Teaching Bayesian Model Comparison with the Three-sided Coin

B. Blais[1,2] S. Kuindersma[3]

[1]Department of Science and Technology
Bryant University

[2]Institute for Brain and Neural Systems
Brown University

[3]Department of Computer Science
University of Massachusetts, Amherst

Bryant Faculty Research Day, January 2007

Abstract

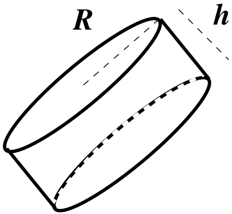
In the present work we introduce the problem of determining the probability that a rotating and bouncing cylinder (i.e. flipped coin) will land and come to rest on its edge. We present this problem and analysis as a practical, nontrivial example to introduce the reader to Bayesian model comparison. Several models are presented, each of which take into consideration different physical aspects of the problem and the relative effects on the edge landing probability. The Bayesian formulation of model comparison is then used to compare the models and their predictive agreement with data from hand-flipped cylinders of several sizes. Keywords: probability theory, log-likelihood, parameter estimation, marginalization, Occam's razor.

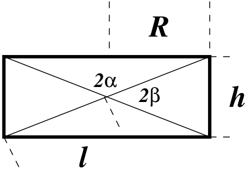
Outline

1. What is a Three-Sided Coin?
   - Example Models
2. What Does Bayesian Mean?
   - Bayesian Model Comparison
   - Occam Factor
3. Model Comparison with the Three-sided Coin

Outline

Coin





Definitions

- $\eta \equiv h/R$
- $l = \sqrt{R^2 + (h/2)^2}$
- $\beta \equiv \text{atan}\left(\frac{h}{2R}\right) = \text{atan}(\eta/2)$

The Question

What is the probability, p_{edge}, for the coin to land on the edge, as a function of the radius, R, and the height, h?

Restricted version of the question for a "fair" coin:

What values of h and R yield $p_{\text{edge}}(h, R) = 1/3$?

The Question

What is the probability, p_{edge}, for the coin to land on the edge, as a function of the radius, R, and the height, h?

Restricted version of the question for a “fair” coin:

What values of h and R yield $p_{\text{edge}}(h, R) = 1/3$?

Example: Surface Area

- Probability proportional to the surface area

$$p_{\text{edge}}(h,R) = \frac{2\pi Rh}{2\pi R^2 + 2\pi Rh}$$

$$p_{\text{edge}}(\eta) = \frac{\eta}{1+\eta} \quad (\eta \equiv h/R)$$

For the fair coin we obtain

$$\frac{\eta}{1+\eta} = \frac{1}{3}$$

$$\eta = \frac{1}{2}$$

Example: Surface Area

- Probability proportional to the surface area

$$p_{\text{edge}}(h,R) = \frac{2\pi Rh}{2\pi R^2 + 2\pi Rh}$$

$$p_{\text{edge}}(\eta) = \frac{\eta}{1+\eta} \quad (\eta \equiv h/R)$$

For the fair coin we obtain

$$\frac{\eta}{1+\eta} = \frac{1}{3}$$

$$\eta = \frac{1}{2}$$

Example: Surface Area

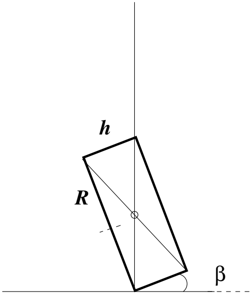
- Probability proportional to the surface area

$$
\begin{aligned}
p_{\mathrm{edge}}(h,R) &= \frac{2\pi Rh}{2\pi R^2 + 2\pi Rh} \\
p_{\mathrm{edge}}(\eta) &= \frac{\eta}{1+\eta} \quad (\eta \equiv h/R)
\end{aligned}
$$

For the fair coin we obtain

$$
\begin{aligned}
\frac{\eta}{1+\eta} &= \frac{1}{3} \\
\eta &= \frac{1}{2}
\end{aligned}
$$

Example: Center of Mass



β determines tip-over direction

$$\beta \equiv \mathrm{atan}\left(\eta/2\right)$$

- if $0 < \theta < \beta$ then the coin will land on edge
- if $\beta < \theta < \pi/2$ then the coin will land on the heads

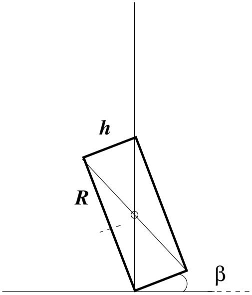
$$p_{\mathrm{edge}}(h,R) = \frac{\beta}{\pi/2} \equiv p_e$$

For a fair coin we obtain

$$p_{\mathrm{edge}}(h,R) = \frac{1}{3} = \frac{\beta}{\pi/2} \Rightarrow \alpha = 60^\circ$$

$$h/R = \frac{2}{\sqrt{3}} \approx 1.155$$

Example: Center of Mass



β determines tip-over direction

$$\beta \equiv \text{atan}\left(\eta/2\right)$$

- if $0 < \theta < \beta$ then the coin will land on edge
- if $\beta < \theta < \pi/2$ then the coin will land on the heads

