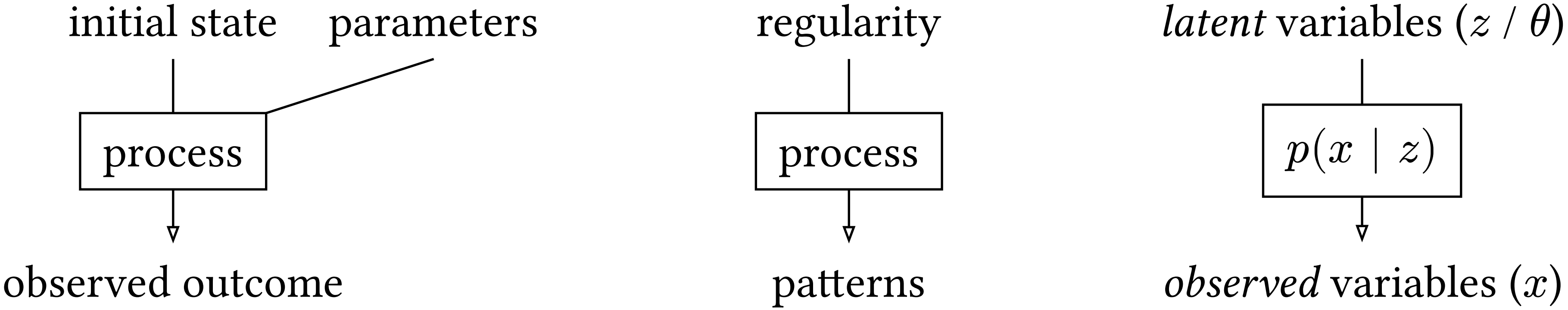
Inference



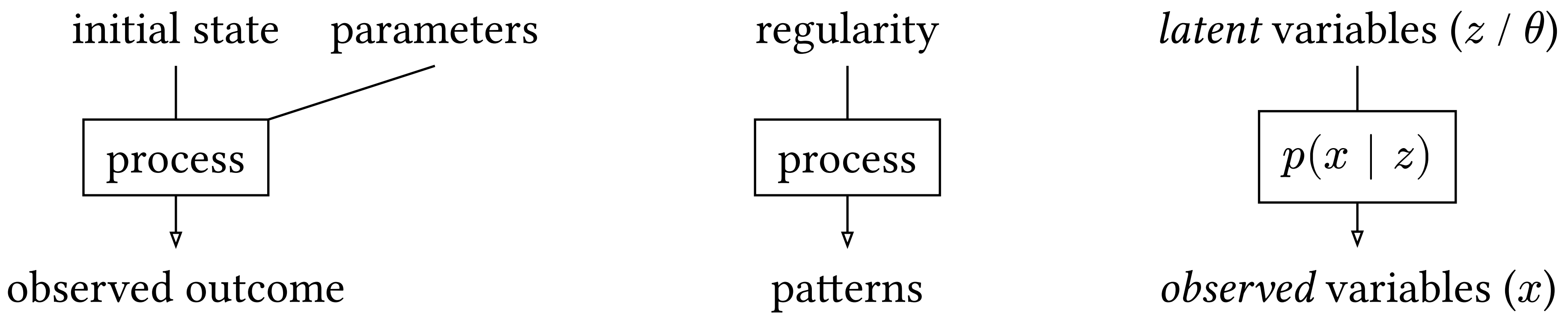
Inference


initial state
parameters
process
observed outcome
regularity
process
patterns

Inference


initial state
parameters
process
observed outcome
regularity
process
patterns
latent variables (z / θ)
$p(x \mid z)$
observed variables (x)

