# And the Bayesians and the frequentists shall lie down together...

Keith Winstein

MIT CSAIL

October 16, 2013



#### Axioms of Probability (1933)

S: a finite set (the sample space)
A: any subset of S (an event)
P(A): the probability of A satisfies

If S infinite, axiom becomes: for an infinite sequence of disjoint subsets  $A_1, A_2, \ldots$ ,

$$P\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i)$$

▲□▶ ▲□▶ ▲臣▶ ▲臣▶ 三臣 - のへで

#### Some Theorems

$$\blacktriangleright P(\overline{A}) = 1 - P(A)$$

- $P(\emptyset) = 0$
- $P(A) \leq P(B)$  if  $A \subset B$
- $P(A) \leq 1$
- $\blacktriangleright P(A \cup B) = P(A) + P(B) P(A \cap B)$

▲ロト ▲圖 ▶ ▲ 臣 ▶ ▲ 臣 ▶ ● 臣 ● のへで

•  $P(A \cup B) \leq P(A) + P(B)$ 

#### Joint & Conditional Probability

- ▶ For  $A, B \subseteq S$ ,  $P(A \cap B)$  is joint probability of A and B.
- The conditional probability of A given B in:

$$P(A|B) = rac{P(A \cap B)}{P(B)}$$

- A and B are **independent** iff  $P(A \cap B) = P(A)P(B)$ .
- A, B independent  $\rightarrow P(A|B) = P(A)$ .

#### Bayes' Theorem

We have:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Therefore:  $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$ 

Bayes' Theorem:

$$P(A|B) = rac{P(B|A)P(A)}{P(B)}$$

On the islands of Ste. Frequentiste and Bayesienne...

▲□▶ ▲圖▶ ▲圖▶ ▲圖▶ ▲□▶

#### On the islands of Ste. Frequentiste and Bayesienne...



The king has been poisoned!

< ロ > < 同 > < 回 > < 回 >

э

Dear Governor: Attached is a blood test for proximity to the poison. It has a 0% rate of false negative and a 1% rate of false positive. Jail those responsible.

BUT REMEMBER THE NATIONWIDE LAW: You must be 95% certain to send a citizen to jail.

▲□▶ ▲□▶ ▲□▶ ▲□▶ □ ののの

## On Ste. Frequentiste:

Test has a 0% rate of false negative and a 1% rate of false positive. You must be 95% certain to send a citizen to jail.

▲□▶ ▲□▶ ▲□▶ ▲□▶ □ ののの

- P(Positive | GUILTY) = 1
- P(Negative | GUILTY) = 0
- P(Positive | INNOCENT) = 0.01
- ► *P*(*Negative* | INNOCENT) = 0.99

#### How to interpret the law? "We must be 95% certain" $\rightarrow$

## On Ste. Frequentiste:

Test has a 0% rate of false negative and a 1% rate of false positive. You must be 95% certain to send a citizen to jail.

- P(Positive | GUILTY) = 1
- P(Negative | GUILTY) = 0
- P(Positive | INNOCENT) = 0.01
- ► *P*(*Negative* | INNOCENT) = 0.99

#### How to interpret the law? "We must be 95% certain" $\rightarrow P(Jail | INNOCENT) \leq 0.05$

▲□▶ ▲□▶ ▲□▶ ▲□▶ □ ののの

## On Ste. Frequentiste:

Test has a 0% rate of false negative and a 1% rate of false positive. You must be 95% certain to send a citizen to jail.

- P(Positive | GUILTY) = 1
- P(Negative | GUILTY) = 0
- P(Positive | INNOCENT) = 0.01
- P(Negative | INNOCENT) = 0.99

How to interpret the law? "We must be 95% certain"  $\rightarrow P(Jail | INNOCENT) \leq 0.05$ 

#### Can Positive $\rightarrow$ Jail? Yes.

#### On Isle Bayesienne:

Test has a 0% rate of false negative and a 1% rate of false positive. You must be 95% certain to send a citizen to jail.

▲□▶ ▲□▶ ▲□▶ ▲□▶ □ ののの

- P(Positive | GUILTY) = 1
- P(Negative | GUILTY) = 0
- P(Positive | INNOCENT) = 0.01
- ► *P*(*Negative* | INNOCENT) = 0.99

#### How to interpret the law? "We must be 95% certain" $\rightarrow$

#### On Isle Bayesienne:

Test has a 0% rate of false negative and a 1% rate of false positive. You must be 95% certain to send a citizen to jail.

- P(Positive | GUILTY) = 1
- P(Negative | GUILTY) = 0
- P(Positive | INNOCENT) = 0.01
- ► *P*(*Negative* | INNOCENT) = 0.99

#### How to interpret the law? "We must be 95% certain" $\rightarrow P(INNOCENT | Jail) \leq 0.05$

- ▶ "We must be 95% certain"  $\rightarrow$  P(INNOCENT | Jail)  $\leq$  0.05
- Can Positive  $\rightarrow$  Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Positive}) = \frac{P(\text{Positive} | \text{INNOCENT}) P(\text{INNOCENT})}{P(\text{Positive})}$$

- ▶ "We must be 95% certain"  $\rightarrow$  P(INNOCENT | Jail)  $\leq$  0.05
- Can Positive  $\rightarrow$  Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Positive}) = \frac{P(\text{Positive} | \text{INNOCENT}) P(\text{INNOCENT})}{P(\text{Positive})}$$

- ▶ "We must be 95% certain"  $\rightarrow$  P(INNOCENT | Jail)  $\leq$  0.05
- Can Positive  $\rightarrow$  Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{P(\text{Positive} | \text{INNOCENT}) P(\text{INNOCENT})}{P(\text{Positive})}$$

- ▶ "We must be 95% certain"  $\rightarrow$   $\mathsf{P}(\text{INNOCENT} \mid \mathsf{Jail}) \leq 0.05$
- ► Can Positive → Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{P(\text{Positive} | \text{INNOCENT}) P(\text{INNOCENT})}{P(\text{Positive})}$$