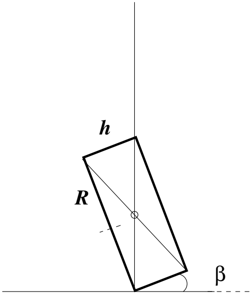
$$p_{\text{edge}}(h, R) = \frac{\beta}{\pi/2} \equiv p_e$$

For a fair coin we obtain

$$p_{\text{edge}}(h, R) = \frac{1}{3} = \frac{\beta}{\pi/2} \Rightarrow \alpha = 60^\circ$$

$$h/R = \frac{2}{\sqrt{3}} \approx 1.155$$

Example: Center of Mass



β determines tip-over direction

$$\beta \equiv \text{atan}\left(\eta/2\right)$$

- if $0 < \theta < \beta$ then the coin will land on edge
- if $\beta < \theta < \pi/2$ then the coin will land on the heads

$$p_{\text{edge}}(h,R) = \frac{\beta}{\pi/2} \equiv p_e$$

For a fair coin we obtain

$$p_{\text{edge}}(h,R) = \frac{1}{3} = \frac{\beta}{\pi/2} \Rightarrow \alpha = 60^\circ$$

$$h/R = \frac{2}{\sqrt{3}} \approx 1.155$$

Models

- Surface Area
- Cross-Sectional Length
- Center of Mass (no-bounce)
- Solid Angle(Mosteller, 1987; Pegg, 1997)
- Dynamic Model(Murray and Teare, 1993)
- Simple Bounce Model (Blais)
- Extended Bounce Model (Kuindersma,Blais)
- Correlated Bounce Model ((Jaynes, 2003), Blais)
- ⋮

Question

Given data, how do we determine which model is correct?

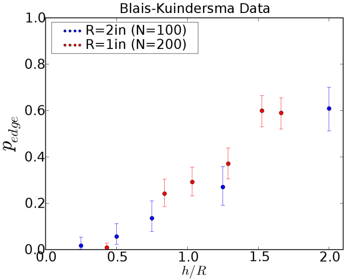
Models

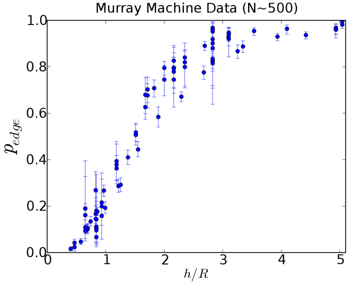
- Surface Area
- Cross-Sectional Length
- Center of Mass (no-bounce)
- Solid Angle(Mosteller, 1987; Pegg, 1997)
- Dynamic Model(Murray and Teare, 1993)
- Simple Bounce Model (Blais)
- Extended Bounce Model (Kuindersma,Blais)
- Correlated Bounce Model ((Jaynes, 2003), Blais)
- ⋮

Question

Given data, how do we determine which model is correct?

Some Data





Question

Given data, how do we determine which model is correct?

Outline

Bayesian Inference

Probability of our Proposition Given the Data

$$p(\text{proposition}|\text{data}, I) \equiv p(M|\text{data}, I)$$

Applying Bayes' Theorem

$$\overbrace{p(\text{M}|\text{data}, I)}^{\text{posterior}} = \frac{\overbrace{p(\text{data}|\text{M}, I)}^{\text{likelihood}}\overbrace{p(\text{M}|I)}^{\text{prior}}}{\underbrace{p(\text{data}|I)}_{\text{normalization}}}$$

$$\propto \overbrace{p(\text{data}|\text{M}, I)}^{\text{likelihood}}\overbrace{p(\text{M}|I)}^{\text{prior}}$$

Bayesian Inference

Probability of our Proposition Given the Data

$$p(\text{proposition}|\text{data}, I) \equiv p(M|\text{data}, I)$$

Applying Bayes' Theorem

$$\overbrace{p(\text{M}|\text{data}, I)}^{\text{posterior}} = \frac{\overbrace{p(\text{data}|\text{M}, I)}^{\text{likelihood}}\overbrace{p(\text{M}|I)}^{\text{prior}}}{\underbrace{p(\text{data}|I)}_{\text{normalization}}}$$

$$\propto \overbrace{p(\text{data}|\text{M}, I)}^{\text{likelihood}}\overbrace{p(\text{M}|I)}^{\text{prior}}$$

Two Schools of Thought on Probability

Frequentist Statistical Inference

$p(A) =$long-run relative frequency with which A occurs in identical repeats of an experiment.
"A" restricted to propositions about random variables.

Bayesian Inference

$p(A|B) =$a real number measure of the plausibility of a proposition/hypothesis A, given (conditional on) the truth of the information represented by proposition B.
"A" can be any logical proposition, *not* restricted to propositions about random variables.

