A Foundational Introduction to Bayesian Statistics

Joseph F. Lucke

Research Institute on Addictions
State University of New York at Buffalo

Quantitative Methods Forum
Psychology Department
York University
November 17, 2014

Schools of Statistics

Two Major Schools

- Frequentist:— significance, power, p-values, hypothesis testing
- Bayesian: — subjective probability, prior & posterior distributions; inference, decision theory;

Schools of Statistics

Two Major Schools

- Frequentist:— significance, power, p-values, hypothesis testing
- Bayesian: — subjective probability, prior & posterior distributions; inference, decision theory;

Additional Minor Schools

- Neyman-Pearson: Neyman, Lehmann — significance, power, rejection region, decision between hypotheses.
- Likelihood: Edwards, Royall — likelihood, support, Bayes factor.
- Fisher: Fisher, Mayo — rejection of hypotheses, probabilistic falsification;
- Fiducial: Fisher — Posterior probabilities without priors.

Interpretations of Probability

Probability has had multiple interpretations:

- Mathematical — Measure theory, not interpreted;

Interpretations of Probability

Probability has had multiple interpretations:

- Mathematical — Measure theory, not interpreted;
- Frequency — Relative frequencies of events, infinite sequences;

Interpretations of Probability

Probability has had multiple interpretations:

- Mathematical — Measure theory, not interpreted;
- Frequency — Relative frequencies of events, infinite sequences;
- Propensity — Causal, single case, relative frequency;

Interpretations of Probability

Probability has had multiple interpretations:

- Mathematical — Measure theory, not interpreted;
- Frequency — Relative frequencies of events, infinite sequences;
- Propensity — Causal, single case, relative frequency;
- Logical — Partial entailment among propositions;

Interpretations of Probability

Probability has had multiple interpretations:

- Mathematical — Measure theory, not interpreted;
- Frequency — Relative frequencies of events, infinite sequences;
- Propensity — Causal, single case, relative frequency;
- Logical — Partial entailment among propositions;
- Classical — Equally possible alternatives;

Interpretations of Probability

Probability has had multiple interpretations:

- Mathematical — Measure theory, not interpreted;
- Frequency — Relative frequencies of events, infinite sequences;
- Propensity — Causal, single case, relative frequency;
- Logical — Partial entailment among propositions;
- Classical — Equally possible alternatives;
- Subjective — Logic of uncertain beliefs, opinions, judgements;

Interpretations of Probability

Probability has had multiple interpretations:

- Mathematical — Measure theory, not interpreted;
- Frequency — Relative frequencies of events, infinite sequences;
- Subjective — Logic of uncertain beliefs, opinions, judgements;

Why Consider Bayesian Statistical Theory?

Pragmatic Reasons

- Solve more statistical problems
- Implement more realistic models
- Less concern with sample size issues
- Solve technical problems, e.g., negative variances, ill-conditioning, non-standard distributions, complex estimation
- Supplement the likelihood function with additional constraints

Why Consider Bayesian Statistical Theory?

Pragmatic Reasons

- Solve more statistical problems
- Implement more realistic models
- Less concern with sample size issues
- Solve technical problems, e.g., negative variances, ill-conditioning, non-standard distributions, complex estimation
- Supplement the likelihood function with additional constraints

Radical Reasons

- Satisfactory interpretation of probability axioms.
- Unified approach to probability and statistics.
- Incorporate prior information.
- Conceptual difficulties with foundations of frequentist approach.

Frequency Theory of Probability

The probability of an attribute ω_i in a reference set $\{\omega_1, \ldots, \omega_K\}$ is p.

means exactly (no more and no less) that

The limit of the relative frequency of occurrences of ω_i that would be obtained were the reference set $\{\omega_1, \ldots, \omega_K\}$ realized infinitely often is p, i.e.,

$$\lim_{N \to \infty} \frac{\#(\omega_i)}{N} = p.$$

or equivalently,

$$\forall \epsilon > 0 \, \exists N \; \forall n > N \left| \frac{\#(\omega_i)}{n} - p \right| < \epsilon.$$

Advantages of Frequency Definition

$$\lim_{N \to \infty} \frac{\#(\omega_i)}{N} = p.$$

1. Defined: Limit of a sequence.
2. Empirical: Based on observations.
3. Operational: Procedure to define a specific probability.
4. Objective: Everyone can agree on the probability of an event.
5. Mathematical: Satisfies the (Kolmogorov) axioms of probability.