- ▶ "We must be 95% certain"  $\rightarrow$  P(INNOCENT | Jail)  $\leq$  0.05
- Can Positive  $\rightarrow$  Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{P(\text{Positive} | \text{INNOCENT}) P(\text{INNOCENT})}{P(\text{Positive})}$$

- ▶ "We must be 95% certain"  $\rightarrow$  P(INNOCENT | Jail)  $\leq$  0.05
- Can Positive  $\rightarrow$  Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{(0.01) P(\text{INNOCENT})}{P(\text{Positive})}$$

- ▶ "We must be 95% certain"  $\rightarrow$   $\mathsf{P}(\text{INNOCENT} \mid \mathsf{Jail}) \leq 0.05$
- ► Can Positive → Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{(0.01) P(\text{INNOCENT})}{P(\text{Positive})}$$

- ▶ "We must be 95% certain"  $\rightarrow$   $\mathsf{P}(\text{INNOCENT} \mid \mathsf{Jail}) \leq 0.05$
- ► Can Positive → Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{(0.01) P(\text{INNOCENT})}{P(\text{Positive})}$$

- ▶ "We must be 95% certain"  $\rightarrow$   $\mathsf{P}(\text{INNOCENT} \mid \mathsf{Jail}) \leq 0.05$
- ► Can Positive → Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{(0.01) \quad P(\text{INNOCENT})}{P(\text{Positive}|\text{INNOCENT}) \quad P(\text{INNOCENT})} + P(\text{Positive}|\text{Guilty}) \quad P(\text{Guilty})$$

- ▶ "We must be 95% certain"  $\rightarrow$   $\mathsf{P}(\text{INNOCENT} \mid \mathsf{Jail}) \leq 0.05$
- ► Can Positive → Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{(0.01) \quad P(\text{INNOCENT})}{P(\text{Positive}|\text{INNOCENT}) \quad P(\text{INNOCENT})} + P(\text{Positive}|\text{Guilty}) \quad P(\text{Guilty})$$

- ▶ "We must be 95% certain"  $\rightarrow$   $\mathsf{P}(\text{INNOCENT} \mid \mathsf{Jail}) \leq 0.05$
- ► Can Positive → Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{(0.01) \quad P(\text{INNOCENT})}{P(\text{Positive}|\text{INNOCENT}) \quad P(\text{INNOCENT})} + P(\text{Positive}|\text{Guilty}) \quad P(\text{Guilty})$$

- ▶ "We must be 95% certain"  $\rightarrow$  P(INNOCENT | Jail)  $\leq$  0.05
- ► Can Positive → Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{(0.01) \quad P(\text{INNOCENT})}{1 - (0.99) \quad P(\text{INNOCENT})}$$

- ▶ "We must be 95% certain"  $\rightarrow$   $\mathsf{P}(\text{INNOCENT} \mid \mathsf{Jail}) \leq 0.05$
- ► Can Positive → Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} \mid \text{Jail}) = \frac{(0.01) \quad P(\text{INNOCENT})}{1 - (0.99) \quad P(\text{INNOCENT})}$$

- ▶ "We must be 95% certain"  $\rightarrow$   $\mathsf{P}(\text{INNOCENT} \mid \mathsf{Jail}) \leq 0.05$
- ► Can Positive → Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{(0.01) \quad P(\text{INNOCENT})}{1 - (0.99) \quad P(\text{INNOCENT})}$$

- ▶ "We must be 95% certain"  $\rightarrow$   $\mathsf{P}(\text{INNOCENT} \mid \mathsf{Jail}) \leq 0.05$
- ► Can Positive → Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{(0.01) \quad P(\text{INNOCENT})}{1 - (0.99) \quad P(\text{INNOCENT})}$$

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

P(Innocent) = ???

- ▶ "We must be 95% certain"  $\rightarrow$   $\mathsf{P}(\text{INNOCENT} \mid \mathsf{Jail}) \leq 0.05$
- ► Can Positive → Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{(0.01) \quad P(\text{INNOCENT})}{1 - (0.99) \quad P(\text{INNOCENT})}$$

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

• P(Innocent) =  $0.9 \rightarrow$ 

- ▶ "We must be 95% certain"  $\rightarrow$   $\mathsf{P}(\text{INNOCENT} \mid \mathsf{Jail}) \leq 0.05$
- ► Can Positive → Jail?
- Apply Bayes' theorem

$$P(\text{INNOCENT} | \text{Jail}) = \frac{(0.01) \quad P(\text{INNOCENT})}{1 - (0.99) \quad P(\text{INNOCENT})}$$

▶  $P(Innocent) = 0.9 \rightarrow P(Innocent | Jail) \approx 0.08$  !!

On the islands of Ste. Frequentiste and Bayesienne...

- More than 1% of Ste. Frequentiste goes to jail.
- On Isle Bayesienne, 10% are assumed guilty, but nobody goes to jail.
- ▶ The disagreement wasn't about math or how to interpret P().

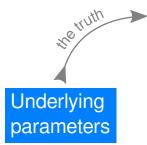
What was it about?

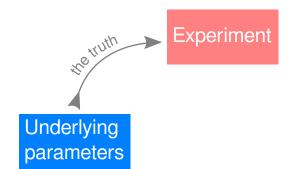
#### The islanders' concerns

- Frequentist cares about the rate of jailings among innocent people. Concern: overall rate of false positive
- Bayesian cares about the rate of innocence among jail inmates. Concern: rate of error among positives
- The Bayesian had to make an assumption about the overall probability of innocence.