Objective versus Subjective

- Bayesian inference is often labeled as *subjective*, because the probability is a measure of a state of knowledge, and not directly observable like a relative frequency
- Loredo 1990: “In this sense, Bayesian Probability Theory is ‘subjective,’ it describes states of knowledge, not states of nature. But it is ‘objective’ in that we insist that equivalent states of knowledge be represented by equal probabilities, and that problems be well-posed: enough information must be provided to allow unique, unambiguous probability assignments.”
- Jaynes Consistency

Objective versus Subjective

- Bayesian inference is often labeled as *subjective*, because the probability is a measure of a state of knowledge, and not directly observable like a relative frequency
- Loredo 1990: "In this sense, Bayesian Probability Theory is 'subjective,' it describes states of knowledge, not states of nature. But it is 'objective' in that we insist that equivalent states of knowledge be represented by equal probabilities, and that problems be well-posed: enough information must be provided to allow unique, unambiguous probability assignments."
- Jaynes Consistency

Bayesian Model Comparison

Compare Posterior Probabilities

$$\frac{p(M_1|\text{data}, I)}{p(M_2|\text{data}, I)}$$

Data

$$\begin{aligned} h &= 1\text{in} \\ R &= 1\text{in} \\ N_{\text{flips}} &= 20 \\ N_{\text{edge}} &= 7 \end{aligned}$$

Models

- Surface Area:

$$p_{\text{edge}} = 0.5$$

- Center of Mass:

$$p_{\text{edge}} = 0.3$$

Bayesian Model Comparison

Compare Posterior Probabilities

$$\frac{p(M_1|\text{data}, I)}{p(M_2|\text{data}, I)}$$

Data

$$h = 1\text{in}$$
$$R = 1\text{in}$$
$$N_{\text{flips}} = 20$$
$$N_{\text{edge}} = 7$$

Models

- Surface Area:

$$p_{\text{edge}} = 0.5$$

- Center of Mass:

$$p_{\text{edge}} = 0.3$$

Bayesian Model Comparison

Compare Posterior Probabilities

$$\frac{p(M_1|\text{data}, I)}{p(M_2|\text{data}, I)}$$

Data

$$\begin{aligned} h &= 1\text{in} \\ R &= 1\text{in} \\ N_{\text{flips}} &= 20 \\ N_{\text{edge}} &= 7 \end{aligned}$$

Models

- Surface Area:

$$p_{\text{edge}} = 0.5$$

- Center of Mass:

$$p_{\text{edge}} = 0.3$$

Bayesian Model Comparison

Surface Area (SA)

$$
\begin{aligned}
p_{\text{edge}} &= 0.5 \\
p(M_{\text{SA}}|N=20, N_e=7, I) &\propto p(N=20, N_e=7|M_{\text{SA}}, I) \\
&= \frac{20!}{7!13!}(0.5)^7(1-0.5)^{13} \\
&= 0.074
\end{aligned}
$$

Center Of Mass (COM)

$$
\begin{aligned}
p_{\text{edge}} &= 0.3 \\
p(M_{\text{COM}}|N=20, N_e=7, I) &\propto p(N=20, N_e=7|M_{\text{COM}}, I) \\
&= \frac{20!}{7!13!}(0.3)^7(1-0.3)^{13} \\
&= 0.164
\end{aligned}
$$

Bayesian Model Comparison

Surface Area (SA)

$$
\begin{aligned}
p_{\text{edge}} &= 0.5 \\
p(M_{\text{SA}}|N = 20, N_e = 7, I) &\propto p(N = 20, N_e = 7|M_{\text{SA}}, I) \\
&= \frac{20!}{7!13!}(0.5)^7(1 - 0.5)^{13} \\
&= 0.074
\end{aligned}
$$

Center Of Mass (COM)

$$
\begin{aligned}
p_{\text{edge}} &= 0.3 \\
p(M_{\text{COM}}|N = 20, N_e = 7, I) &\propto p(N = 20, N_e = 7|M_{\text{COM}}, I) \\
&= \frac{20!}{7!13!}(0.3)^7(1 - 0.3)^{13} \\
&= 0.164
\end{aligned}
$$

Bayesian Model Comparison

Surface Area (SA) versus Center Of Mass (COM)

$$\frac{p(M_{\text{COM}}|\text{data}, I)}{p(M_{\text{SA}}|\text{data}, I)} = \frac{0.164}{0.074} = 2.2$$

Given the data, Center Of Mass (COM) is more likely than Surface Area (SA).

Bayesian Model Comparison

Surface Area (SA) versus Center Of Mass (COM)

$$\frac{p(M_{\rm COM}|{\rm data},I)}{p(M_{\rm SA}|{\rm data},I)} = \frac{0.164}{0.074} = 2.2$$

Given the data, Center Of Mass (COM) is more likely than Surface Area (SA).