What's Wrong with the Frequency Theory?

$$\lim_{N \to \infty} \frac{\#(\omega_i)}{N} = p.$$

1. Not defined: Convergence of a physical process is not well-defined.

What's Wrong with the Frequency Theory?

$$\lim_{N \to \infty} \frac{\#(\omega_i)}{N} = p.$$

1. Not defined: Convergence of a physical process is not well-defined.
2. Not empirical, but hypothetical: Not based on finite set observations, but on unobserved limit.

What's Wrong with the Frequency Theory?

$$\lim_{N \to \infty} \frac{\#(\omega_i)}{N} = p.$$

1. Not defined: Convergence of a physical process is not well-defined.
2. Not empirical, but hypothetical: Not based on finite set observations, but on unobserved limit.
3. Counterfactual: Based on what would happen in an infinite sequence.

What's Wrong with the Frequency Theory?

$$\lim_{N\to\infty} \frac{\#(\omega_i)}{N} = p.$$

1. Not defined: Convergence of a physical process is not well-defined.
2. Not empirical, but hypothetical: Not based on finite set observations, but on unobserved limit.
3. Counterfactual: Based on what would happen in an infinite sequence.
4. Not operational: No finite sequence yields any information regarding the hypothetical limit.

What's Wrong with the Frequency Theory?

$$\lim_{N\to\infty} \frac{\#(\omega_i)}{N} = p.$$

1. Not defined: Convergence of a physical process is not well-defined.
2. Not empirical, but hypothetical: Not based on finite set observations, but on unobserved limit.
3. Counterfactual: Based on what would happen in an infinite sequence.
4. Not operational: No finite sequence yields any information regarding the hypothetical limit.
5. Not satisfactory: Does not satisfy axioms: not countably additive, does not form a σ-algebra.

Fixing the Frequency Theory

Additional assumptions are required to address the defects in the Frequency Theory.

- Note that one cannot use the LLNs to fix defects. (Probability not yet defined.)
- Postulate of (non-mathematical) Convergence: The sequence converges and does so rapidly.
- Postulate of (non-probabilistic) Randomness: Any (recursively computable) subsequence of the sequence converges to the same limit.

Frequency theory requires additional, non-testable, subjective assumptions.

Comments on “Objective” Probability

- Ramsey — “There are no such things as objective chances . . . Chances must be defined by degrees of belief.” (1931)

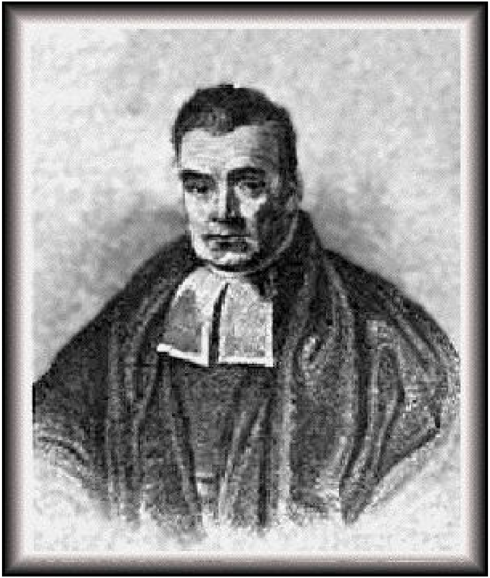
Comments on “Objective” Probability

- Ramsey — “There are no such things as objective chances . . . Chances must be defined by degrees of belief.” (1931)
- de Finetti — “[Objective] probability does not exist!” (1972)