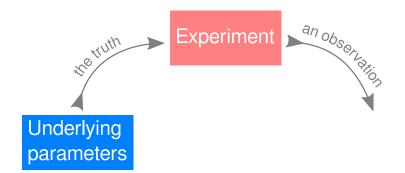
Underlying parameters



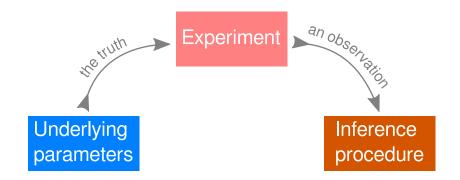






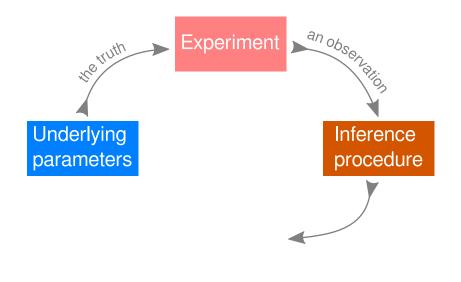


#### Quantifying uncertainty



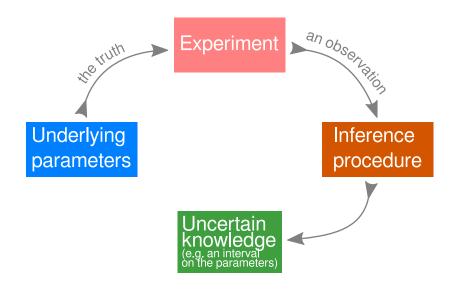
▲ロト ▲圖 ▶ ▲ 臣 ▶ ▲ 臣 ▶ ● 臣 ● のへで

#### Quantifying uncertainty



◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 三臣 - のへで

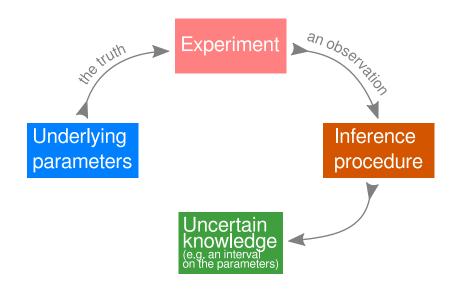
#### Quantifying uncertainty

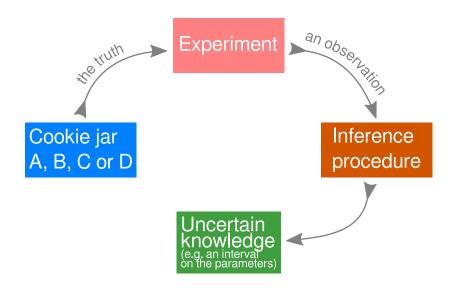


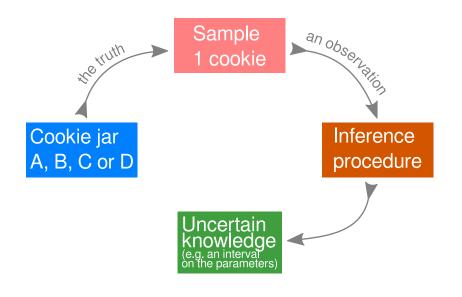
▲ロト ▲御 ▶ ▲ 臣 ▶ ▲ 臣 ▶ ● 臣 ■ ● の Q (2)

#### Jewel's Cookies

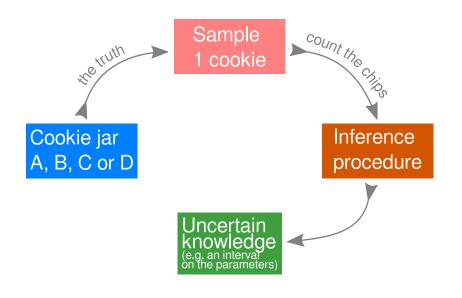
P( chips   jar )	Α	В	C	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
total	100%	100%	100%	100%

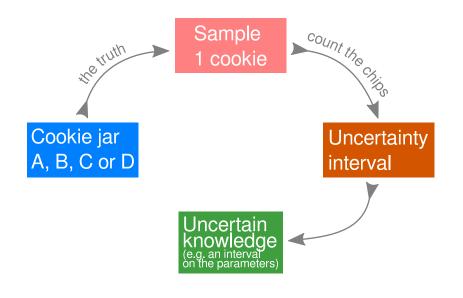




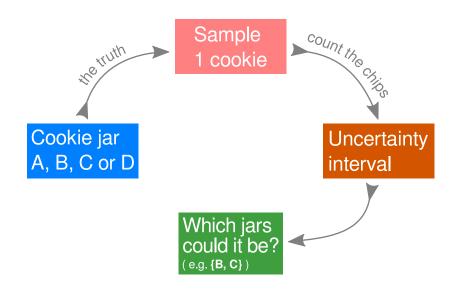


▲□▶▲圖▶▲圖▶▲圖▶ 圖 のQ@





▲□▶ ▲□▶ ▲□▶ ▲□▶ □ ののの



◆□> ◆□> ◆豆> ◆豆> ・豆 ・ のへで

#### Frequentist inference

## A 70% **confidence** interval method includes the correct jar with at least 70% probability **in the worst case, no matter what**.

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

P( chips   jar )	Α	В	C	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
coverage				

P( chips   jar )	Α	В	C	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
coverage	70%			

P( chips   jar )	Α	В	C	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
coverage	70%	25%		

P( chips   jar )	Α	В	C	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
coverage	70%	49%		

Cookie jars **A**, **B**, **C**, **D** have 100 cookies each, but different numbers of chocolate chips per cookie:

P( chips   jar )	Α	В	C	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
coverage	70%	69%		

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

Cookie jars **A**, **B**, **C**, **D** have 100 cookies each, but different numbers of chocolate chips per cookie:

P( chips   jar )	Α	В	C	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
coverage	70%	88%		

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

P( chips   jar )	Α	В	С	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
coverage	70%	88%	67%	

Cookie jars **A**, **B**, **C**, **D** have 100 cookies each, but different numbers of chocolate chips per cookie:

P( chips   jar )	Α	В	C	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
coverage	70%	88%	87%	

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

P( chips   jar )	Α	В	С	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
coverage	70%	88%	87%	70%

# A 70% credible interval has at least 70% conditional probability of including the correct jar, given the observation and the prior assumptions.