Advantages

- Intuitive: probabilities of hypotheses
- Simple: everything in the posterior

$$p(M|\text{data}, I)$$

- Estimate parameters: maximum posterior, median, etc.
- Estimate confidence (credible) intervals: width of posterior
- Compare models with different complexity: built-in Occam factor

$$p(M|\text{data}, I) = \int_{\Delta\gamma} d\gamma\, p(M, \gamma|\text{data}, I)$$

Advantages

- Intuitive: probabilities of hypotheses
- Simple: everything in the posterior

$$p(M|\text{data}, I)$$

- Estimate parameters: maximum posterior, median, etc.
- Estimate confidence (credible) intervals: width of posterior
- Compare models with different complexity: built-in Occam factor

$$p(M|\text{data}, I) = \int_{\Delta\gamma} d\gamma\, p(M, \gamma|\text{data}, I)$$

Advantages

- Intuitive: probabilities of hypotheses
- Simple: everything in the posterior

$$p(M|\text{data}, I)$$

- Estimate parameters: maximum posterior, median, etc.
- Estimate confidence (credible) intervals: width of posterior
- Compare models with different complexity: built-in Occam factor

$$p(M|\text{data}, I) = \int_{\Delta\gamma} d\gamma\, p(M, \gamma|\text{data}, I)$$

Advantages

- Intuitive: probabilities of hypotheses
- Simple: everything in the posterior

$$p(M|\text{data}, I)$$

- Estimate parameters: maximum posterior, median, etc.
- Estimate confidence (credible) intervals: width of posterior
- Compare models with different complexity: built-in Occam factor

$$p(M|\text{data}, I) = \int_{\Delta\gamma} d\gamma p(M, \gamma|\text{data}, I)$$

Advantages

- Intuitive: probabilities of hypotheses
- Simple: everything in the posterior

$$p(M|\text{data}, I)$$

- Estimate parameters: maximum posterior, median, etc.
- Estimate confidence (credible) intervals: width of posterior
- Compare models with different complexity: built-in Occam factor

$$p(M|\text{data}, I) = \int_{\Delta\gamma} d\gamma p(M, \gamma|\text{data}, I)$$

Example of Occam factor

Data

$$\begin{aligned} h &= 1\text{in} \\ R &= 1\text{in} \\ N_{\text{flips}} &= 20 \\ N_{\text{edge}} &= 7 \end{aligned}$$

Models

- Surface Area: $p_{\text{edge}} = 0.5$
- Center of Mass: $p_{\text{edge}} = 0.3$
- Model 3: $0.1 \leq p_e \leq 0.5$ (one free parameter)

$$\begin{aligned} p(M_{\text{SA}}|\text{data}, I) &\propto p(\text{data}|M_{\text{SA}}, I) \\ &= \frac{20!}{7!13!}(0.5)^7(1-0.5)^{13} = 0.074 \\ p(M_{\text{COM}}|\text{data}, I) &\propto p(\text{data}|M_{\text{COM}}, I) \\ &= \frac{20!}{7!13!}(0.3)^7(1-0.3)^{13} = 0.164 \end{aligned}$$

To compare with Model 3 we need $p(M_3|\text{data}, I)$, but we need to integrate over the free parameter

Example of Occam factor

Data

$$h = 1\text{in}$$
$$R = 1\text{in}$$
$$N_{\text{flips}} = 20$$
$$N_{\text{edge}} = 7$$

Models

- Surface Area: $p_{\text{edge}} = 0.5$
- Center of Mass: $p_{\text{edge}} = 0.3$
- Model 3: $0.1 \leq p_e \leq 0.5$ (one free parameter)

$$\begin{aligned} p(M_{\text{SA}}|\text{data}, I) &\propto p(\text{data}|M_{\text{SA}}, I) \\ &= \frac{20!}{7!13!}(0.5)^7(1-0.5)^{13} = 0.074 \\ p(M_{\text{COM}}|\text{data}, I) &\propto p(\text{data}|M_{\text{COM}}, I) \\ &= \frac{20!}{7!13!}(0.3)^7(1-0.3)^{13} = 0.164 \end{aligned}$$

To compare with Model 3 we need $p(M_3|\text{data}, I)$, but we need to integrate over the free parameter

Example of Occam factor

Data

$$h = 1\text{in}$$
$$R = 1\text{in}$$
$$N_{\text{flips}} = 20$$
$$N_{\text{edge}} = 7$$

Models

- Surface Area: $p_{\text{edge}} = 0.5$
- Center of Mass: $p_{\text{edge}} = 0.3$
- Model 3: $0.1 \leq p_e \leq 0.5$ (one free parameter)

$$\begin{aligned} p(M_{\text{SA}}|\text{data}, I) &\propto p(\text{data}|M_{\text{SA}}, I) \\ &= \frac{20!}{7!13!}(0.5)^7(1-0.5)^{13} = 0.074 \\ p(M_{\text{COM}}|\text{data}, I) &\propto p(\text{data}|M_{\text{COM}}, I) \\ &= \frac{20!}{7!13!}(0.3)^7(1-0.3)^{13} = 0.164 \end{aligned}$$

To compare with Model 3 we need $p(M_3|\text{data}, I)$, but we need to integrate over the free parameter

Example of Occam factor

Data

$$h = 1\text{in}$$
$$R = 1\text{in}$$
$$N_{\text{flips}} = 20$$
$$N_{\text{edge}} = 7$$