Comments on "Objective" Probability

- Ramsey — "There are no such things as objective chances . . . Chances must be defined by degrees of belief." (1931)
- de Finetti — "[Objective] probability does not exist!" (1972)
- Laplace — Probability is "only the expression of our ignorance of the true causes." (1814)

Thomas Bayes



First Look at Bayesian Analysis

- Bayesian Statistical Theory (BST) is radically different from frequentist (Neyman-Pearson + Fisher) theory statistics
- BST is distinguished by the fact it uses subjective probability and Bayes's Theorem for inference.
- BST is not just another class of statistical models like structural equation models or multilevel models.
- BST can in principle analyze any statistical model.
- Even though the obtained numbers may be the same as in frequentist theory, the interpretation will be different.
- Inferential reasoning is more natural in BST than in frequentist.

Comparison of Bayesian and frequentist Theories

Feature	Bayesian	frequentist
Content	Beliefs	Decisions
Unifying Principle	Coherence	Inductive behavior
Probability	Subjective	Objective
Repeated Events	Exchangeability	Independence
Data	Fixed	Random
Parameters	Random	Fixed, unknown
Inference	Bayes's Theorem	Unbiased, MLE, MSE, etc.
Confidence interval	Fixed	Random
Hypothesis testing	Posterior	NHST,Significance, power

Three Pillars of Bayesian Statistical Theory

There are three theorems that form the foundations of BST.

1. Coherence: Logic of subjective probability;

Three Pillars of Bayesian Statistical Theory

There are three theorems that form the foundations of BST.

1. Coherence: Logic of subjective probability;
2. Exchangeability: Repeated events (or measurements);

Three Pillars of Bayesian Statistical Theory

There are three theorems that form the foundations of BST.

1. Coherence: Logic of subjective probability;
2. Exchangeability: Repeated events (or measurements);
3. Bayes's Theorem: Inference;

F. P. Ramsey

Subjective Probability: Coherence

- Probability is the logic of uncertain beliefs, judgements, or opinions.
- Your opinion can be represented as a set of subjectively fair bets on an event.
- Events may be unique. No repetition is required.
- Coherence principle: Avoid sets of bets that entrain a guaranteed loss. A form of pragmatic consistency.

Consistency and Coherence

- Deductive logic is the logic of certainty $\leftrightarrow$ Probability is the logic of uncertainty.

Consistency and Coherence

- Deductive logic is the logic of certainty $\leftrightarrow$ Probability is the logic of uncertainty.
- Deductive logic preserves consistency $\leftrightarrow$ Probability preserves coherence.

Consistency and Coherence

- Deductive logic is the logic of certainty $\leftrightarrow$ Probability is the logic of uncertainty.
- Deductive logic preserves consistency $\leftrightarrow$ Probability preserves coherence.
- Deductive logic is content-free. $\leftrightarrow$ Probability is content-free. (Very important!)

Consistency and Coherence

- Deductive logic is the logic of certainty $\leftrightarrow$ Probability is the logic of uncertainty.
- Deductive logic preserves consistency $\leftrightarrow$ Probability preserves coherence.
- Deductive logic is content-free. $\leftrightarrow$ Probability is content-free. (Very important!)
- Deductive logic does not establish truth but just transmits it $\leftrightarrow$ Probability does not establish uncertainty but just transmits it.

Consistency and Coherence

- Deductive logic is the logic of certainty $\leftrightarrow$ Probability is the logic of uncertainty.
- Deductive logic preserves consistency $\leftrightarrow$ Probability preserves coherence.
- Deductive logic is content-free. $\leftrightarrow$ Probability is content-free. (Very important!)
- Deductive logic does not establish truth but just transmits it $\leftrightarrow$ Probability does not establish uncertainty but just transmits it.

Subjective probability is as objective as deductive logic.

Example of Incoherence

Suppose you give judgements regarding a manned mission to Mars.

Events	$(2015 : 2020]$	$(2020 : 2040]$	$(2040 : 2060]$

Example of Incoherence

Suppose you give judgements regarding a manned mission to Mars.

