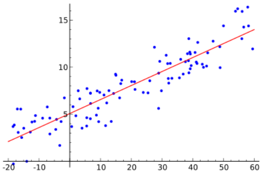
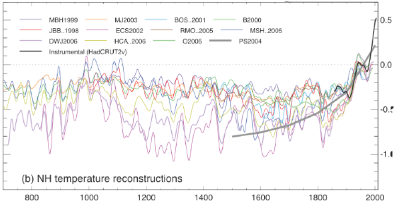
Classical vs. Bayesian Statistics: A Short Introduction

Conor Mayo-Wilson

University of Washington

Summer school in mathematical philosophy for women
July 27th, 2015





Statistics is used in virtually every scientific discipline.

- During the scientific revolution, epistemology and science were intimately connected.
- In the 20th century, statistics has replaced epistemology in its influence on working scientists.

Question: So why should philosophers be interested in statistics?

Answer:

- As the mathematical study of inductive inference, statistics grapples with the many of the same difficult philosophical questions as epistemologists.
- So there are several philosophies of statistics that deserve philosophical scrutiny.

Three Schools

Three "schools" of statistical inference:

1. Classical
2. Bayesian
3. Likelihoodism

Central Difference:

- Bayesians argue that an experimenter's subjective degrees of beliefs and subjective utility functions do influence (and should influence) his or her inferences.
- Classical statisticians argue otherwise.

Subjectivity in statistics

Question: Why should subjective degrees of belief and utility play or not play a role in statistical inference?

Question: Why should subjective degrees of belief and utility play or not play a role in statistical inference?

Classical Statistician's Answer: Subjectivity threatens the objectivity of science

- E.g., A medical researcher ought not tell the FDA that, given her personal values and prior convictions, a new potentially dangerous drug seems acceptable for widespread use.

Question: Why should subjective degrees of belief and utility play or not play a role in statistical inference?

Bayesian Answer:

- All statistical procedures require subjective elements; the difference is Bayesian techniques make the reliance explicit.
- Rationality axioms entail that one ought to behave as a Bayesian.
- Subjectivity "washes out" in the long run.

Question: Why does the difference matter for methodology?

Answer:

- Bayesians can calculate a posterior distribution $P(\theta|X_1, \ldots X_n)$ given the data, and minimize expected loss relative to it.

Question: Why does the difference matter for methodology?

Answer:

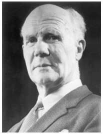
- Bayesians can calculate a posterior distribution $P(\theta|X_1, \ldots X_n)$ given the data, and minimize expected loss relative to it.
- Because there is no well-defined prior probability of θ in the classical statistician's framework, the posterior distribution is also not well-defined.
    - So classical statisticians have a number of criteria used to assess the reliability of different estimators.

Today:

- An illustration of the distinction between classical and Bayesian hypothesis testing
- Discussion of philosophical differences

Classical Hypothesis Tests

vs.



"Classical" hypothesis testing is really two different sets of techniques.

- Fisher
- Neyman-Pearson

Fisher vs. Neyman-Pearson

	Fisher	Neyman-Pearson
Alternative hypothesis?	No	Yes
Key Concepts	P-value	Size and Power
Normative Upshot	Evidential	Behavioral
One-shot?	Single Experiment	Long Run

To review the differences between two traditions, here's a toy example:

- Suppose a coin factory produces two types of coins, one with bias $\frac{1}{4}$ and the other with bias $\frac{3}{4}$.
  - The bias of a coin is the objective probability that the coin will land heads when tossed.
  - This objective probability might be a long-run frequency, a "logical" fact, or a propensity.

Hypothesis Testing - Toy Example

- Suppose a coin factory produces two types of coins, one with bias $\frac{1}{4}$ and the other with bias $\frac{3}{4}$.
  - The bias of a coin is the objective probability that the coin will land heads when tossed.
  - This objective probability might be a long-run frequency, a "logical" fact, or a propensity.
- Suppose your null-hypothesis Θ_0 is that the coin has bias $\frac{1}{4}$.

Hypothesis Testing - Toy Example

Suppose you have flipped the coin 52 times, and observed 26 heads.

- For each data sequence x, there is a set of "more extreme values" E_x.
  - E.g., If x has n_x heads, then E_x might be all data sequences containing at least n_x many heads.

