Errors and inferences

Readings for today

- Mayo, D. G. (1997). Error statistics and learning from error: making a virtue of necessity. Philosophy of Science, 64(S4), S195-S212.

Topics

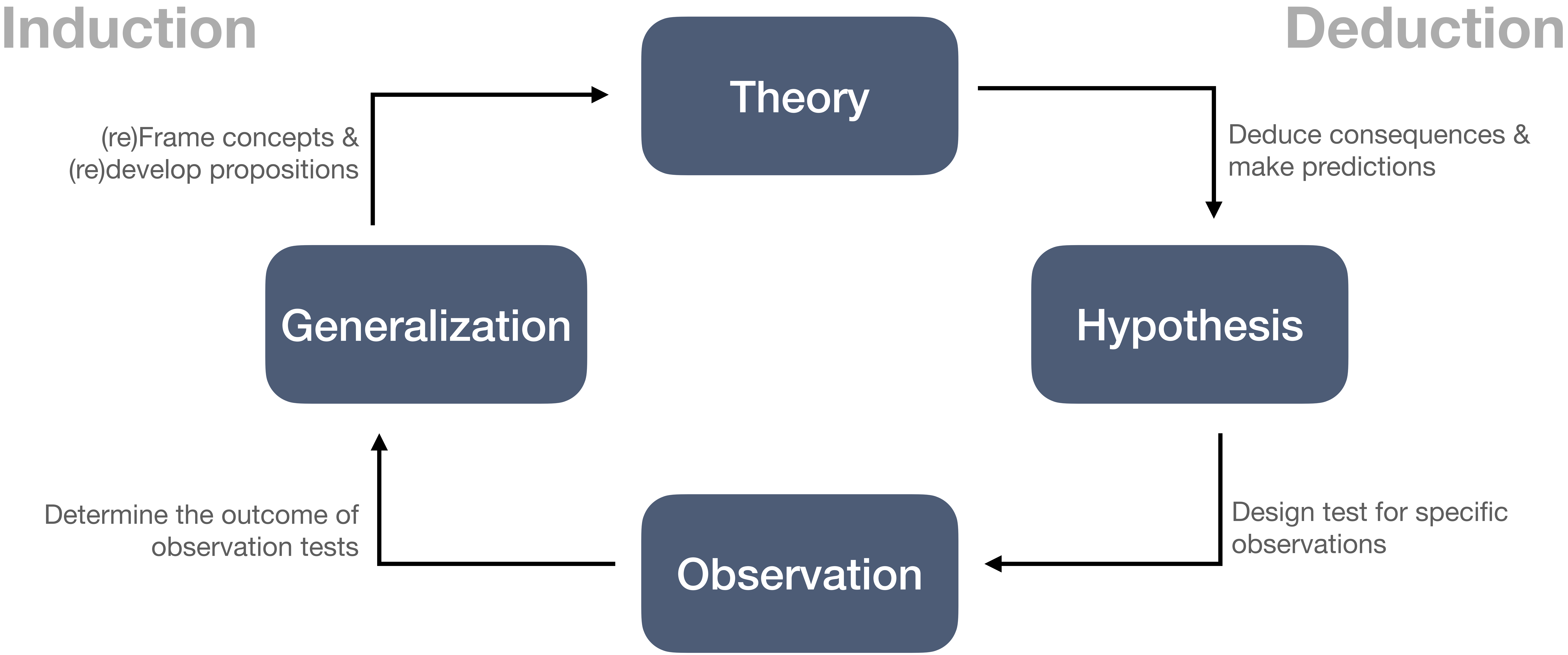
1. Evidentiary vs. existential approaches

2. A framework of inquiry

3. A case example

Evidentiary vs. Existential approaches

Hypothetico-deductive model of science



Wallace, W. L. (1971). *The Logic of Science in Sociology*.

(Hempel 1966; Popper 1935, 1963)

The problem of demarcation

Problem: What distinguishes a scientific (aka- empirical) theory from a metaphysical theory (aka- non-empirical)?

Modus ponens (induction)

$$P \rightarrow Q, P \vdash Q$$

If P then Q, thus if P is true Q is also true.

Example: "There are black swans."

Exhaustive search required.

Modus tollens (deduction)

$$P \rightarrow Q, \neg Q \vdash \neg P$$

If P then Q, thus if Q is false P is also false.

Example: "All swans are white."

Just 1 counter point to disprove.