Events	$(2015 : 2020]$	$(2020 : 2040]$	$(2040 : 2060]$
Odds Against	$8 : 1$	$2 : 1$	$1 : 1$

Example of Incoherence

Suppose you give judgements regarding a manned mission to Mars.

Events	$(2015 : 2020]$	$(2020 : 2040]$	$(2040 : 2060]$	
Odds Against	$8 : 1$	$2 : 1$	$1 : 1$	
Bets ($)	2	6	9	Gain

Example of Incoherence

Suppose you give judgements regarding a manned mission to Mars.

Events	$(2015:2020]$	$(2020:2040]$	$(2040:2060]$	
Odds Against	$8:1$	$2:1$	$1:1$	
Bets ($)	2	6	9	Gain
$(2015:2020]$	-16	$+6$	$+9$	-1

Example of Incoherence

Suppose you give judgements regarding a manned mission to Mars.

Events	$(2015:2020]$	$(2020:2040]$	$(2040:2060]$	
Odds Against	$8:1$	$2:1$	$1:1$	
Bets (\$)	2	6	9	Gain
$(2015:2020]$	-16	$+6$	$+9$	-1
$(2020:2040]$	$+2$	-12	$+9$	-1

Example of Incoherence

Suppose you give judgements regarding a manned mission to Mars.

Events	$(2015 : 2020]$	$(2020 : 2040]$	$(2040 : 2060]$	
Odds Against	$8 : 1$	$2 : 1$	$1 : 1$	
Bets ($)	2	6	9	Gain
$(2015 : 2020]$	-16	$+6$	$+9$	-1
$(2020 : 2040]$	$+2$	-12	$+9$	-1
$(2040 : 2060]$	$+2$	$+6$	-9	-1

Example of Incoherence

Suppose you give judgements regarding a manned mission to Mars.

Events	$(2015:2020]$	$(2020:2040]$	$(2040:2060]$	
Odds Against	$8:1$	$2:1$	$1:1$	
Bets ($)	2	6	9	Gain
$(2015:2020]$	-16	$+6$	$+9$	-1
$(2020:2040]$	$+2$	-12	$+9$	-1
$(2040:2060]$	$+2$	$+6$	-9	-1

- No matter what the outcome, I lose $\$1$. My judgements are irrational (incoherent).

Example of Incoherence

Suppose you give judgements regarding a manned mission to Mars.

Events	$(2015:2020]$	$(2020:2040]$	$(2040:2060]$	
Odds Against	$8:1$	$2:1$	$1:1$	
Bets ($)	2	6	9	Gain
$(2015:2020]$	-16	$+6$	$+9$	-1
$(2020:2040]$	$+2$	-12	$+9$	-1
$(2040:2060]$	$+2$	$+6$	-9	-1

- No matter what the outcome, I lose $\$1$. My judgements are irrational (incoherent).
- Corresponding probabilities: $\frac{1}{2}+\frac{1}{3}+\frac{1}{9}=\frac{17}{18}<1$.

Example of Coherence

Suppose instead you change your bets to:

Events	$(2015 : 2020]$	$(2020 : 2040]$	$(2040 : 2060]$

Example of Coherence

Suppose instead you change your bets to:

Events	$(2015 : 2020]$	$(2020 : 2040]$	$(2040 : 2060]$
Odds Against	$5 : 1$	$2 : 1$	$1 : 1$

Example of Coherence

Suppose instead you change your bets to:

Events	$(2015:2020]$	$(2020:2040]$	$(2040:2060]$	
Odds Against	$5:1$	$2:1$	$1:1$	
Bets ($)	2	6	9	Gain

Example of Coherence

Suppose instead you change your bets to:

Events	$(2015:2020]$	$(2020:2040]$	$(2040:2060]$	
Odds Against	$5:1$	$2:1$	$1:1$	
Bets ($)	2	6	9	Gain
$(2015:2020]$	-10	$+6$	$+9$	$+5$

Example of Coherence

Suppose instead you change your bets to:

Events	$(2015 : 2020]$	$(2020 : 2040]$	$(2040 : 2060]$	
Odds Against	$5 : 1$	$2 : 1$	$1 : 1$	
Bets ($)	2	6	9	Gain
$(2015 : 2020]$	-10	$+6$	$+9$	$+5$
$(2020 : 2040]$	$+2$	-12	$+9$	-1

Example of Coherence

Suppose instead you change your bets to:

Events	$(2015:2020]$	$(2020:2040]$	$(2040:2060]$	
Odds Against	$5:1$	$2:1$	$1:1$	
Bets ($)	2	6	9	Gain
$(2015:2020]$	-10	$+6$	$+9$	$+5$
$(2020:2040]$	$+2$	-12	$+9$	-1
$(2040:2060]$	$+2$	$+6$	-9	-1

Example of Coherence

Suppose instead you change your bets to:

Events	$(2015:2020]$	$(2020:2040]$	$(2040:2060]$	
Odds Against	$5:1$	$2:1$	$1:1$	
Bets ($)	2	6	9	Gain
$(2015:2020]$	-10	$+6$	$+9$	$+5$
$(2020:2040]$	$+2$	-12	$+9$	-1
$(2040:2060]$	$+2$	$+6$	-9	-1

- No guaranteed loss or win

Example of Coherence

Suppose instead you change your bets to:

Events	$(2015:2020]$	$(2020:2040]$	$(2040:2060]$	
Odds Against	$5:1$	$2:1$	$1:1$	
Bets ($)	2	6	9	Gain
$(2015:2020]$	-10	$+6$	$+9$	$+5$
$(2020:2040]$	$+2$	-12	$+9$	-1
$(2040:2060]$	$+2$	$+6$	-9	-1

- No guaranteed loss or win
- My judgements are coherent.

Example of Coherence

Suppose instead you change your bets to:

Events	$(2015:2020]$	$(2020:2040]$	$(2040:2060]$	
Odds Against	$5:1$	$2:1$	$1:1$	
Bets ($)	2	6	9	Gain
$(2015:2020]$	-10	$+6$	$+9$	$+5$
$(2020:2040]$	$+2$	-12	$+9$	-1
$(2040:2060]$	$+2$	$+6$	-9	-1

- No guaranteed loss or win
- My judgements are coherent.
- Corresponding probabilities: $\frac{1}{2}+\frac{1}{3}+\frac{1}{6}=1$.

Ramsey-de Finetti (1931) Theorem

A set of uncertain judgements is coherent if and only if their probabilistic representation satisfy the axioms of probability.

Ramsey-de Finetti (1931) Theorem

A set of uncertain judgements is coherent if and only if their probabilistic representation satisfy the axioms of probability.

- It is crucial to understand that coherence applies strictly to a set of judgements.

Ramsey-de Finetti (1931) Theorem

A set of uncertain judgements is coherent if and only if their probabilistic representation satisfy the axioms of probability.

- It is crucial to understand that coherence applies strictly to a set of judgements.
- Coherence merely insures that vindication of an action based on such judgements cannot be sabotaged in advance.

Ramsey-de Finetti (1931) Theorem

A set of uncertain judgements is coherent if and only if their probabilistic representation satisfy the axioms of probability.

- It is crucial to understand that coherence applies strictly to a set of judgements.
- Coherence merely insures that vindication of an action based on such judgements cannot be sabotaged in advance.
- My judgements (e.g., There are living pterodactyls in Papua New Guinea.) need not be based on reality.

Bruno de Finetti

Exchangeability for finite case

- The concept of exchangeability is the subjectivist's equivalent to random sampling.
- Given a set of N events.
- Let $\pi : \{1, \ldots, N\} \mapsto \{1, \ldots, N\}$ be a permutation function.
- The set is exchangeable if any sample of size $n \leq N$ is judged to have the same distribution as any other sample of size n.
- $\mathrm{pr}\,(x_1) = \mathrm{pr}\left(x_{\pi(1)}\right)$
- $\mathrm{pr}\,(x_1, x_2) = \mathrm{pr}\left(x_{\pi(1)}, x_{\pi(2)}\right)$
- $\ldots$
- $\mathrm{pr}\,(x_1, x_2, \ldots, x_n) = \mathrm{pr}\left(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}\right)$