Models

- Surface Area: $p_{\text{edge}} = 0.5$
- Center of Mass: $p_{\text{edge}} = 0.3$
- Model 3: $0.1 \leq p_e \leq 0.5$ (one free parameter)

$$p(M_3|\text{data}, I) \propto p(\text{data}|M_3, I)$$
$$= \int_{\text{valid } \theta} d\theta p(\text{data}|M_3, \theta, I) \times p(\theta|M_3, I)$$
$$= \int_{0.1}^{0.5} d\theta \frac{20!}{7!13!} (\theta)^7 (1-\theta)^{13} \times \frac{1}{0.5-0.1}$$
$$= 0.108$$

Example of Occam factor

Data

$$
\begin{aligned}
h &= 1\text{in} \\
R &= 1\text{in} \\
N_{\text{flips}} &= 20 \\
N_{\text{edge}} &= 7
\end{aligned}
$$

Models

- Surface Area: $p_{\text{edge}} = 0.5$
- Center of Mass: $p_{\text{edge}} = 0.3$
- Model 3: $0.1 \leq p_e \leq 0.5$ (one free parameter)

$$
\begin{aligned}
p(M_3|\text{data}, I) &\propto p(\text{data}|M_3, I) \\
&= \int_{\text{valid } \theta} d\theta p(\text{data}|M_3, \theta, I) \times p(\theta|M_3, I) \\
&= \int_{0.1}^{0.5} d\theta \frac{20!}{7!13!} (\theta)^7 (1-\theta)^{13} \times \frac{1}{0.5-0.1} \\
&= 0.108
\end{aligned}
$$

Example of Occam factor

Data

$$h = 1\text{in}$$
$$R = 1\text{in}$$
$$N_{\text{flips}} = 20$$
$$N_{\text{edge}} = 7$$

Models

- Surface Area: $p_{\text{edge}} = 0.5$
- Center of Mass: $p_{\text{edge}} = 0.3$
- Model 3: $0.1 \leq p_e \leq 0.5$ (one free parameter)

$$\begin{aligned} p(M_3|\text{data}, I) &\propto p(\text{data}|M_3, I) \\ &= \int_{\text{valid } \theta} d\theta p(\text{data}|M_3, \theta, I) \times p(\theta|M_3, I) \\ &= \int_{0.1}^{0.5} d\theta \frac{20!}{7!13!} (\theta)^7 (1-\theta)^{13} \times \frac{1}{0.5 - 0.1} \\ &= 0.108 \end{aligned}$$

Example of Occam factor

Data

$$h = 1\text{in}$$
$$R = 1\text{in}$$
$$N_{\text{flips}} = 20$$
$$N_{\text{edge}} = 7$$

Models

- Surface Area: $p_{\text{edge}} = 0.5$
- Center of Mass: $p_{\text{edge}} = 0.3$
- Model 3: $0.1 \leq p_e \leq 0.5$ (one free parameter)

$$\begin{aligned} p(M_3|\text{data}, I) &\propto p(\text{data}|M_3, I) \\ &= \int_{\text{valid } \theta} d\theta p(\text{data}|M_3, \theta, I) \times p(\theta|M_3, I) \\ &= \int_{0.1}^{0.5} d\theta \frac{20!}{7!13!} (\theta)^7 (1-\theta)^{13} \times \frac{1}{0.5 - 0.1} \\ &= 0.108 \end{aligned}$$

Example of Occam factor

Data $h = 1$in, $R = 1$in, $N = 20, N_{\text{edge}} = 7$

$$\frac{p(M_3|\text{data}, I)}{p(M_{\text{SA}}|\text{data}, I)} = \frac{0.108}{0.074} = 1.46$$

Model 3 is more likely than Surface Area.

$$\frac{p(M_3|\text{data}, I)}{p(M_{\text{COM}}|\text{data}, I)} = \frac{0.108}{0.164} = 0.658$$

Model 3 ($0.1 \leq p_e \leq 0.5$) *less* likely than Center of Mass($p_e = 0.3$), even though Model 3 does contain the better performing $p_e = 0.35$ value.

Example of Occam factor

Data $h = 1$in, $R = 1$in, $N = 20, N_{\text{edge}} = 7$

$$\frac{p(M_3|\text{data}, I)}{p(M_{\text{SA}}|\text{data}, I)} = \frac{0.108}{0.074} = 1.46$$

Model 3 is more likely than Surface Area.

$$\frac{p(M_3|\text{data}, I)}{p(M_{\text{COM}}|\text{data}, I)} = \frac{0.108}{0.164} = 0.658$$

Model 3 ($0.1 \le p_e \le 0.5$) *less* likely than Center of Mass($p_e = 0.3$), even though Model 3 does contain the better performing $p_e = 0.35$ value.

Outline

Comparison

- For convenience, use the log of the posterior:
  - less-negative $\Rightarrow$ more probable
  - differences of the logs $\Rightarrow$ ratios of posteriors

Model	**Log-Posterior (unnormalized)**	
	Kuindersma-Blais	Murray's Machine
Surface Area	-229	-4747
Cross-sectional Length	-162	-7370
Solid Angle	-155	-2826
Center of Mass	-91	-3313
Simple Bounce	-78	-1947
Dynamic Model	-49	-764
Extended Bounce	-42	-1018

Comments? Questions?

