The new riddle of induction
the “old” riddle
DEDUCTION

general

specific
DEDUCTION

all swan are white

A is a swan
DEDUCTION

general

specific

all swan are white

A is a swan

A is white
INDUCTION

specific

A is white
A is a swan
all swan are white

general
all swan are white
A is a swan
B is a swan
____________________
all swan are white
the sun came up today
the sun came up yesterday
the sun came up the day before yesterday
the sun will come up tomorrow

the sun came up today

the sun came up yesterday

the sun came up the day before yesterday

the sun will come up tomorrow
is a tufa
is a tufa

is a tufa
is a tufa

is a tufa

is a tufa

is a tufa
is a tufa

is a tufa

is a tufa

is a tufa

is a tufa
the “old” riddle:

how are these kinds of inferences justified?
the “old” riddle:

how are these kinds of inferences justified?

uniformity of nature
the “old” riddle:

how are these kinds of inferences justified?

uniformity of nature

how is uniformity of nature justified?
the new riddle
that is, are emeralds green or grue?
the new riddle
that is, are emeralds green or grue?

Nelson Goodman 1983
Fact, Fiction and Forecast
a green emerald
a green emerald
a green emerald
a green emerald
a green emerald
a green emerald
a green emerald

a green emerald

a green emerald

emeralds are green  (evidence supports theory)
grue
if observed before t, green; else, blue

bleen
if observed before t, blue; else, green
a grue emerald
a grue emerald

a grue emerald
a grue emerald
a grue emerald
a grue emerald
a grue emerald
a grue emerald
a grue emerald

emeralds are grue  (evidence supports theory)
green

grue

if observed before t, green; else, blue
DICTIONARY

**green**
- green

**grue**
- if observed before $t$, green; else, blue

elaborate definition, use Occam’s razor
green

- if observed before t, green; else, blue

grue

- if observed before t, grue; else, bleen

MARTIAN’S DICTIONARY

DICTIONARY
**DICTIONARY**

- **green**
  - green

- **grue**
  - if observed before t, green; else, blue

**MARTIAN’S DICTIONARY**

- **green**
  - if observed before t, grue; else, bleen

- **grue**
  - grue

  [elaborate definition]
• “The US forces were always commanded by George Washington, hence they will be commanded by him in the future”
• “The US forces were always commanded by George Washington, hence they will be commanded by him in the future”

• “The US forces were always commanded by the US president, hence they will be commanded by him in the future”
• “The US forces were always commanded by George Washington, hence they will be commanded by him in the future”

• “The US forces were always commanded by the US president, hence they will be commanded by him in the future”

• “Mary Ball Washington was always the mother of George Washington, hence she will be his mother in the future”
• “The US forces were always commanded by George Washington, hence they will be commanded by him in the future”

• “The US forces were always commanded by the US president, hence they will be commanded by him in the future”

• “Mary Ball Washington was always the mother of George Washington, hence she will be his mother in the future”

• “Mary Ball Washington was always the mother of the US president, hence she will be his mother in the future”
• inductive inference is relative to the language it is formulated in
BAYESIAN MODEL SELECTION
BAYESIAN MODEL SELECTION

H1: emeralds are green

H2: emeralds are grue
BAYESIAN MODEL SELECTION

posterior over models

\[ P(\mathcal{H}_i | D) \propto P(D | \mathcal{H}_i)P(\mathcal{H}_i) \]
Bayesian Model Selection

Evidence compatible with both grue and green

\[ P(\mathcal{H}_i \mid D) \propto P(D \mid \mathcal{H}_i)P(\mathcal{H}_i) \]

Posterior over models
BAYESIAN MODEL SELECTION

\[ P(\mathcal{H}_i \mid D) \propto P(D \mid \mathcal{H}_i) P(\mathcal{H}_i) \]

- posterior over models
- evidence compatible with both grue and green
- model prior decides
MINIMUM DESCRIPTION LENGTH

the best model is the one that leads to the best compression of the observed data
MINIMUM DESCRIPTION LENGTH

\[ L(D, \mathcal{H}) = L(\mathcal{H}) + L(D \mid \mathcal{H}) \]
MINIMUM DESCRIPTION LENGTH

\[ L(D, \mathcal{H}) = L(\mathcal{H}) + L(D | \mathcal{H}) \]
MINIMUM DESCRIPTION LENGTH

\[ L(D, \mathcal{H}) = L(\mathcal{H}) + L(D | \mathcal{H}) \]
MINIMUM DESCRIPTION LENGTH

\[ L(D, \mathcal{H}) = L(\mathcal{H}) + L(D \mid \mathcal{H}) \]
KOLMOGOROV COMPLEXITY

10101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101010
KOLMOGOROV COMPLEXITY

\[
\text{description}_{\text{PYTHON}} = (\text{program}_{\text{PYTHON}}, \text{input})
\]

\[
K_{\text{PYTHON}} = 13 \quad \text{depends on language}
\]

\[
K = L(\text{description}) = L(\text{input}) + \text{const}
\]
\[ L(D, \mathcal{H}) = L(\mathcal{H}) + L(D \mid \mathcal{H}) \]

Minimum Description Length

\[ L(D, \mathcal{H}) = - \log P(\mathcal{H}) - \log (P(D \mid \mathcal{H}) \delta D) \]
\[ = - \log P(\mathcal{H} \mid D) + \text{const}. \]

Bayesian inference
\[ L(D, \mathcal{H}) = L(\mathcal{H}) + L(D | \mathcal{H}) \]

Minimum Description Length

Kraft inequality

\[ P(x) = 2^{-L(x)}, \quad L(x) = - \log_2 P(x) \]

Bayesian inference

\[
L(D, \mathcal{H}) = - \log P(\mathcal{H}) - \log (P(D | \mathcal{H}) \delta D) \\
= - \log P(\mathcal{H} | D) + \text{const.}
\]
\[ L(D, \mathcal{H}) = \boxed{L(\mathcal{H})} + L(D \mid \mathcal{H}) \]

**Minimum Description Length**

**Kraft inequality**

\[ P(x) = 2^{-L(x)}, \quad L(x) = -\log_2 P(x) \]

**Bayesian inference**

\[
\begin{align*}
L(D, \mathcal{H}) &= -\log P(\mathcal{H}) - \log (P(D \mid \mathcal{H})\delta D) \\
&= -\log P(\mathcal{H} \mid D) + \text{const.}
\end{align*}
\]
“the subject must pick a \((universal)\) Turing machine whose operations describe the basic operations believed to represent "simplicity" by the subject.

However, one could always choose a Turing machine with a simple operation that happened to construct one’s entire theory and would hence score highly under the razor.”
• Goodman’s problem - inductive inference is relative to the language it is formulated in
• Goodman’s problem - inductive inference is relative to the language it is formulated in

• this problem appears in the formal approaches as well
• Goodman’s problem - inductive inference is relative to the language it is formulated in

• this problem appears in the formal approaches as well

• how should we choose the language?
SOURCES

http://plato.stanford.edu/entries/induction-problem/#GruParNewRidInd

David McKay - Information Theory, Inference, and Learning Algorithms

http://jeremykun.com/2012/04/21/kolmogorov-complexity-a-primer/

http://en.wikipedia.org/wiki/Occam's_razor
• What is the difference between those generalizations that are supported by their instances and those that are not?

• Which generalizations support counterfactual conditionals?

• How are lawlike generalizations to be distinguished from accidental generalizations?