Exchangeability for infinite case

- An infinite set of events is infinitely exchangeable if any arbitrarily large finite sample of those events is exchangeable.
- DeFinetti's Representation Theorem (1937): If an infinite set of events is infinitely exchangeable, then the events can be modeled as if they were independent and identically distributed events conditional upon some "unknown" parameter.
- If $\{x_1, \ldots, x_n, \ldots\}$ are infinitely exchangeable, then

$$\mathrm{pr}\{x_1, \ldots, x_n\} = \int \mathrm{pr}(x_1 \mid \theta), \ldots, \mathrm{pr}(x_n \mid \theta)\ \ \mathrm{pr}(\theta) d\theta.$$

- Exchangeable events are a mixture of conditionally independent events.

Exchangeability Explicates Relative Frequency

If $\{x_1, \ldots, x_n, \ldots\}$ are infinitely exchangeable Bernoulli random quantities, then

$$\mathrm{pr}(x_1, \ldots, x_n) = \int_0^1 \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}\, \mathrm{pr}(\theta) d\theta,$$

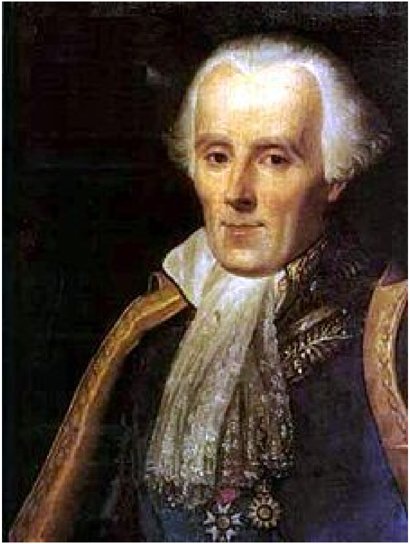
where $\theta = \lim_{n \to \infty} \sum x_i / n$. and $\mathrm{pr}(\theta)$ is the density of the "unknown" parameter θ.

1. The randomness and convergence of relative frequencies is a mathematical result from our judgement of exchangeability regarding the random variables!
2. Independent and identically distributed (IID) random variables conditional an uncertain parameter.

Extensions of Exchangeability

1. The indexing of events underlying exchangeability can be subtle and complicated.
2. Multidimensional indices.
3. Partial exchangeability: Exchangeability with respect to covariate indices.
4. Markov exchangeability: Exchangeability over adjacent pairs of time indices.
5. Multilevel exchangeability: Exchangeability within hierarchy of level indices.

Pierre-Simon Laplace



Bayes's Theorem

1. Let θ be a parameter of a model.
2. Let x be the obtained observation.
3. Let $\mathrm{pr}(\theta)$ be the prior probability (density) of θ.
4. Let $\mathrm{pr}(x|\theta)$ be the likelihood of observing x given θ.
5. Then the posterior distribution of the parameter θ given the data x is:
$$\mathrm{pr}(\theta|x) = \frac{\mathrm{pr}(x|\theta)\,\mathrm{pr}(\theta)}{\int \mathrm{pr}(x|\theta)\,\mathrm{pr}(\theta)\,d\theta}.$$
6. Interpretation: What your uncertainty regarding θ should be were you to observe x.
7. Note that frequentist inference only uses $\mathrm{pr}(x|\theta)$.

Posterior Analyses

The posterior distribution contains all the information regarding the impact of the observed data on the model parameters. Any characteristic of the distribution can be examined. Analyses ands summaries of the posterior convey the results of the analyses.

- Location—mean, median, model
- Spread—variance, quantiles
- Transformations of parameters
- Credible intervals—the probability that the parameter falls within a fixed interval.
- Hypothesis tests—the probability that a model fits the data.

Bayesian Analysis of PTCA vs Stent for MI

RCT for percutaneous transluminal coronary angioplasty (PTCA) versus provisional stenting (Stent) for reducing rates of myocardial infarction (MI) or death (Savage, 1997).