Bibliography

Jaynes, E. T. (2003).
Probability Theory: The Logic of Science.
Cambridge University Press, Cambridge.
Edited by G. Larry Bretthorst.

Mosteller, F. (1987).
Fifty Challenging Problems in Probability with Solutions.
Dover Publications.

Murray, D. B. and Teare, S. W. (1993).
Probability of a tossed coin landing on edge.
Physical Review E, 48(4):2547–2552.

Pegg, E. (1997).
A Complete List of Fair Dice.
Masters Thesis.

Outline

History: R. T. Cox and E. T. Jaynes

Axioms for Probability Theory

1. Degrees of plausibility are represented by real numbers
2. Qualitative correspondence with common sense. Consistent with deductive logic in the limit of true and false propositions.
3. Consistency
   1. If a conclusion can be reasoned out in more than one way, every possible way must lead to the same result
   2. The theory must use all of the information provided
   3. Equivalent states of knowledge must be represented by equivalent plausibility assignments

Bayesian formulation uniquely satisfies these criteria

History: R. T. Cox and E. T. Jaynes

Axioms for Probability Theory

1. Degrees of plausibility are represented by real numbers
2. Qualitative correspondence with common sense. Consistent with deductive logic in the limit of true and false propositions.
3. Consistency
    1. If a conclusion can be reasoned out in more than one way, every possible way must lead to the same result
    2. The theory must use all of the information provided
    3. Equivalent states of knowledge must be represented by equivalent plausibility assignments

Bayesian formulation uniquely satisfies these criteria

History: E. T. Jaynes

Generalization of the Principle of Indifference

- Maximum Entropy
  - Measure of the uncertainty, H, of a distribution, $(p_1, p_2, \ldots, p_n)$, called the entropy
  - Prior probabilities are assigned as those with the maximum entropy, given the initial information of the problem
- Transformation groups
  - Equal states of knowledge yield equal probability assignments

History: E. T. Jaynes

Generalization of the Principle of Indifference

- Maximum Entropy
  - Measure of the uncertainty, H, of a distribution, $(p_1, p_2, \ldots, p_n)$, called the entropy
  - Prior probabilities are assigned as those with the maximum entropy, given the initial information of the problem
- Transformation groups
  - Equal states of knowledge yield equal probability assignments

History: E. T. Jaynes

Generalization of the Principle of Indifference

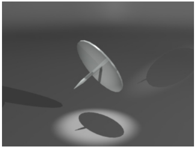
- Maximum Entropy
  - Measure of the uncertainty, H, of a distribution, $(p_1, p_2, \ldots, p_n)$, called the entropy
  - Prior probabilities are assigned as those with the maximum entropy, given the initial information of the problem
- Transformation groups
  - Equal states of knowledge yield equal probability assignments

Outline

4 Jaynes
- Probability
- Priors

5 Flipping a Tack

Lindley (1976): Flipping a Tack



- Flipped thumbtack onto the table
- Data:

  UUUDUDUUUUUD - (9 Ups, and 3 Downs)

- Question:

  Is there good evidence that this tack is (or is not) unbiased (50-50 chance of U or D)?

Flipping a Tack: Frequentist Solution

- Obtain a p-value: “the chance of the observed result or more extreme results given infinite number of identical repetitions”
- For 12 flips, these results are
  - 9 U + 3 D (12 flips)
  - 10 U + 2 D (12 flips)
  - 11 U + 1 D (12 flips)
  - 12 U + 0 D (12 flips)
- Using the standard binomial distribution, with $N = 12$, we get

$$p = \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{2}\left(\frac{1}{2}\right)^{12} + \binom{12}{1}\left(\frac{1}{2}\right)^{12} + \binom{12}{0}\left(\frac{1}{2}\right)^{12} = 7.30\%$$

- Not significant at the 5% level

Flipping a Tack: Frequentist Solution

- Obtain a p-value: “the chance of the observed result or more extreme results given infinite number of identical repetitions”
- For 12 flips, these results are
  - 9 U + 3 D (12 flips)
  - 10 U + 2 D (12 flips)
  - 11 U + 1 D (12 flips)
  - 12 U + 0 D (12 flips)
- Using the standard binomial distribution, with $N = 12$, we get

$$p = \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{2}\left(\frac{1}{2}\right)^{12} + \binom{12}{1}\left(\frac{1}{2}\right)^{12} + \binom{12}{0}\left(\frac{1}{2}\right)^{12} = 7.30\%$$

- Not significant at the 5% level

Flipping a Tack: Frequentist Solution

- Obtain a p-value: “the chance of the observed result or more extreme results given infinite number of identical repetitions”
- For 12 flips, these results are
  - 9 U + 3 D (12 flips)
  - 10 U + 2 D (12 flips)
  - 11 U + 1 D (12 flips)
  - 12 U + 0 D (12 flips)
- Using the standard binomial distribution, with $N = 12$, we get

$$p = \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{2}\left(\frac{1}{2}\right)^{12} + \binom{12}{1}\left(\frac{1}{2}\right)^{12} + \binom{12}{0}\left(\frac{1}{2}\right)^{12} = 7.30\%$$