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

#### Uniform prior

### Our prior assumption: jars A, B, C, and D have equal probability.

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

#### Conditional probabilities

P( chips   jar )	A	В	C	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
total	100%	100%	100%	100%

#### Conditional probabilities

P( chips   jar )	Α	В	C	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
total	100%	100%	100%	100%

Joint probabilities under uniform prior

$P(\ chips \ \cap \ jar \ )$	Α	В	С	D	P(chips)
0	1/4	<b>12</b> /4	<b>13</b> /4	<b>27</b> /4	13.25%
1	1/4	<b>19</b> /4	<b>20</b> /4	<b>70</b> /4	27.50%
2	70/4	<b>24</b> /4	0/4	1/4	23.75%
3	28/4	20/4	0/4	1/4	12.25%
4	0/4	<b>25</b> /4	<b>67</b> /4	1/4	23.25%
total <i>P(jar</i> )	25%	25%	25%	25%	100%

◆□ ▶ <圖 ▶ < E ▶ < E ▶ E • 9 < 0</p>

Joint probabilities under uniform prior

P( chips $\cap$ jar )	A	В	С	D	P(chips)
0	1/4	<b>12</b> /4	<b>13</b> /4	27/4	13.25%
1	1/4	<b>19</b> /4	<b>20</b> /4	<b>70</b> /4	27.50%
2	70/4	<b>24</b> /4	0/4	1/4	23.75%
3	28/4	20/4	0/4	1/4	12.25%
4	0/4	<b>25</b> /4	<b>67</b> /4	1/4	23.25%
total <i>P(jar</i> )	25%	25%	25%	25%	100%

◆□ ▶ <圖 ▶ < E ▶ < E ▶ E • 9 < 0</p>

	A	В	С	D	P(chips)
0	1/4	12/4	<b>13</b> /4	27/4	13.25%
1	1/4	<b>19</b> /4	20/4	70/4	27.50%
2	70/4	<b>24</b> /4	0/4	1/4	23.75%
3	28/4	20/4	0/4	1/4	12.25%
4	0/4	25/4	<b>67</b> /4	1/4	23.25%
					I

	A	В	С	D	P(chips)
0	1.9	22.6	24.5	50.9	100%
1	1/4	<b>19</b> /4	<b>20</b> /4	<b>70</b> /4	27.50%
2	70/4	<b>24</b> /4	0/4	1/4	23.75%
3	28/4	<b>20</b> /4	0/4	1/4	12.25%
4	0/4	<b>25</b> /4	<b>67</b> /4	1/4	23.25%

	A	В	С	D	P(chips)
0	1.9	22.6	24.5	50.9	100%
1	1/4	<b>19</b> /4	<b>20</b> /4	<b>70</b> /4	27.50%
2	70/4	<b>24</b> /4	0/4	1/4	23.75%
3	28/4	<b>20</b> /4	0/4	1/4	12.25%
4	0/4	<b>25</b> /4	67/4	1/4	23.25%
					'

	Α	В	С	D	P(chips)
0	1.9	22.6	24.5	50.9	100%
1	0.9	17.3	18.2	63.6	100%
2	70/4	<b>24</b> /4	0/4	1/4	23.75%
3	28/4	<b>20</b> /4	0/4	1/4	12.25%
4	0/4	<b>25</b> /4	<b>67</b> /4	1/4	23.25%
					,

	A	В	С	D	P(chips)
0	1.9	22.6	24.5	50.9	100%
1	0.9	17.3	18.2	63.6	100%
2	70/4	<b>24</b> /4	0/4	1/4	23.75%
3	28/4	<b>20</b> /4	0/4	1/4	12.25%
4	0/4	<b>25</b> /4	<b>67</b> /4	1/4	23.25%
					,

	A	В	С	D	P(chips)
0	1.9	22.6	24.5	50.9	100%
1	0.9	17.3	18.2	63.6	100%
2	73.7	25.3	0.0	1.1	100%
3	28/4	20/4	0/4	1/4	12.25%
4	0/4	<b>25</b> /4	<b>67</b> /4	1/4	23.25%

	A	В	С	D	P(chips)
0	1.9	22.6	24.5	50.9	100%
1	0.9	17.3	18.2	63.6	100%
2	73.7	25.3	0.0	1.1	100%
3	28/4	<b>20</b> /4	0/4	1/4	12.25%
4	0/4	<b>25</b> /4	<b>67</b> /4	1/4	23.25%

	Α	В	С	D	P(chips)
0	1.9	22.6	24.5	50.9	100%
1	0.9	17.3	18.2	63.6	100%
2	73.7	25.3	0.0	1.1	100%
3	57.1	40.8	0.0	2.0	100%
4	0/4	<b>25</b> /4	<b>67</b> /4	1/4	23.25%

	A	В	С	D	P(chips)
0	1.9	22.6	24.5	50.9	100%
1	0.9	17.3	18.2	63.6	100%
2	73.7	25.3	0.0	1.1	100%
3	57.1	40.8	0.0	2.0	100%
4	0/4	<b>25</b> /4	67/4	1/4	23.25%

	A	В	С	D	P(chips)
0	1.9	22.6	24.5	50.9	100%
1	0.9	17.3	18.2	63.6	100%
2	73.7	25.3	0.0	1.1	100%
3	57.1	40.8	0.0	2.0	100%
4	0.0	26.9	72.0	1.1	100%

Posterior probabilities under uniform prior

P( jar   chips )	Α	В	С	D	P(chips)
0	1.9	22.6	24.5	50.9	100%
1	0.9	17.3	18.2	63.6	100%
2	73.7	25.3	0.0	1.1	100%
3	57.1	40.8	0.0	2.0	100%
4	0.0	26.9	72.0	1.1	100%

<□ > < @ > < E > < E > E のQ @

P( jar   chips )	Α	В	С	D	credibility
0	1.9	22.6	24.5	50.9	
1	0.9	17.3	18.2	63.6	
2	73.7	25.3	0.0	1.1	
3	57.1	40.8	0.0	2.0	
4	0.0	26.9	72.0	1.1	

P( jar   chips )	A	В	С	D	credibility
0	1.9	22.6	24.5	50.9	51%
1	0.9	17.3	18.2	63.6	
2	73.7	25.3	0.0	1.1	
3	57.1	40.8	0.0	2.0	
4	0.0	26.9	72.0	1.1	

P( jar   chips )	A	В	С	D	credibility
0	1.9	22.6	24.5	50.9	75%
1	0.9	17.3	18.2	63.6	
2	73.7	25.3	0.0	1.1	
3	57.1	40.8	0.0	2.0	
4	0.0	26.9	72.0	1.1	