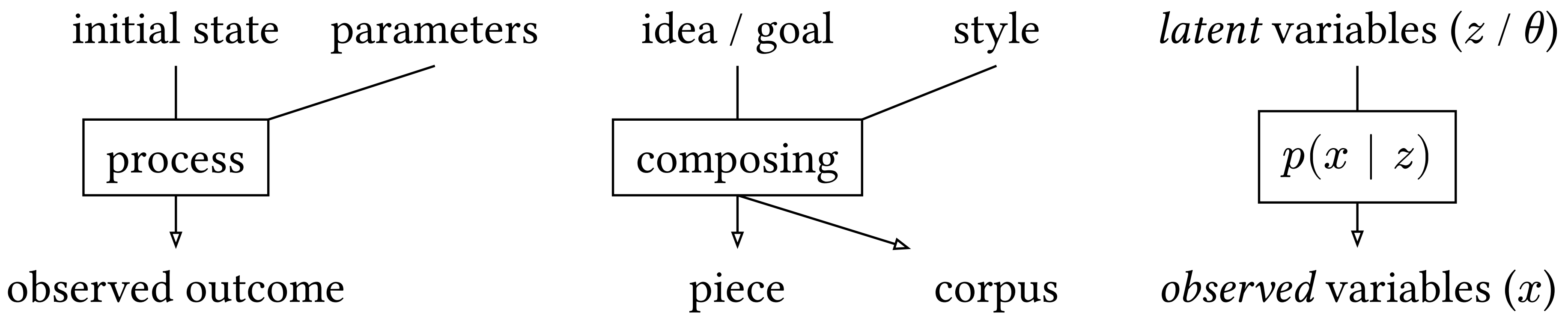
Inference



The Inference Button™:

$$\underbrace{p(x, z)}_{\substack{\text{model} \\ p(z) \cdot p(x|z)}} \longrightarrow \underbrace{p(z \mid x)}_{\substack{\text{posterior} \\ \text{distribution}}}$$

Inference



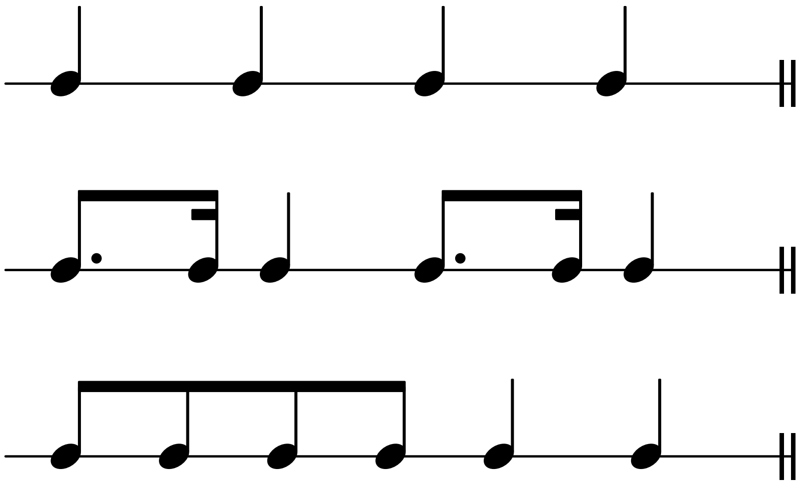
The Inference Button™:

$$\underbrace{p(x, z)}_{\substack{\text{model} \\ p(z) \cdot p(x|z)}} \longrightarrow \underbrace{p(z \mid x)}_{\substack{\text{posterior} \\ \text{distribution}}}$$

Back to our Corpus Study

Variables:

- regularity $r = ?$
- corpus C



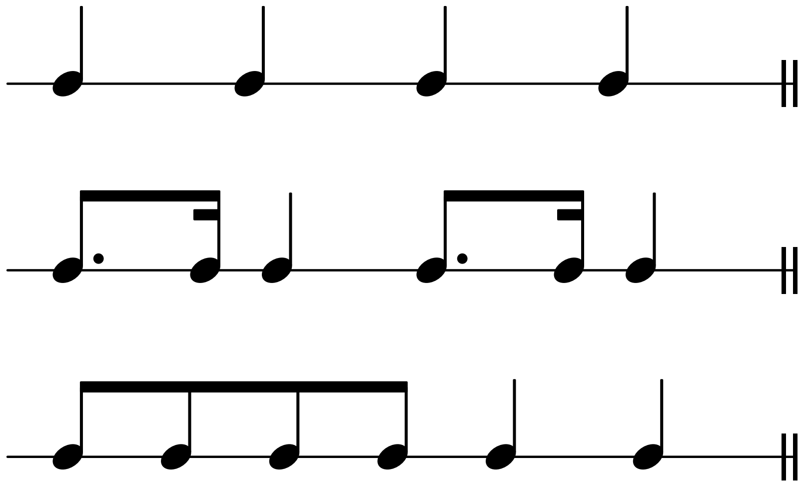
Back to our Corpus Study

Variables:

- regularity $r = ?$
- corpus C

Model:

- choose r
- for each pattern:
  - beat 1 random
  - for beat 2-4:
    - flip coin (r):
      - heads: repeat
      - tails: don't repeat



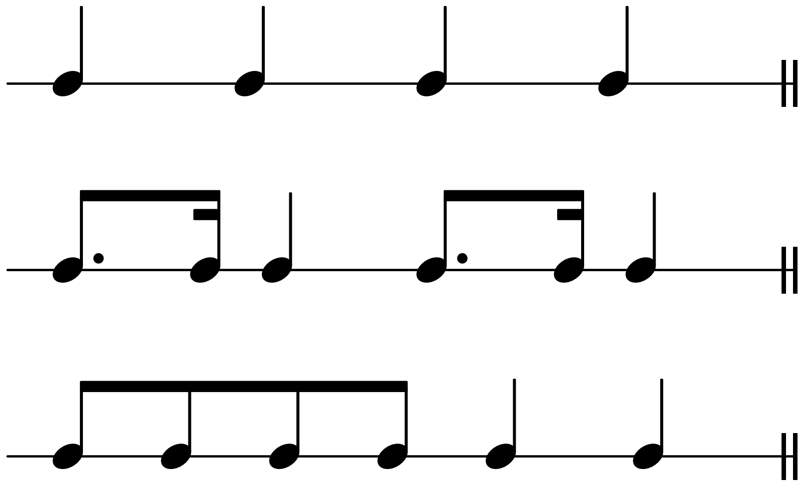
Back to our Corpus Study

Variables:

- regularity $r = ?$
- corpus $C = [$`*rrr`, `*nnn`, `*rnr`$]$

Model:

- choose r
- for each pattern:
  - beat 1 random
  - for beat 2-4:
    - flip coin (r):
      - heads: repeat
      - tails: don't repeat



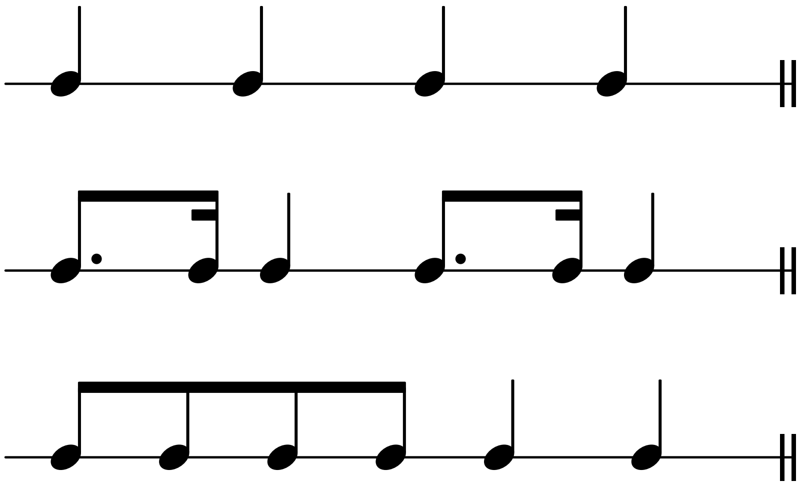
Back to our Corpus Study

Variables:

- regularity $r = ?$
- corpus $C = [\texttt{*rrr}, \texttt{*nnn}, \texttt{*rnr}]$