- Not significant at the 5% level

Flipping a Tack: Frequentist Solution

- Obtain a p-value: "the chance of the observed result or more extreme results given infinite number of identical repetitions"
- For 12 flips, these results are
  - 9 U + 3 D (12 flips)
  - 10 U + 2 D (12 flips)
  - 11 U + 1 D (12 flips)
  - 12 U + 0 D (12 flips)
- Using the standard binomial distribution, with $N = 12$, we get

$$p = \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{2}\left(\frac{1}{2}\right)^{12} + \binom{12}{1}\left(\frac{1}{2}\right)^{12} + \binom{12}{0}\left(\frac{1}{2}\right)^{12} = 7.30\%$$

- Not significant at the 5% level

Flipping a Tack: Frequentist Solution

- Obtain a p-value: “the chance of the observed result or more extreme results given infinite number of identical repetitions”
- For 12 flips, these results are
  - 9 U + 3 D (12 flips)
  - 10 U + 2 D (12 flips)
  - 11 U + 1 D (12 flips)
  - 12 U + 0 D (12 flips)
- Using the standard binomial distribution, with $N = 12$, we get

$$p = \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{2}\left(\frac{1}{2}\right)^{12} + \binom{12}{1}\left(\frac{1}{2}\right)^{12} + \binom{12}{0}\left(\frac{1}{2}\right)^{12} = 7.30\%$$

- Not significant at the 5% level

Flipping a Tack: Frequentist Solution

- Obtain a p-value: "the chance of the observed result or more extreme results given infinite number of identical repetitions"
- For 12 flips, these results are
  - 9 U + 3 D (12 flips)
  - 10 U + 2 D (12 flips)
  - 11 U + 1 D (12 flips)
  - 12 U + 0 D (12 flips)
- Using the standard binomial distribution, with $N = 12$, we get

$$
\begin{aligned}
p &= \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{2}\left(\frac{1}{2}\right)^{12} + \\
&\quad \binom{12}{1}\left(\frac{1}{2}\right)^{12} + \binom{12}{0}\left(\frac{1}{2}\right)^{12} = 7.30\%
\end{aligned}
$$

- Not significant at the 5% level

Frequentist Solution

. . . BUT. . .

Flipping a Tack: Frequentist Solution

- What if the experimenter decided to stop measuring when he reached 3 Down?
- For 3D (3 Down), results at least as extreme are
  - 9 U + 3 D (12 flips)
  - 10 U + 3 D (13 flips)
  - 11 U + 3 D (14 flips)
  - 12 U + 3 D (15 flips)
  - 13 U + 3 D (16 flips)
  - ⋮
- Using the negative binomial distribution, with $D = 3$, we get

$$p = \binom{11}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{13}{1}\left(\frac{1}{2}\right)^{12} + \binom{14}{3}\left(\frac{1}{2}\right)^{12} + \cdots = 3.27\%$$

- Is significant at the 5% level

Flipping a Tack: Frequentist Solution

- What if the experimenter decided to stop measuring when he reached 3 Down?
- For 3D (3 Down), results at least as extreme are
  - 9 U + 3 D (12 flips)
  - 10 U + 3 D (13 flips)
  - 11 U + 3 D (14 flips)
  - 12 U + 3 D (15 flips)
  - 13 U + 3 D (16 flips)
  - $\vdots$
- Using the negative binomial distribution, with $D = 3$, we get

$$p = \binom{11}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{13}{1}\left(\frac{1}{2}\right)^{12} + \binom{14}{3}\left(\frac{1}{2}\right)^{12} + \cdots = 3.27\%$$

- Is significant at the 5% level

Flipping a Tack: Frequentist Solution

- What if the experimenter decided to stop measuring when he reached 3 Down?
- For 3D (3 Down), results at least as extreme are
  - 9 U + 3 D (12 flips)
  - 10 U + 3 D (13 flips)
  - 11 U + 3 D (14 flips)
  - 12 U + 3 D (15 flips)
  - 13 U + 3 D (16 flips)
  - $\vdots$
- Using the negative binomial distribution, with $D = 3$, we get

$$p = \binom{11}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{13}{1}\left(\frac{1}{2}\right)^{12} + \binom{14}{3}\left(\frac{1}{2}\right)^{12} + \cdots = 3.27\%$$

- Is significant at the 5% level

Flipping a Tack: Frequentist Solution

- What if the experimenter decided to stop measuring when he reached 3 Down?
- For 3D (3 Down), results at least as extreme are
  - 9 U + 3 D (12 flips)
  - 10 U + 3 D (13 flips)
  - 11 U + 3 D (14 flips)
  - 12 U + 3 D (15 flips)
  - 13 U + 3 D (16 flips)
  - $\vdots$
- Using the negative binomial distribution, with $D = 3$, we get

$$p = \binom{11}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{13}{1}\left(\frac{1}{2}\right)^{12} + \binom{14}{3}\left(\frac{1}{2}\right)^{12} + \cdots = 3.27\%$$