P( jar   chips )	A	В	С	D	credibility
0	1.9	22.6	24.5	50.9	75%
1	0.9	17.3	18.2	63.6	64%
2	73.7	25.3	0.0	1.1	
3	57.1	40.8	0.0	2.0	
4	0.0	26.9	72.0	1.1	
					,

P( jar   chips )	A	В	С	D	credibility
0	1.9	22.6	24.5	50.9	75%
1	0.9	17.3	18.2	63.6	82%
2	73.7	25.3	0.0	1.1	
3	57.1	40.8	0.0	2.0	
4	0.0	26.9	72.0	1.1	
					,

P( jar   chips )	A	В	С	D	credibility
0	1.9	22.6	24.5	50.9	75%
1	0.9	17.3	18.2	63.6	82%
2	73.7	25.3	0.0	1.1	74%
3	57.1	40.8	0.0	2.0	
4	0.0	26.9	72.0	1.1	

P( jar   chips )	Α	В	С	D	credibility
0	1.9	22.6	24.5	50.9	75%
1	0.9	17.3	18.2	63.6	82%
2	73.7	25.3	0.0	1.1	74%
3	57.1	40.8	0.0	2.0	57%
4	0.0	26.9	72.0	1.1	

P( jar   chips )	A	В	С	D	credibility
0	1.9	22.6	24.5	50.9	75%
1	0.9	17.3	18.2	63.6	82%
2	73.7	25.3	0.0	1.1	74%
3	57.1	40.8	0.0	2.0	98%
4	0.0	26.9	72.0	1.1	

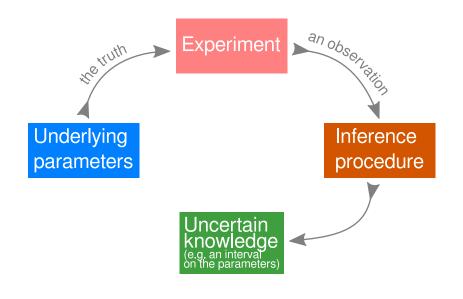
P( jar   chips )	A	В	С	D	credibility
0	1.9	22.6	24.5	50.9	75%
1	0.9	17.3	18.2	63.6	82%
2	73.7	25.3	0.0	1.1	74%
3	57.1	40.8	0.0	2.0	98%
4	0.0	26.9	72.0	1.1	72%

### Confidence & credible intervals together

confidence	Α	В	C	D	credibility
0	1	12	13	27	0%
1	1	19	20	70	99%
2	70	24	0	1	99%
3	28	20	0	1	41%
4	0	25	67	1	99%
coverage	70%	88%	87%	70%	

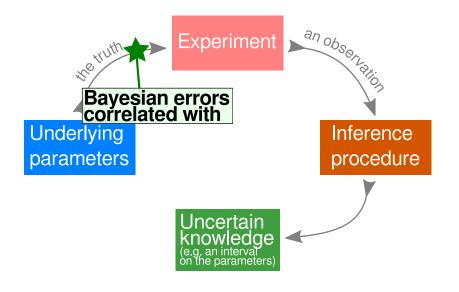
credible	A	В	C	D	credibility
0	1	12	13	27	75%
1	1	19	20	70	82%
2	70	24	0	1	74%
3	28	20	0	1	98%
4	0	25	67	1	72%
coverage	98%	20%	100%	97%	

### Correlation of error

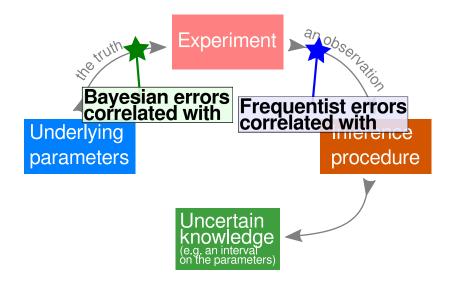


▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□ ● ● ●

### Correlation of error



### Correlation of error



▲口▶▲圖▶▲≣▶▲≣▶ = 悪 - 釣A@

### Criticism of frequentist style

### Essay

# Why Most Published Research Findings Are False

John P. A. Ioannidis

#### Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding factors that influence this problem and some corollaries thereof.

#### Modeling the Framework for False Positive Findings

### It can be proven that most claimed research findings are false.

yet ill-founded strategy of claiming conclusive research findings solely on is characteristic of the vary a lot depending o field targets highly like or searches for only on true relationships amo and millions of hypoth be postulated. Let us a for computational sim circumscribed fields wi is only one true relatio many that can be hypo

Why Most Published Research Findings Are False, Ioannidis JPA, PLOS MEDICINE Vol. 2, No. 8, e124 doi:10.1371/journal.pmed.0020124

### Criticism of Bayesian style

# ECONOMETRICA

VOLUME 47

November, 1979

NUMBER 6

### THE IMPOSSIBILITY OF BAYESIAN GROUP DECISION MAKING WITH SEPARATE AGGREGATION OF BELIEFS AND VALUES

### By AANUND HYLLAND AND RICHARD ZECKHAUSER<sup>1</sup>

Bayesian theory for rational individual decision making under uncertainty prescribes that the decision maker define independently a set of beliefs (probability assessments for the states of the world) and a system of values (utilities). The decision is then made by maximizing expected utility. We attempt to generalize the model to group decision making. It is assumed that the group's belief depends only on individual beliefs and the group's values only on individual values, that the belief aggregation procedure respects unanimity, and that the entire procedure guarantees Pareto optimality. We prove that only trivial (dictatorial) aggregation procedures for beliefs are possible.

#### 1. INTRODUCTION

MANY DECISIONS MADE under uncertainty, indeed many important ones, are made by a group, be it a collection of friends, the Congress of the United States, or

### Disagreement in the real world

- ► Avandia: world's #1 diabetes drug, approved in 1999.
- Sold by GlaxoSmithKline PLC
- Sales: \$3 billion in 2006
- ▶ In 2004, GSK releases results of many small studies.