Model:

- choose $r \sim$ unknown distribution
- for each pattern:
  - beat 1 random
  - for beat 2-4:
    - flip coin (r):
      - heads: repeat
      - tails: don't repeat



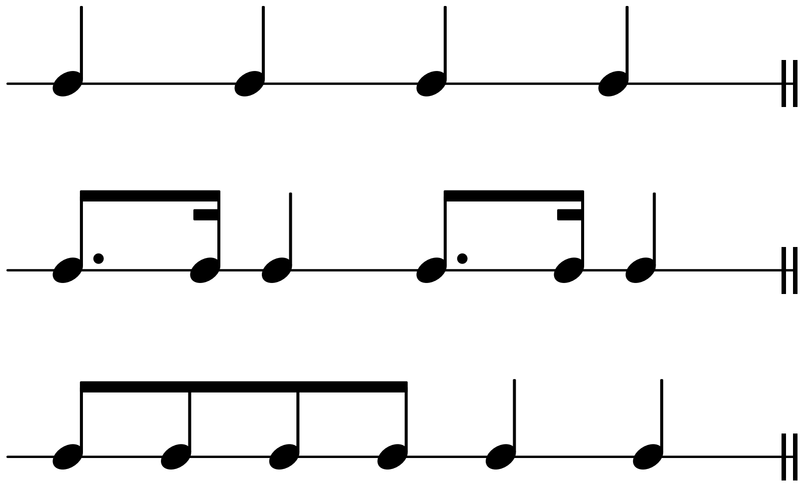
Back to our Corpus Study

Variables:

- regularity $r = ?$
- corpus $C = [$`*rrr`, `*nnn`, `*rnr`$]$

Model:

- choose $r \sim$ unknown distribution
- for each pattern:
  - beat 1 random
  - for beat 2-4:
    - flip coin (r): $\sim \text{Bernoulli}(r)$
      - heads: repeat
      - tails: don't repeat



Looking at the Likelihood

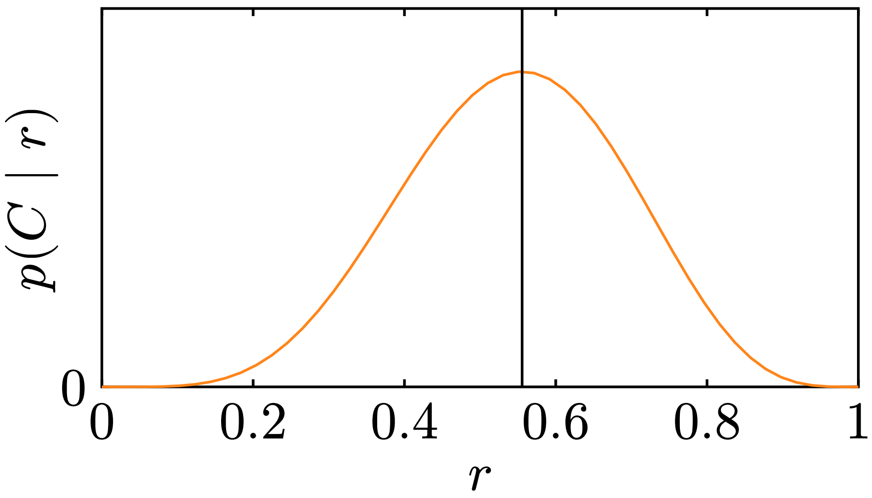
The probability of the data depends on r:

$$p(C \mid r) = p([*\mathtt{rrr}, *\mathtt{nnn}, *\mathtt{rnr}] \mid r) = r^5 \cdot (1 - r)^4$$

Looking at the Likelihood

The probability of the data depends on r:

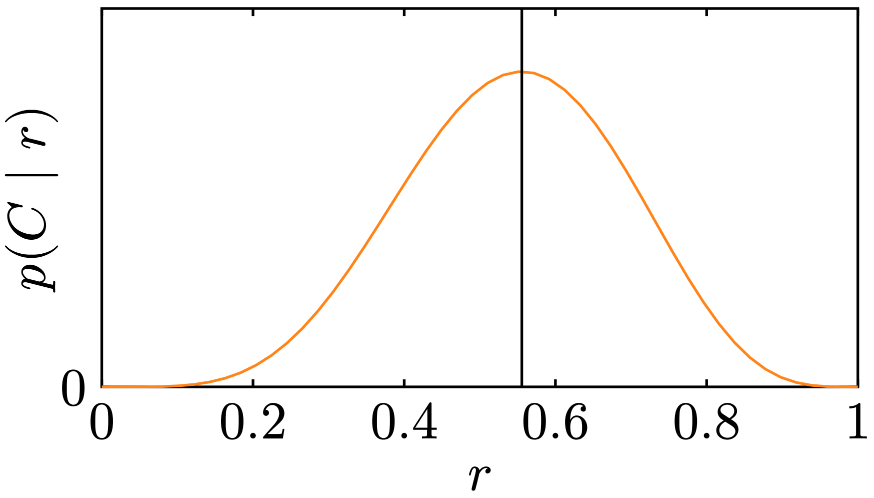
$$p(C \mid r) = p([\texttt{*rrr}, \texttt{*nnn}, \texttt{*rnr}] \mid r) = r^5 \cdot (1-r)^4$$



Looking at the Likelihood

The probability of the data depends on r:

$$p(C \mid r) = p([\texttt{*rrr}, \texttt{*nnn}, \texttt{*rnr}] \mid r) = r^5 \cdot (1-r)^4$$



$$\arg\max_r p(C \mid r) = \frac{5}{9}$$

Make it Bayesian

$$\underbrace{p(r, C)}_{p(r)\cdot p(C|r)} \longrightarrow p(r \mid C)$$

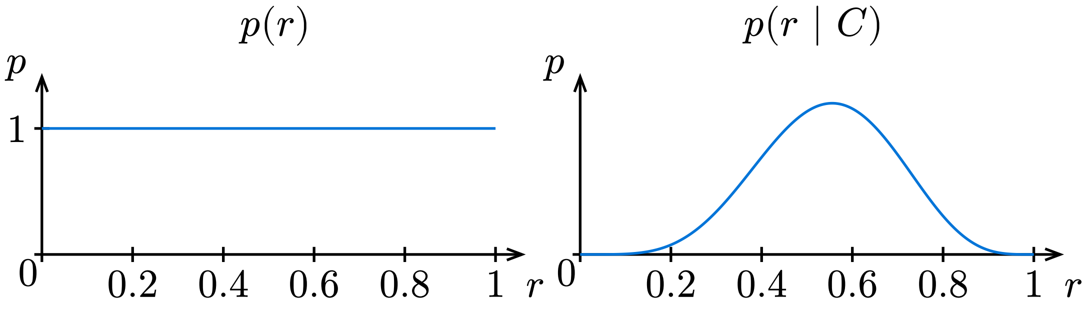
Make it Bayesian

$$\underbrace{p(r, C)}_{p(r) \cdot p(C | r)} \longrightarrow p(r \mid C)$$

Model:

- choose $r \sim \text{Uniform}(0, 1)$
- for each pattern i:
  - choose $\vec{c}_i \sim 3 \times \text{Bernoulli}(r)$

Make it Bayesian



$$\underbrace{p(r, C)}_{p(r) \cdot p(C|r)} \longrightarrow p(r \mid C)$$

Model:

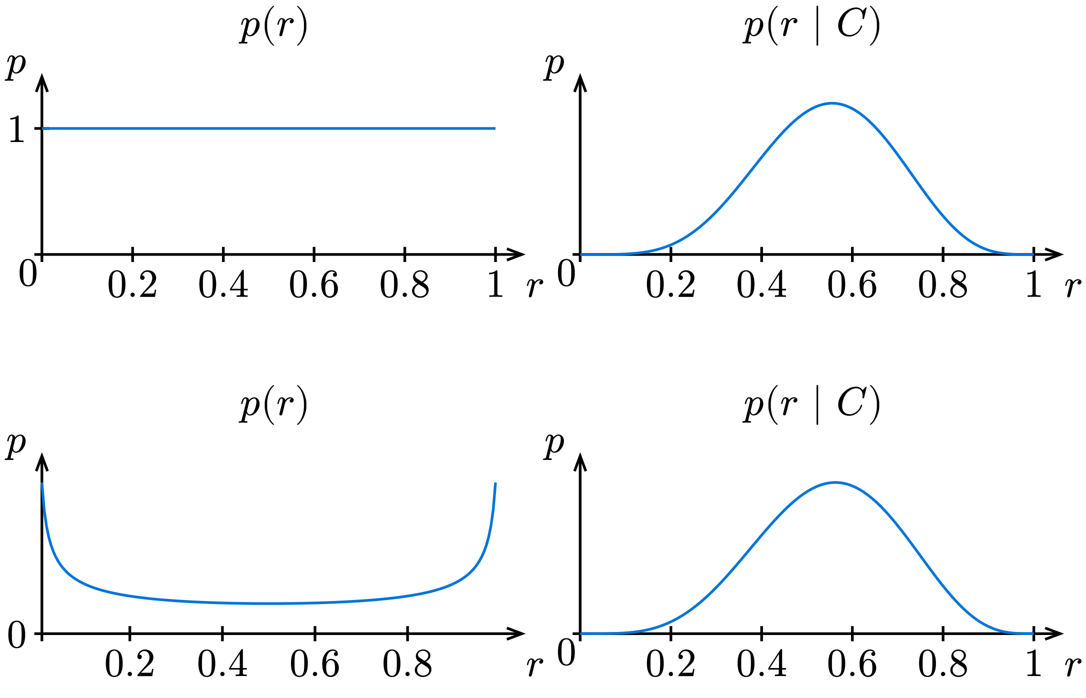
- choose $r \sim \text{Uniform}(0, 1)$
- for each pattern i:
  - choose $\vec{c}_i \sim 3 \times \text{Bernoulli}(r)$

Make it Bayesian

$$\underbrace{p(r, C)}_{p(r) \cdot p(C|r)} \longrightarrow p(r \mid C)$$

Model:

- choose $r \sim \text{Beta}(0.5, 0.5)$
- for each pattern i:
  - choose $\vec{c}_i \sim 3 \times \text{Bernoulli}(r)$