(Popper 1959)

Two philosophies for evaluating hypotheses

Evidentiary (inductive, Bayesian)

How well does evidence confirm or support a hypothesis?

evidential-relation (ER)

Existential (deductive, frequentist)

How does the evidence discriminate between hypotheses?

error statistical

(Mayo 1997)

Bayes Factor

What is the ratio of evidence for two hypotheses?

$$BF_{01} = \frac{P(H_0 \mid Y)}{P(H_1 \mid Y)} = \underbrace{\boxed{\frac{P(Y \mid H_0)}{P(Y \mid H_1)}}}_{\text{likelihood}} \cdot \underbrace{\boxed{\frac{P(H_0)}{P(H_1)}}}_{\text{prior}}$$

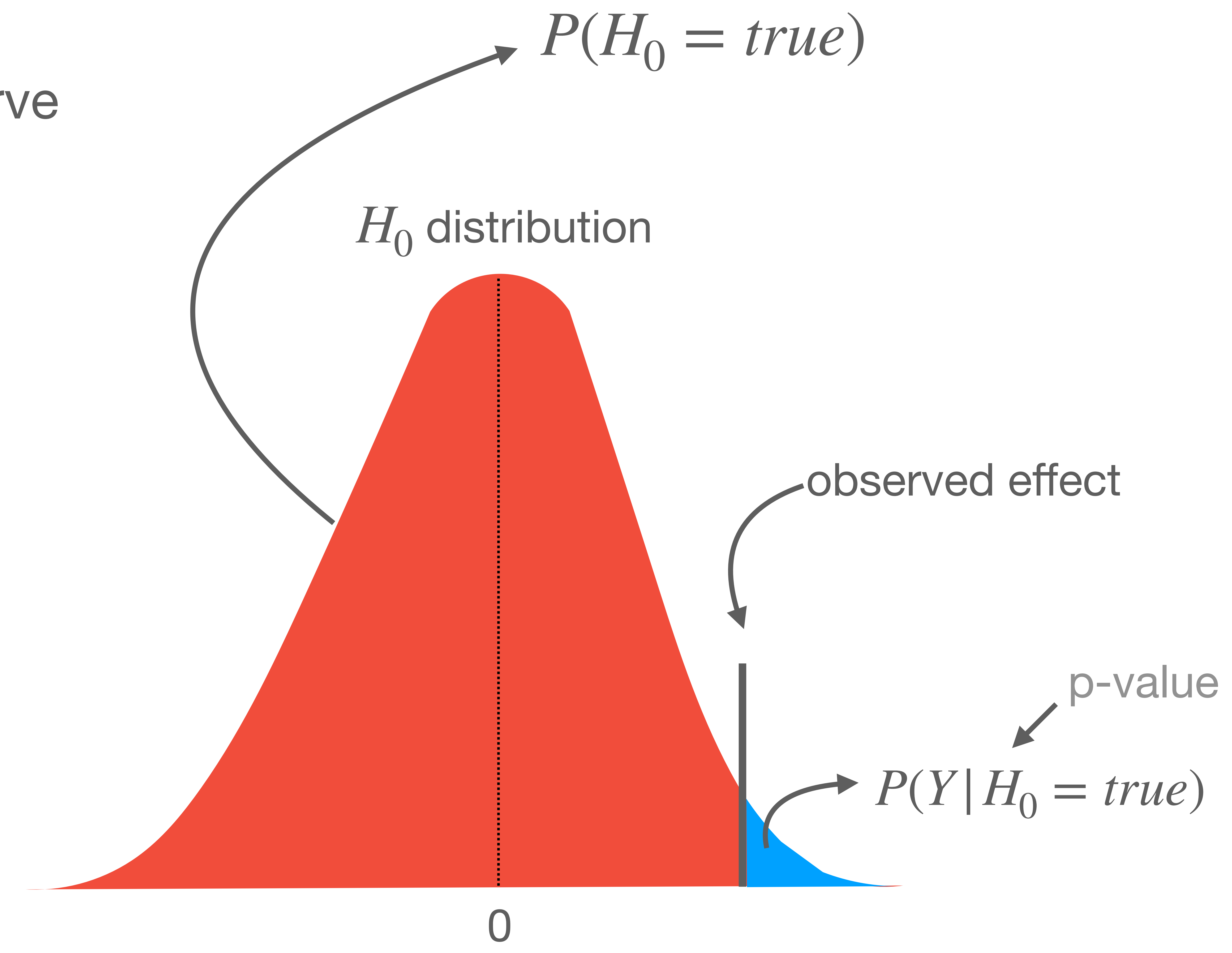
Bayes Factor for H_0 (against H_1)

- Determine the relative evidence for one hypothesis against the other.
- BF_{ij} identifies whether the observed data are more likely to arise from hypothesis i (H_i) than from hypothesis j (H_j).