### GSK releases 42 small studies

Study	Avandia heart attacks	Control heart attacks
49632-020	2/391	1/207
49653-211	5/110	2/114
DREAM	15/2635	9/2634
49653-134	0/561	2/276
49653-331	0/706	0/325
:	:	•
•	·	·

### In 2007, Dr. Nissen crashes the party



ESTABLISHED IN 1812

JUNE 14, 2007

VOL. 356 NO. 24

Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes

Steven E. Nissen, M.D., and Kathy Wolski, M.P.H.

#### ABSTRACT

#### BACKGROUND

Rosiglitazone is widely used to treat patients with type 2 diabetes mellitus, but its effect on cardiovascular morbidity and mortality has not been determined.

#### METHODS

We conducted searches of the published literature, the Web site of the Food and Drug Administration, and a clinical-trials registry maintained by the drug manufacturer (GlaxoSmithKline). Criteria for inclusion in our mera-analysis included a study duration of more than 24 weeks, the use of a randomized control group nor receiving rosiglitazone, and the availability of ourcome data for myocardial infarction and death from cardiovascular causes. Of 116 potentially relevant studies, 42 trials met the inclusion criteria. We tabulated all occurrences of myocardial infarction and death from cardiovascular causes.

#### RESULTS

Data were combined by means of a fixed-effects model. In the 42 trials, the mean age of the subjects was approximately 56 years, and the mean baseline glycared hemoglobin level was approximately 8.2%. In the rosiglitazone group, as compared with the control group, the odds ratio for myocardial infarction was 1.43 (05% confidence interval [CI], 1.03 to 1.98; P=0.05), and the odds ratio for death from cardiovascular causes was 1.64 (95% CI, 0.98 to 2.74; P=0.06).

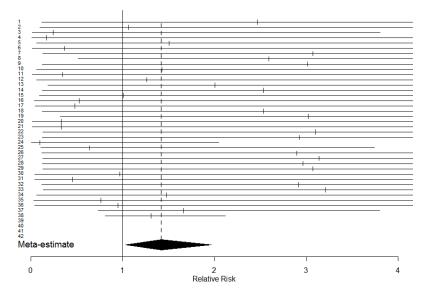
From the Cleveland Clinic, Cleveland. Address reprint requests to Dr. Nissen at the Department of Cardiovascular Medicine, Cleveland Clinic, 9500 Euclid Ave., Cleveland, OH 44195, or at nissens@ccf. org.

This article (10.1056/NEJMoa072761) was published at www.nejm.org on May 21, 2007.

N Engl J Med 2007;356:2457-71. Copyright © 2007 Massachusetts Medical Society.



### Frequentist inference



....

# THE WALL STREET JOURNAL

DOWJONES

TUESDAY, MAY 22, 2007 - VOL. CCXLIX NO. 119

DJA 13542.88 ¥13.65 -0.1% NASDAO 2578.79 & 0.8% NIKKEI 17556.87 & 0.9% DJ STOXX 50 3905.70 ¥ 0.3% 10-YR TREAS & 4/32, vield 4.79% OL 566.27 & 51.33 GOLD 5662.90 & 51.90 EURO 51.3470 YEN 121.45

. Lowe's posted a 12% profit

What's News-

. Iraq's military is drawing up

60 the same are as Mr. Voerein.

### Sequel for Vioxx Critic: Attack on Diabetes Pill

Glaxo Shares Plunge As Dr. Nissen Sees Risk To Heart From Avandia

#### By ANNA WILDE MATHEWS

MEDICAL DETECTIVE

An analysis linking the widely used diabetes drug Avandia to higher risk of heart attacks represents a serious blow to GlaxoSmithkline PLC and underscores how outside critics have been empowered to challenge big-selling drugs after the outcry over the withdrawn painkiller Vioxx Glaxo rang up



more than \$3 billion in world-wide sales of Avandia last year. Its share price fell more than 7% after the New England Journal of Medicine released the analysis by prominent cardiologist Steven Nissen of the Cleveland Clinic, who helped

raise early safety concerns about Vioxx. The analysis suggested that peonle on Avandia have a 43% higher chance of suffering a heart attack. Glaxo said it "strongly disagrees"

with his conclusions which come from

#### Drug in Demand Sales of GlavoSmithKline's Avandia. in billions of pounds:



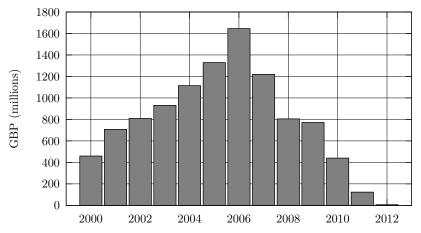
Note: E1 = \$1.97 at the current rate; includes sales of Source: the company

and Drug Administration should have acted faster to alert the public about possible risk from Avandia. Glaxo performed its own meta-analysis, which also showed a potential danger. It shared an early version of it with the FDA in September 2005 and a more complete one in August 2006. The findings weren't reflected on the U.S. label, which is supposed to give a comprehensive review of the drug's risks. Robert Meyer, head of the FDA office

that oversees diabetes drugs, said the agency is still working on its analysis. "We have other data that suggests we

<sup>\*\*\*\* \$1.00</sup> 

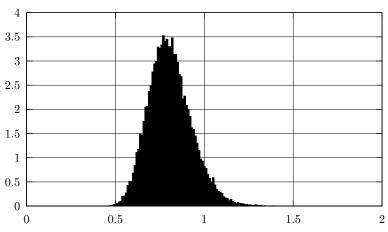
### GlaxoSmithKline loses \$12 billion



Avandia worldwide sales

▲□▶ ▲□▶ ▲臣▶ ★臣▶ 三臣 - のへで

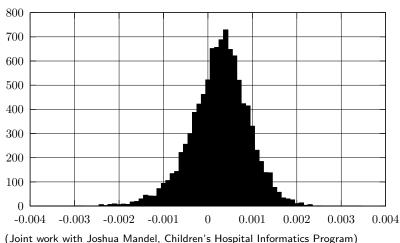
### Bayesian inference disagrees, for risk ratio



P.D.F. on Avandia's risk ratio for heart attack

(Joint work with Joshua Mandel, Children's Hospital Informatics Program)

### Or does it? Here, risk difference model



P.D.F. on Avandia's risk difference for heart attack

◆□ → ◆□ → ◆ 三 → ◆ 三 → のへぐ

### The TAXUS ATLAS Experiment

- Boston Scientific proposed to show that new heart stent was not "inferior" to old heart stent, with 95% confidence.
- Inferior means three percentage points more "bad" events.
  - ▶ Control 7% vs. Treatment  $10.5\% \Rightarrow inferior$
  - ▶ CONTROL 7% vs. TREATMENT  $9.5\% \Rightarrow$  non-inferior.