- For each data sequence x, there is a set of "more extreme values" E_x.
  - E.g., If x has n_x heads, then E_x might be all data sequences containing at least n_x many heads.
- In general, the P-value of the observed outcome x is defined as:

$$\sup_{\theta \in \Theta_0} P_\theta(E_x)$$

where Θ_0 is the null hypothesis.

- In the example, if E_x is all sequences involving more heads those observed, then the P-value is

$$P_{\frac{1}{4}}(\text{Number Heads} \geq 26) = 0.000090219654714 00772$$

- Typically, P-values are computed using some test statistic.

Test Statistics

- Typically, P-values are computed using some test statistic.
- A test statistic T "summarizes" the observed data:
  - Sample mean: Average your observations

Test Statistics

- Typically, P-values are computed using some test statistic.
- A test statistic T "summarizes" the observed data:
  - Sample mean: Average your observations
  - Sample Variance: See how "spread out" your observations are.

Test Statistics

- Typically, P-values are computed using some test statistic.
- A test statistic T "summarizes" the observed data:
  - Sample mean: Average your observations
  - Sample Variance: See how "spread out" your observations are.
  - The Sample Itself!

Fisherian Testing

- Fisher interprets a low P-value as strong evidence against the null hypothesis.
  - Notice the "evidential" interpretation.
  - Notice the evidence is **against** the null hypothesis, not for some alternative.
- Reject the null hypothesis if the P-value is low.

Objection 1: The choice of null hypothesis is arbitrary

- In the example, the sample may seem like strong evidence against the null, but if the coin factory produces only the two types of coins, it's equally strong evidence against $\Theta_1 = \{\frac{3}{4}\}$.

Objection 2: The choice of test-statistic is arbitrary.

- Why the number of heads? Why not the number of heads on even tosses?

Objection 3: The choice of the set of extreme values is arbitrary.

- Suppose 13 heads rather than 26 heads had been observed in the example. Would Fisher reject the null hypothesis under the reasoning the data is "too good to be true"?
- Note: Fisher did something similar in rejecting Mendel's data.

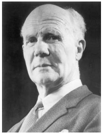
Objection 4: Different P-values can be obtained from the same evidence, and hence, the P-value cannot be a measure of the strength of evidence again the null hypothesis.

Responses:

- There are ways of responding to each objection, but we have to move on.

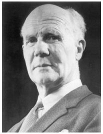
Responses:

- There are ways of responding to each objection, but we have to move on.
- Neyman and Pearson tests, however, already address these three objections (at least in special cases) ...

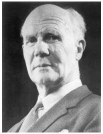


- Neyman and Pearson argue that a null-hypothesis Θ_0 ought to always be tested against an alternative Θ_1.

Neyman and Pearson



- Neyman and Pearson argue that a null-hypothesis Θ_0 ought to always be tested against an alternative Θ_1.
- Consequently, they aim to minimize two types of error:
  - **Type I Error:** Rejecting the null when it's true.
  - **Type II Error:** Accepting the null when it's false.

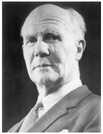


- Suppose you employ a test that rejects the null hypothesis if the data $\mathbf{X}$ belongs to a set R called the rejection region.
- The size of a test is the greatest chance of committing a Type I error:

$$\sup_{\theta \in \Theta_0} P_\theta(\mathbf{X} \in R)$$

- The Power is the greatest chance of committing a Type II error:

$$\sup_{\theta \in \Theta_1} P_\theta(\mathbf{X} \notin R)$$



Consider the example.

- Suppose your rejection region R is the set of coin flips containing more heads than tails:
- The size of this test is

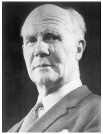
$$P_{\frac{1}{4}}(\text{Num Heads} \geq 27)$$

- The Power of this test is:

$$P_{\frac{3}{4}}(\text{Num Heads} < 27)$$

- Clearly, there is a tradeoff between size and power.
- You can minimize the chance of Type I error by always retaining the null hypothesis.
- You can minimize the chance of Type II error by always rejecting the null.

How do Neyman and Pearson navigate this tradeoff?



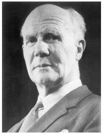
- First, fix the size α of the test (customary is .05).
- Then find a rejection region R that maximizes the power if the size is α.

Question: Aren't Neyman and Pearson subject to the same objections as Fisher?

Answer: Not always.