(Wagenmakers 2007)

The null hypothesis (H_0)

$\underline{H_0}$: The probability that you observe your data if your predicted relationships are not true.



(Wasserstein et al. 2019)

A difference in aims

Evidentiary (inductive, Bayesian)

Update the degree of belief in a hypothesis based on the evidence and prior probabilities.

Existential (deductive, frequentist)

Use errors as a means of evaluating the existence of a given hypothesis.

Virtues

- Includes evaluations *before* evidence is collected.
- Does *not* equate scientific inference with a direct application of a statistical test.

(Mayo 1997)

The error statistical approach

Goal

Determine the *error probability* of individual hypotheses.

- Origins trace back to the Neyman-Pearson model of statistics.
- Reject use of prior probabilities.
- Evaluate models in how well they discriminate between alternative hypotheses.
- A model is only as good as how well it facilitates learning from error.
- One can ***only*** deduce incorrect models.

(Mayo 1997)

A framework of inquiry

A framework of inquiry

Models of experimental inquiry

Models of the ***primary question of interest***. The high-level, theoretical hypotheses being evaluated.

Models of the ***statistical form*** that link the primary models to the data itself.

Models of the ***assumptions*** about the form of the data (e.g., distribution assumptions)

(Mayo 1997)

Learning from errors

1) After-trial checking

After data are collected, develop methods for evaluating how well the data matches the assumptions of the data model you started with (e.g., if you assume Gaussian data distributions, run a test for whether your data are, in fact, Gaussian).

2) Before-trial planning

Knowledge of past mistakes can influence the experimental design in order to prevent those mistakes in the future (e.g., introducing “catch” trials to confirm a participant is following instructions).

(Mayo 1997)

Learning from errors

3) An error repertoire

Generate a list of known errors that have occurred in the past to either avoid (before-trial) or check (after-trial) for in the future (e.g., data cleansing steps).

4) The effects of mistakes

Knowledge of the *effects* of mistakes allows for discriminating effects from artifacts (e.g., fatigue reduces reaction times).

5) Simulating errors

Use simulations (real or artificial) to determine the effects of mistakes on target outcomes (e.g., simulating drift in reaction times over time to see how it impacts predicted group differences).

(Mayo 1997)

Learning from errors

6) Amplifying & listening to error patterns

Push systems to detect small errors by magnifying their effects (e.g., shorten trial windows to force people towards random guessing).

7) Robustness

When do violations of assumptions not impact the accuracy of subsequent inferences?

8) Severely probing error

Point 1-7 provide a basis for detecting errors that, when put together, provide an arsenal for thoroughly evaluating a hypothesis.

(Mayo 1997)

Arguing from error

Severe error probe

“After learning enough about certain types of mistakes, we may construct a testing procedure with an overwhelmingly good chance of revealing the presence of a specific error, if it exists-but not otherwise.”

- If an effect is observed after all possible sources of error have been accounted for or eliminated, then there is a very high probability that the true effect exists.

A case example

Example: breast cancer screening

Question: ***Does a patient have breast cancer?***

Scenario

e : The evidence (an abnormal reading in the cancer screen)

H : Breast cancer is present. $\rightarrow H_0$

J : Breast disease is absent

$P(e \mid H : \text{breast cancer}) = \sim 1$

$P(e \mid J) = \sim 0.01$

If a person is disease free, then the probability of an abnormal reading is low.

(Mayo 1997)

Example: breast cancer screening

Question: ***Does a patient have breast cancer?***

Bayesian criticism

- e is taken as failing to reject H, while rejecting J.
- H passes a sever test, thus H is indicated as having a high likelihood.
- But the *base rate* (i.e., prior) for breast cancer is low and thus, according to Bayes rule, the posterior probability of H is low.
- Therefore H is not indicated, but J is.
- Therefore, the existential inference is unsound.

(Mayo 1997)

Example: breast cancer screening

Question: ***Does a patient have breast cancer?***

Error statistician reply

- e itself is not an indication on H directly, but depends on potential sources of error.
- e is a *poor indication* of H if the probability of e is more extensive than the disease prevalence rate, d (i.e., when $P(e \mid \neg H) > d$).
- e is a *good indication* of H if the probability of e is more extensive than the disease prevalence rate, d (i.e., when $P(e \mid \neg H) < d$).

Severe test

(Mayo 1997)

Take home message

- Inferring the probability of a hypothesis requires a systematic awareness of the sources of errors, at multiple levels of inquiry.
- The evidentiary (Bayesian) and existential (frequentist) approaches differ in their aims, with the latter being more directly accessible to evaluating sources of error.