- Is significant at the 5% level

Flipping a Tack: Frequentist Solution

- What if the experimenter decided to stop measuring when he reached 3 Down?
- For 3D (3 Down), results at least as extreme are
  - 9 U + 3 D (12 flips)
  - 10 U + 3 D (13 flips)
  - 11 U + 3 D (14 flips)
  - 12 U + 3 D (15 flips)
  - 13 U + 3 D (16 flips)
  - $\vdots$
- Using the negative binomial distribution, with $D = 3$, we get

$$p = \binom{11}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{13}{1}\left(\frac{1}{2}\right)^{12} + \binom{14}{3}\left(\frac{1}{2}\right)^{12} + \cdots = 3.27\%$$

- Is significant at the 5% level

Flipping a Tack: Frequentist Solution

- What if the experimenter decided to stop measuring when he reached 3 Down?
- For 3D (3 Down), results at least as extreme are
  - 9 U + 3 D (12 flips)
  - 10 U + 3 D (13 flips)
  - 11 U + 3 D (14 flips)
  - 12 U + 3 D (15 flips)
  - 13 U + 3 D (16 flips)
  - $\vdots$
- Using the negative binomial distribution, with $D = 3$, we get

$$p = \binom{11}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{13}{1}\left(\frac{1}{2}\right)^{12} + \binom{14}{3}\left(\frac{1}{2}\right)^{12} + \cdots = 3.27\%$$

- Is significant at the 5% level

Flipping a Tack: Frequentist Solution

- What if the experimenter decided to stop measuring when he reached 3 Down?
- For 3D (3 Down), results at least as extreme are
  - 9 U + 3 D (12 flips)
  - 10 U + 3 D (13 flips)
  - 11 U + 3 D (14 flips)
  - 12 U + 3 D (15 flips)
  - 13 U + 3 D (16 flips)
  - ⋮
- Using the negative binomial distribution, with $D = 3$, we get

$$p = \binom{11}{3}\left(\frac{1}{2}\right)^{12} + \binom{12}{3}\left(\frac{1}{2}\right)^{12} + \binom{13}{1}\left(\frac{1}{2}\right)^{12} + \binom{14}{3}\left(\frac{1}{2}\right)^{12} + \cdots = 3.27\%$$

- Is significant at the 5% level

Flipping a Tack: Bayesian Solution

- Posterior: β-dist

$$p(\theta|D,U,I) = \frac{(D+U+1)!}{D!U!}\theta^D(1-\theta)^U$$

$$p(\theta|D,U,I) = \frac{13!}{3!9!}\theta^3(1-\theta)^9$$

- Median value: $\theta_{\text{median}} = 0.275$
- Probability for the chance of D less than 50-50: integrate the posterior

$$\int_0^{0.5} d\theta p(\theta|D,U,I) = 0.954$$

- Is significant at the 5% level, and doesn't depend on choice of experiment

Flipping a Tack: Bayesian Solution

- Posterior: β-dist

$$p(\theta|D,U,I) = \frac{(D+U+1)!}{D!U!}\theta^D(1-\theta)^U$$

$$p(\theta|D,U,I) = \frac{13!}{3!9!}\theta^3(1-\theta)^9$$

- Median value: $\theta_{\text{median}} = 0.275$
- Probability for the chance of D less than 50-50: integrate the posterior

$$\int_0^{0.5} d\theta p(\theta|D,U,I) = 0.954$$

- Is significant at the 5% level, and doesn't depend on choice of experiment

Flipping a Tack: Bayesian Solution

- Posterior: β-dist

$$p(\theta|D,U,I) = \frac{(D+U+1)!}{D!U!}\theta^D(1-\theta)^U$$

$$p(\theta|D,U,I) = \frac{13!}{3!9!}\theta^3(1-\theta)^9$$

- Median value: $\theta_{\text{median}} = 0.275$
- Probability for the chance of D less than 50-50: integrate the posterior

$$\int_0^{0.5} d\theta p(\theta|D,U,I) = 0.954$$

- Is significant at the 5% level, and doesn't depend on choice of experiment

Flipping a Tack: Bayesian Solution

- Posterior: β-dist

$$p(\theta|D,U,I) = \frac{(D+U+1)!}{D!U!}\theta^D(1-\theta)^U$$

$$p(\theta|D,U,I) = \frac{13!}{3!9!}\theta^3(1-\theta)^9$$

- Median value: $\theta_{\text{median}} = 0.275$
- Probability for the chance of D less than 50-50: integrate the posterior

$$\int_0^{0.5} d\theta p(\theta|D,U,I) = 0.954$$

- Is significant at the 5% level, and doesn't depend on choice of experiment

Flipping a Tack: Bayesian Solution

- Posterior: β-dist

$$p(\theta|D,U,I) = \frac{(D+U+1)!}{D!U!}\theta^D(1-\theta)^U$$

$$p(\theta|D,U,I) = \frac{13!}{3!9!}\theta^3(1-\theta)^9$$

- Median value: $\theta_{\mathrm{median}} = 0.275$
- Probability for the chance of D less than 50-50: integrate the posterior

$$\int_0^{0.5} d\theta p(\theta|D,U,I) = 0.954$$

- Is significant at the 5% level, and doesn't depend on choice of experiment