Objections

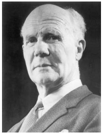
vs.



Objection 1: The choice of the null hypothesis is arbitrary.
Answer: Recall the difference between Fisher and Neyman and Pearson's interpretations of hypothesis tests ...

But we may look at the purpose of tests from another view-point. Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour with regard to them, in following which we insure that, in the long run of experience, we shall not be too often wrong. Here, for example, would be such a "rule of behaviour": to decide whether a hypothesis, H, of a given type be rejected or not, calculate a specified character, x, of the observed facts ; if $x > x_0$, reject H, if $x \leq x_0$,, accept H. **Such a rule tells us nothing as to whether in a particular case H is true when $x \leq x_0$, or false when $x > x_0$.** *But it may often be proved that if we behave according to such a rule, then in the long run me shall reject H when it is true not more, say, than once in a hundred times, and in addition we may have evidence that we shall reject H sufficiently often when it is false.*

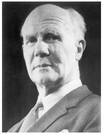
vs.



Objection 1: The choice of the null hypothesis is arbitrary.
Answer: The null hypothesis is chosen with respect to one's long-term goals.

- Since one hypothesis may be more important than another, it may be more important to minimize Type I error in the long run.

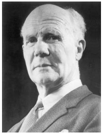
vs.



Objection 1: The choice of the null hypothesis is arbitrary.
Answer: The null hypothesis is chosen with respect to one's long-term goals.

- Since one hypothesis may be more important than another, it may be more important to minimize Type I error in the long run.
- So even if the evidence is symmetric between two hypotheses, the choice of the null is not arbitrary.

vs.



Objection 2: The choice of the rejection region is arbitrary.
Answer: For Neyman and Pearson, in the example, the choice of rejection region in the example is uniquely determined if one fixes the size α of the test.

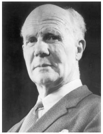
Namely, it is the set of observable data sequences $\mathbf{X}$ such that:

$$\frac{P_{\frac{1}{4}}(\mathbf{X})}{P_{\frac{3}{4}}(\mathbf{X})} \leq k_\alpha$$

where k_α is a constant depending upon α.

Objections

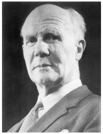
vs.



Objection 3: The choice of the test statistic is arbitrary.

Answer: For Neyman and Pearson, the test statistic should be sufficient. We can't talk about this here.

vs.



Objection 4: The P-value is not a measure of evidential strength.

Answer: That's right. The size of the test, which is closely related to the P-value for Fisher, is a measure of long run correctness.

Bayesian Statistics

Bayesianism: Statistical inference is just like all other decision problems: maximize subjective expected utility!

	Sun	Rain
Read	2	3
Biergarten	4	-2
Watch "Glee"	-10	-10

	Sun	Rain
Read	2	3
Biergarten	4	-2
Watch "Glee"	-10	-10

Decision Matrices: Represent your available actions (in rows) and possible states of the world (in columns).

	Sun	Rain
Read	2	3
Biergarten	4	-2
Watch “Glee”	-10	-10

Decision Matrices: Subjective utilities (i.e., payoffs) to the decision-maker depend upon the unknown state of nature and what action she chooses.

Decision Rule

A decision rule is a method for choosing an action given a decision matrix and one's beliefs about likelihood of various states of the world.

Formally, it is function that

- takes as input (i) a decision matrix and (ii) a probability distribution over states of the world, and
- outputs an action from the decision matrix.

Dominance

	Sun	Rain
Read	2	3
Biergarten	4	-2
Watch “Glee”	-10	-10

Dominance

	Sun	Rain
Read	2	3
Biergarten	4	-2
Watch "Glee"	-10	-10

Dominance: If the outcome of some action a_1 (e.g., Watch Glee) is worse than that of another a_2 (e.g., Read) regardless of the state of the world, do not choose a_1.

Worst-Case

	Sun	Rain
Read	2	3
Biergarten	4	-3

Worst-Case

	Sun	Rain
Read	2	3
Biergarten	4	-3

Worst-Case: Each action has a worst-case payoff. E.g., For Read, it's 2. For Biergarten, it's -3.

	Sun	Rain
Read	2	3
Biergarten	4	-3

	Sun	Rain
Read	2	3
Biergarten	4	-3

Minimax: Pick the action with the best worst-case payoff. Here, it's Read.