### ATLAS Results (May 2006)

May 16, 2006 — NATICK, Mass. and PARIS, May 16 /PRNewswire-FirstCall/ — Boston Scientific Corporation today announced nine-month data from its TAXUS ATLAS clinical trial. [...] **The trial met its primary endpoint** of nine-month target vessel revascularization (TVR), a measure of the effectiveness of a coronary stent in reducing the need for a repeat procedure.

## ATLAS Results (April 2007)

Turco et al., *Polymer-Based, Paclitaxel-Eluting TAXUS Liberté Stent in De Novo Lesions*, Journal of the American College of Cardiology, Vol. 49, No. 16, 2007.

**Results**: The primary non-inferiority end point was met with the 1-sided 95% confidence bound of 2.98% less than the pre-specified non-inferiority margin of 3% (p = 0.0487).

**Statistical methodology**. Student t test was used to compare independent continuous variables, while chi-square or Fisher exact test was used to compare proportions.

### $p < 0.05 \rightarrow 95\%$ confidence interval excludes inferiority

### The problem

	Event	No event	Total
Control	67	889	956
Treatment	68	787	855
Total	135	1676	1811

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

## The problem

	Event	No event	Total
Control	67	889	956
Treatment	68	787	855
Total	135	1676	1811

- ▶ With uniform prior on rates, Pr(inferior | data) ≈ 0.050737979...
- Posterior probability of non-inferiority is less than 95%.

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

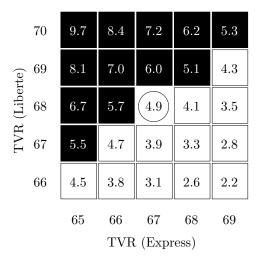
 Confidence interval: approximate each binomial separately with a normal distribution. Known as Wald interval.

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

▶ 
$$p = \int_{0.03}^{\infty} \mathcal{N}\left(\frac{i}{m} - \frac{j}{n}, \frac{i(m-i)}{m^3} + \frac{j(n-j)}{n^3}\right) \approx 0.0487395...$$

•  $p < 0.05 \rightarrow$  success

### The ultimate close call



Wald's area ( $\approx p$ ) with (m, n) = (855, 956)

◆□> ◆□> ◆三> ◆三> 三三 のへで

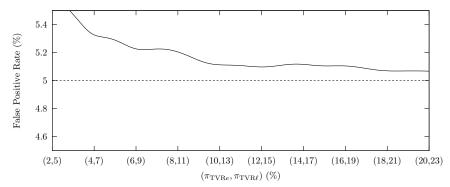
confidence	Α	В	C	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
coverage	70%	88%	87%	70%
			•	

▲□▶ ▲□▶ ▲三▶ ▲三▶ ▲□ ● ● ●

confidence	A	В	C	D
0	1	12	13	27
1	1	19	20	70
2	70	24	0	1
3	28	20	0	1
4	0	25	67	1
coverage	70%	88%	87%	70%
false positive rate	30%	12%	13%	30%

◆□ ▶ ◆■ ▶ ◆ ■ ◆ ● ◆ ● ◆ ● ◆

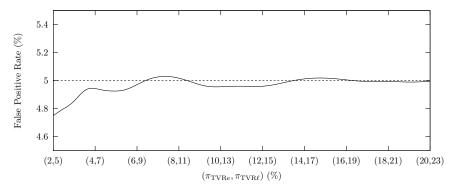
### The Wald interval undercovers



False Positive Rate of ATLAS non-inferiority test along critical line

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 三臣 - のへで

### Better approximation: score interval



False Positive Rate of maximum-likelihood z-test along critical line

▲ロト ▲御 ▶ ▲ 臣 ▶ ▲ 臣 ▶ ● 臣 ■ • の Q (2)

# Other methods all yield failure

Method	<i>p</i> -value or confidence bound	Result
Wald interval	p = 0.04874	Pass
z-test, constrained max likelihood standard error	p = 0.05151	Fail

◆□ ▶ <圖 ▶ < E ▶ < E ▶ E • 9 < 0</p>

# Other methods all yield failure

Method	<i>p</i> -value or confidence bound	Result
Wald interval	p = 0.04874	Pass
z-test, constrained max likelihood standard error	p = 0.05151	Fail
z-test with Yates continuity correction	c = 0.03095	Fail
Agresti-Caffo I <sub>4</sub> interval	p = 0.05021	Fail
Wilson score	c = 0.03015	Fail
Wilson score with continuity correction	c = 0.03094	Fail
Farrington & Manning score	p = 0.05151	Fail
Miettinen & Nurminen score	p = 0.05156	Fail
Gart & Nam score	p = 0.05096	Fail
NCSS's bootstrap method	c = 0.03006	Fail
NCSS's quasi-exact Chen	c = 0.03016	Fail
NCSS's exact double-binomial test	p = 0.05470	Fail
StatXact's approximate unconditional test of non-inferiority	p = 0.05151	Fail
StatXact's exact unconditional test of non-inferiority	p = 0.05138	Fail
StatXact's exact CI based on difference of observed rates	c = 0.03737	Fail
StatXact's approximate CI from inverted 2-sided test	c = 0.03019	Fail
StatXact's exact CI from inverted 2-sided test	c = 0.03032	Fail

## Nerdiest chart contender?

#### **Degree of Certainty**

Medical studies define success or failure in testing a hypothesis by calculating a degree of certainty, known as the p-value. The p-value must be less than 5% for the results to be considered significant. Boston Scientfic's study, which used a statistical method called a Wald Interval, produced a p-value below 5%. But using 16 other methods turned up a p-value greater than 5%. Here are some of the p-values that resulted from the data in the study, using those different methodologies.