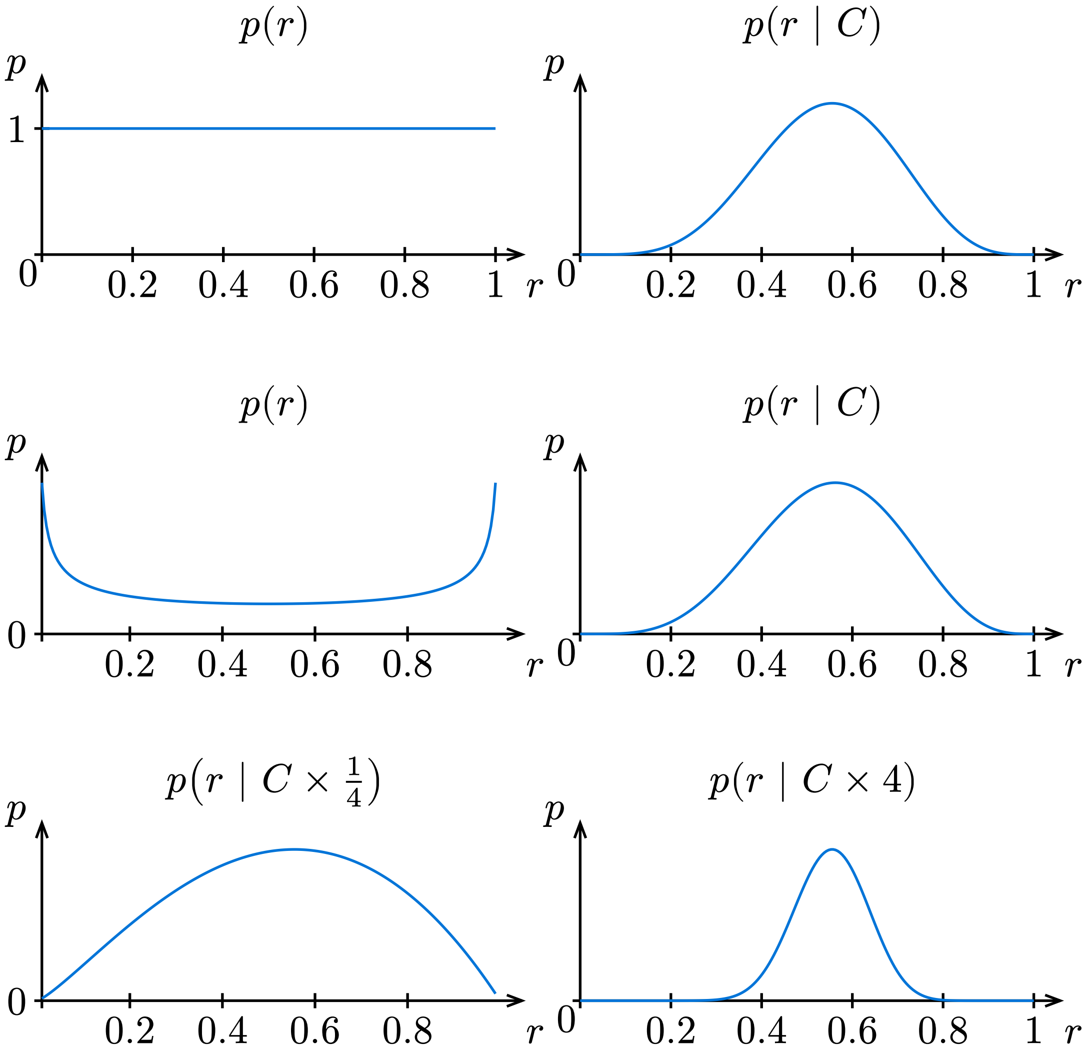


Make it Bayesian

$$\underbrace{p(r, C)}_{p(r) \cdot p(C|r)} \longrightarrow p(r \mid C)$$

Model:

- choose $r \sim \text{Beta}(0.5, 0.5)$
- for each pattern i:
  - choose $\vec{c}_i \sim 3 \times \text{Bernoulli}(r)$