- But suppose you look outside, and it's a beautiful spring day in Munich.

- But suppose you look outside, and it's a beautiful spring day in Munich.
- You read the weather forecast, which claims the chance of rain is .5%.

- But suppose you look outside, and it's a beautiful spring day in Munich.
- You read the weather forecast, which claims the chance of rain is .5%.
- Minimax ignores how likely you think rain is.

- But suppose you look outside, and it's a beautiful spring day in Munich.
- You read the weather forecast, which claims the chance of rain is .5%.
- Minimax ignores how likely you think rain is.
- We'd like some decision rule that simultaneously considers payoffs/losses and likelihood.

Suppose you fully believe the weather forecast, which claims the chance of rain is .5%.

	Sun	Rain
Read	2	3
Biergarten	4	-3

The expected utility of Biergarten is:

$$\begin{aligned} \textsc{seu}(\mathit{Biergarten}) &= p(\mathit{Sun}) \cdot 4 + p(\mathit{Rain}) \cdot -3 \\ &= 995 \cdot 4 + .005 \cdot -3 \\ &= 3.965 \end{aligned}$$

Suppose you fully believe the weather forecast, which claims the chance of rain is .5%.

	Sun	Rain
Read	2	3
Biergarten	4	-3

In contrast, expected utility of Read is:

$$\begin{aligned}
\textsc{seu}(\mathit{Read}) &= p(\mathit{Sun}) \cdot 2 + p(\mathit{Rain}) \cdot 3 \\
&= 995 \cdot 2 + .005 \cdot 3 \\
&= 2.005
\end{aligned}$$

Three Decision Rules

- Maximize (subjective) expected utility (SEU)

THREE DECISION RULES

- Maximize (subjective) expected utility (SEU)
- Dominance
- Minimax

The Standard:

- According to Bayesian statisticians, an agent is rational if she acts as if she were maximizing expected utility.

The Standard:

- According to Bayesian statisticians, an agent is rational if she acts as if she were maximizing expected utility.
- There are a number of arguments for the claim that expected utility maximization is the **unique** rational decision rule; we won't discuss them here.
  - General idea: Postulate axioms on rational preference and show that, if an agent's preferences obey said axioms, she maximizes expected utility.

Hypothesis Testing - Toy Example

Let's reconsider our toy example:

- Suppose a coin factory produces two types of coins, one with bias $\frac{1}{4}$ and the other with bias $\frac{3}{4}$.

Hypothesis Testing - Toy Example

Let's reconsider our toy example:

- Suppose a coin factory produces two types of coins, one with bias $\frac{1}{4}$ and the other with bias $\frac{3}{4}$.
- The null-hypothesis Θ_0 is that the coin has bias $\frac{1}{4}$.

	Null is true	Alternative is true
Accept null	0	-1
Reject	-2	0

We can use decision matrices in scientific contexts as well.

- Note: The payoffs above are arbitrary.

Priors vs. Posteriors:

- Just as in the rain example, you may believe the null is more (or less) likely to be true.
- Your prior distribution summarizes your personal probabilities (i.e., your degrees of belief) **before** the experiment is conducted.
- Your posterior distribution summarizes your personal probabilities **after** experiment is conducted.

Prior Distribution in Toy Example

Example: Prior Distribution

- Suppose before the experiment, you think the null is twice as likely to be true:
  - $P(\Theta_0) = \frac{2}{3}$ and $P(\Theta_1) = \frac{1}{3}$.
- This is your prior distribution.

Example: Posterior Distribution

- Suppose your evidence E is that 26 heads are observed in 52 throws.
- Your update your degrees of belief as follows:

$$P_{new}(\Theta_0) = P(\Theta_0|E) = \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(E)}$$

Posterior Distribution in Toy Example

$$P(\Theta_0|E) = \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(E)}$$

Posterior Distribution in Toy Example

$$
\begin{aligned}
P(\Theta_0|E) &= \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(E)} \\
&= \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(E\&\Theta_0) + P(E\&\Theta_1)}
\end{aligned}
$$

Posterior Distribution in Toy Example

$$
\begin{aligned}
P(\Theta_0|E) &= \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(E)} \\
&= \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(E\&\Theta_0) + P(E\&\Theta_1)} \\
&= \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(\Theta_0) \cdot P(E|\Theta_0) + P(\Theta_1) \cdot P(E|\Theta_1)}
\end{aligned}
$$