EQUATION	PASS ┥	► FAIL
Wald Interval	4.874%	
The Score z-test		5.151%
Agresti-Caffo interval test		5.021
Farrington & Manning score test		5.151
Miettinen & Nurminen score test		5.156
Gart & Nam score		5.096
NCSS LLC's exact double-binomial test		5.470
Cytel Inc.'s StatXact's approximate test		5.151
Cytel Inc.'s StatXact's exact test		5.138

◆□> ◆圖> ◆医> ◆医> 一臣

Source: WSJ research

## Boston Scientific Stent Study Flawed

#### BY KEITH J. WINSTEIN

A tired by Boston Scientured by Boston Scientific Corp. and expecting approval for U.S. sales is backed by flawed research despite the company's claims of success in a clinical trial, according to a Wall Street Journal review of the data.

Boston Scientific submitted the results of the 2006 trial to the Food and Drug Administration to gain U.S. approval for the Tawus Liberte, which already is one of the top-selling stents: abroad. Coronary stents--tury is seficidis tata propogena retries: clogged by heart disease-are one of the most popular methods for treating heart patients, and have been implanted in more than 15 million people world-wide.

But Boston Scientific's claim was based on aflawed statistical equation that favored the Liberts stent; a Journal analysis has found. Using a number of other methods of calculation—including 14 available in off-the-shelf software programs—the Liberte study would have been a failure by the common standards of statistical significance in research.

Boston Scientific isn't the only company to use the equation, known as a Wald interval, which has long been criticized



Boston Scientific is seeking FDA approval for its Taxus Liberte stent.

by statisticians for exaggerating the certainty of research results. Rivals Medtronic Inc. and Abbott Laboratories have used the same equation in stent studies.

But in those cases, any boost provided by the Wald equation wouldn't have changed the outcome of the study. In the Liberte study, the equation's shortcomings meant the difference between success and failure in the study's main goal.

The difference also sheds light on the leeway that device makers have when designing studies for the FDA. Studies designed to satisfy the requirements of the FDA's medical-device branch can be less rigorous than those aimed at winning U.S. approval for drugs. That is partly because of a 1997 federal law aimed at lessening the regulatory requirements on device makers.

The FDA declined to specifically discuss its deliberations of the Liberte, which is still under review by the agency.

Boston Scientific doesn't agree that it made a mistake or that the study failed to reach statistical significance. "We used standard methodology that we discussed with the FDA up front, and then executed," said Dong Viat Baim, Boston Scientific's chief scientific and medical officer. *Hease turn to page B6* 

・ロト ・ 雪 ト ・ ヨ ト

## The statistician says...

- "np > 5, therefore, the Central Limit Theorem applies and a Gaussian approximation is appropriate."
- "We had even more data points than we powered the study for, so there was adequate safety margin."

"'Exact' tests are too conservative."

"Other statistical applications often rely on large-scale assumptions for inferences, risking incorrect conclusions from data sets not normally distributed. StatXact utilizes Cytel's own powerful algorithms to make exact inferences..."

"Other statistical applications often rely on large-scale assumptions for inferences, risking incorrect conclusions from data sets not normally distributed. StatXact utilizes Cytel's own powerful algorithms to make exact inferences..."

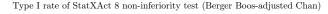
# Graphing the coverage

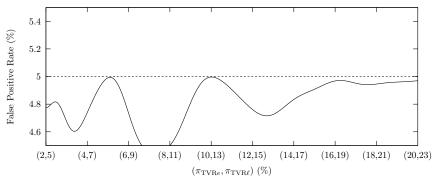
- Problem: hard to calculate a million "exact" p-values
- (StatXact: about 10 minutes each)
- Contribution: method for calculating whole tableau
- Calculates all p-values in time for "hardest" square
- **Trick**: calculate in the right order, cache partial results

"Other statistical applications often rely on large-scale assumptions for inferences, risking incorrect conclusions from data sets not normally distributed. StatXact utilizes Cytel's own powerful algorithms to make exact inferences..."

◆□▶ ◆□▶ ◆ □▶ ◆ □▶ ○ □ ○ ○ ○ ○

"Other statistical applications often rely on large-scale assumptions for inferences, risking incorrect conclusions from data sets not normally distributed. StatXact utilizes Cytel's own powerful algorithms to make exact inferences..."

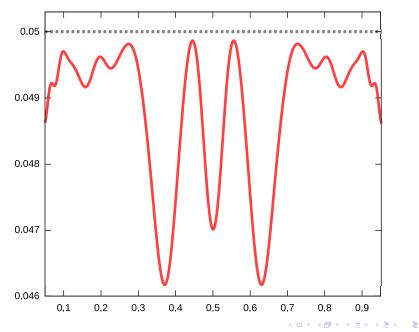




▲□▶ ▲□▶ ▲□▶ ▲□▶ □ ののの

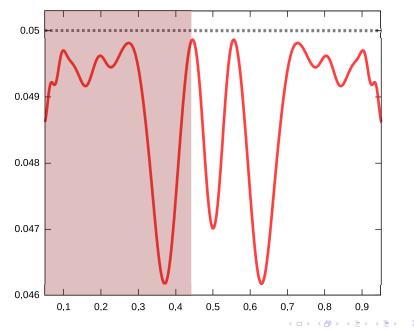
# Animation

Barnard's test for  $N = 256 \times 257$ 



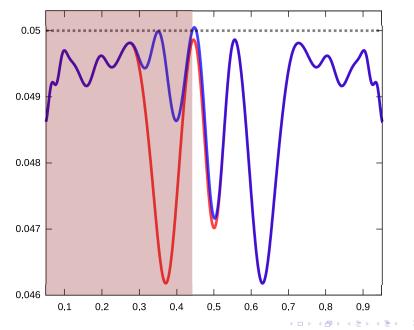
996

# Assume a prior hypothesis...



SAR

# Frequentists can benefit from priors too!



Sac

# Final thoughts

- Bayesian and frequentist schools have much in common.
- If stark disagreement between Bayesian and frequentist methods, probably sign of bigger problems!
- What's important: say what we're trying to infer, how we get there, what we care about.