Make it Bayesian

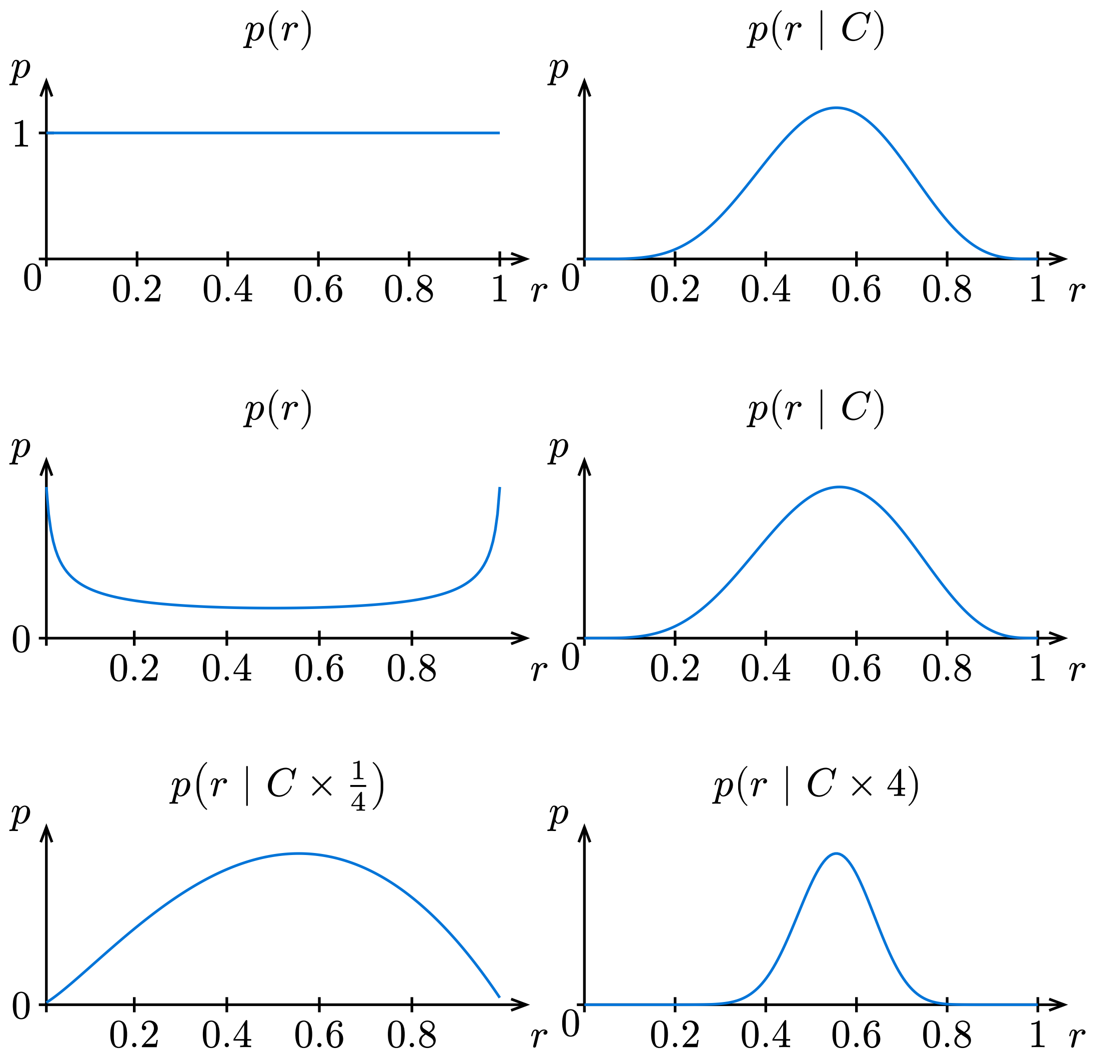
$$\underbrace{p(r, C)}_{p(r) \cdot p(C|r)} \longrightarrow p(r \mid C)$$

Model:

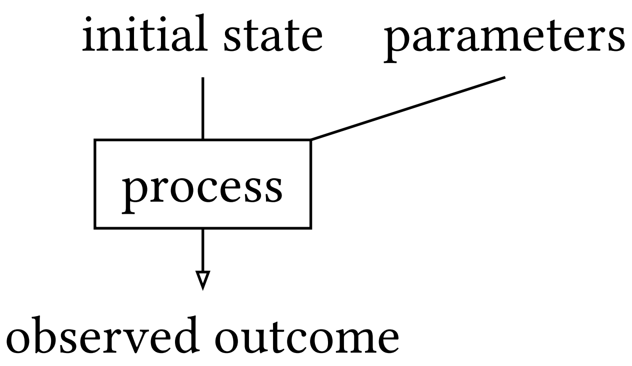
- choose $r \sim \text{Beta}(0.5, 0.5)$
- for each pattern i:
  - choose $\vec{c}_i \sim 3 \times \text{Bernoulli}(r)$

Problem:

How do we compute this?



Summary: The Three Ingredients



$$+ \qquad p(x, z) \qquad \longrightarrow \qquad p(z \mid x)$$

generative process (series of decisions)

decision probabilities

inference: conditioning

Practical Exercises

https://github.com/Amsterdam-Music-Lab/gmth23-bayes-workshop