Posterior Distribution in Toy Example

$$
\begin{aligned}
P(\Theta_0|E) &= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(E)} \\
&= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(E\&\Theta_0)+P(E\&\Theta_1)} \\
&= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(\Theta_0)\cdot P(E|\Theta_0)+P(\Theta_1)\cdot P(E|\Theta_1)} \\
&= \frac{[\frac{1}{4}^{26}\cdot\frac{3}{4}^{26}\cdot\binom{26}{52}]\cdot\frac{2}{3}}{\frac{2}{3}\cdot[\frac{1}{4}^{26}\cdot\frac{3}{4}^{26}\cdot\binom{26}{52}]+\frac{1}{3}\cdot[\frac{1}{4}^{26}\cdot\frac{3}{4}^{26}\cdot\binom{26}{52}]}
\end{aligned}
$$

Posterior Distribution in Toy Example

$$
\begin{aligned}
P(\Theta_0|E) &= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(E)} \\
&= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(E\&\Theta_0)+P(E\&\Theta_1)} \\
&= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(\Theta_0)\cdot P(E|\Theta_0)+P(\Theta_1)\cdot P(E|\Theta_1)} \\
&= \frac{[\frac{1}{4}^{26}\cdot\frac{3}{4}^{26}\cdot\binom{26}{52}]\cdot\frac{2}{3}}{\frac{2}{3}\cdot[\frac{1}{4}^{26}\cdot\frac{3}{4}^{26}\cdot\binom{26}{52}]+\frac{1}{3}\cdot[\frac{1}{4}^{26}\cdot\frac{3}{4}^{26}\cdot\binom{26}{52}]} \\
&= \frac{2}{3}
\end{aligned}
$$

Hence, after observing 26 heads, your new degrees of belief are given by:

$$\begin{aligned} P_{new}(\Theta_0) &= P(\Theta_0|E) = \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(E)} = \frac{2}{3} \\ P_{new}(\Theta_1) &= P(\Theta_0|E) = \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(E)} = \frac{1}{3} \end{aligned}$$

Posterior Distribution in Toy Example

Intuitive Explanation: Since the evidence is equally bad for both hypotheses, your degrees of belief do not change.

Maximizing Hypothesis Testing:

Question: How does a Bayesian decide which hypothesis to accept?

Answer: She maximizes expected utility as before!

Bayesian Hypothesis Testing

	Θ_0	Θ_1
Accept Θ_0	0	-1
Reject Θ_0	-2	0

The expected utility of Accept is:

$$\begin{aligned} \mathrm{SEU}(\mathit{Accept}) &= P_{new}(\Theta_0) \cdot 0 + P_{New}(\Theta_1) \cdot -1 \\ &= -\frac{1}{3} \end{aligned}$$

	Θ_0	Θ_1
Accept Θ_0	0	-1
Reject Θ_0	-2	0

In contrast, expected utility of Reject is:

$$\begin{aligned} \textsc{seu}(\mathit{Reject}) &= P_{new}(\Theta_1) \cdot -2 + P_{New}(\Theta_1) \cdot 0 \\ &= -\frac{2}{3} \end{aligned}$$

	Θ_0	Θ_1
Accept Θ_0	0	-1
Reject Θ_0	-2	0

In contrast, expected utility of Reject is:

$$\begin{aligned} \textsc{seu}(\mathit{Reject}) &= P_{new}(\Theta_1) \cdot -2 + P_{New}(\Theta_1) \cdot 0 \\ &= -\frac{2}{3} \end{aligned}$$

So you ought to accept!

Contrast:

- The Bayesian experimenter's prior distribution and utilities dictate what she chooses to do in this example.
- If the experimenter (you!) had a different prior distribution and utility function, your decision would differ.
- It is for reasons like this that classical statisticians think Bayesian statistics is too subjective.

Example: Posterior Distribution

- Suppose your evidence E is that 39 heads are observed in 52 throws.
- Your update your degrees of belief as follows:

$$P_{new}(\Theta_0) = P(\Theta_0|E) = \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(E)}$$

Posterior Distribution in Toy Example

$$P(\Theta_0|E) = \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(E)}$$

Posterior Distribution in Toy Example

$$
\begin{aligned}
P(\Theta_0|E) &= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(E)} \\
&= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(E\&\Theta_0)+P(E\&\Theta_1)}
\end{aligned}
$$

Posterior Distribution in Toy Example

$$
\begin{aligned}
P(\Theta_0|E) &= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(E)} \\
&= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(E\&\Theta_0)+P(E\&\Theta_1)} \\
&= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(\Theta_0)\cdot P(E|\Theta_0)+P(\Theta_1)\cdot P(E|\Theta_1)}
\end{aligned}
$$

Posterior Distribution in Toy Example

$$
\begin{aligned}
P(\Theta_0|E) &= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(E)} \\
&= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(E\&\Theta_0)+P(E\&\Theta_1)} \\
&= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(\Theta_0)\cdot P(E|\Theta_0)+P(\Theta_1)\cdot P(E|\Theta_1)} \\
&= \frac{[\frac{1}{4}^{39}\cdot\frac{3}{4}^{13}\cdot\binom{13}{52}]\cdot\frac{2}{3}}{\frac{2}{3}\cdot[\frac{1}{4}^{39}\cdot\frac{3}{4}^{13}\cdot\binom{13}{52}]+\frac{1}{3}\cdot[\frac{3}{4}^{39}\cdot\frac{1}{4}^{13}\cdot\binom{13}{52}]}
\end{aligned}
$$

Posterior Distribution in Toy Example

$$
\begin{aligned}
P(\Theta_0|E) &= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(E)} \\
&= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(E\&\Theta_0)+P(E\&\Theta_1)} \\
&= \frac{P(E|\Theta_0)\cdot P(\Theta_0)}{P(\Theta_0)\cdot P(E|\Theta_0)+P(\Theta_1)\cdot P(E|\Theta_1)} \\
&= \frac{[\frac{1}{4}^{39}\cdot\frac{3}{4}^{13}\cdot\binom{13}{52}]\cdot\frac{2}{3}}{\frac{2}{3}\cdot[\frac{1}{4}^{39}\cdot\frac{3}{4}^{13}\cdot\binom{13}{52}]+\frac{1}{3}\cdot[\frac{3}{4}^{39}\cdot\frac{1}{4}^{13}\cdot\binom{13}{52}]} \\
&\approx 7.86\cdot 10^{-13}
\end{aligned}
$$

Posterior Distribution in Toy Example

Hence, after observing 26 heads, your new degrees of belief are given by:

$$
\begin{aligned}
P_{new}(\Theta_0) &= P(\Theta_0|E) = \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(E)} \approx 7.86 \cdot 10^{-13} \approx 0 \\
P_{new}(\Theta_1) &= P(\Theta_0|E) = \frac{P(E|\Theta_0) \cdot P(\Theta_0)}{P(E)} \approx 1
\end{aligned}
$$

Posterior Distribution in Toy Example

Intuitive Explanation: Since the evidence is fairly strong for the alternative hypothesis, your degrees of belief change radically.

Bayesian Hypothesis Testing

	Θ_0	Θ_1
Accept Θ_0	0	-1
Reject	-2	0

$$
\begin{aligned}
\textsc{seu}(\mathit{Accept}) &= P_{new}(\Theta_0)\cdot 0 + P_{New}(\Theta_1)\cdot -1 \approx -1 \\
\textsc{seu}(\mathit{Reject}) &= P_{new}(\Theta_1)\cdot -2 + P_{New}(\Theta_1)\cdot 0 \approx -2
\end{aligned}
$$

Posterior Distribution in Toy Example

Moral: With sufficiently strong evidence, your prior distribution "washes out" and makes no difference in inference.

Central Difference:

- Bayesians argue that an experimenter's subjective degrees of beliefs and subjective utility functions do influence (and should influence) his or her inferences.
- Classical statisticians argue otherwise.

Question: Why should subjective degrees of belief and utility play or not play a role in statistical inference?

- Classical - Subjectivity threatens the objectivity of science

Question: Why should subjective degrees of belief and utility play or not play a role in statistical inference?

- Classical - Subjectivity threatens the objectivity of science
- Bayesian
  - All statistical procedures require subjective elements; the difference is Bayesian techniques make the reliance explicit.
  - Rationality axioms entail that one ought to behave as a Bayesian.
  - Subjectivity "washes out" with enough data.

Question: Where can I learn more?

Answer: I taught a philosophy of statistics course which contains links to many textbooks and papers:

`http://mayowilson.org/Past_Courses.htm`