Group	Sample	Survival	Proportion
PTCA	107	83	.78
Stent	108	90	.83

- $\hat{\delta} = .05$
- $\chi^2(1) = 0.80$, $p = .37$
- $95\%\text{CI} = (-.06 : .17)$

Sources of Priors

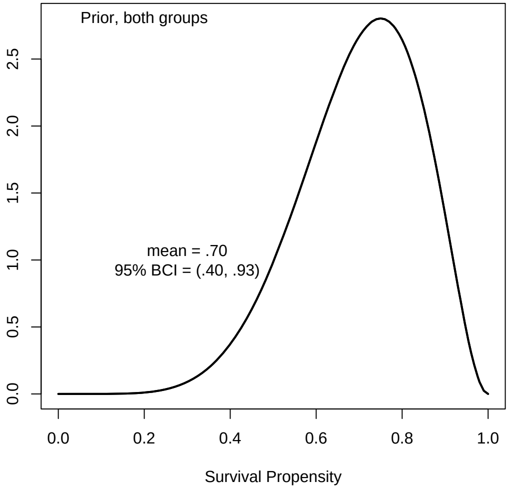
One of the biggest problems in using Bayesian statistics is the requirement of a prior probability for the parameter(s).

- Personal: True beliefs, tacit beliefs, elicitation.
- Expert: My priors $\leftarrow$ expert priors
- Scientific Community: Consensus versus adversarial.
- Previous data: Discounting.
- Theory.
- Technical: Conjugate, approximations.
- Non-informative: Weakly informative, reference (Objective Bayes).

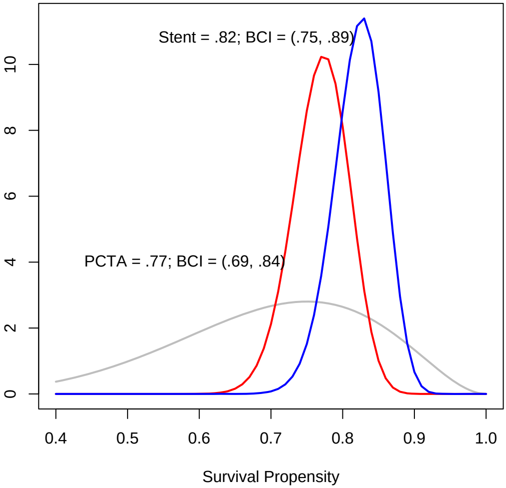
Swamping of Priors

- Dogmatic prior: Little or no uncertainty: prior probability concentrated on very small interval or single point.
- Diffuse prior: Moderate or large amount of uncertainty. Probability is diffused over a region and not concentrated at single points.
- Data Swamping: Sufficient data will ‘swamp’ a diffuse prior (Edwards, Lindman, & Savage, 1963).
- Bayesian “Central Limit Theorem”: In most cases, with sufficient data, the posterior distribution of a parameter will have an approximately normal distribution (Lindley, 1965).

Prior Survival Propensity



Posterior Survival Propensity

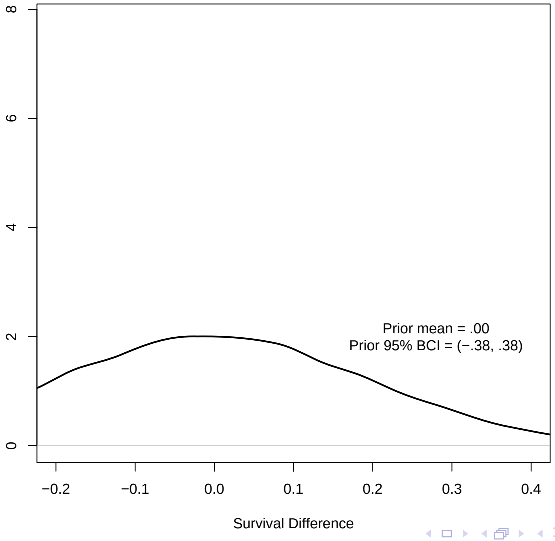


Posterior Analysis: Densities for Difference

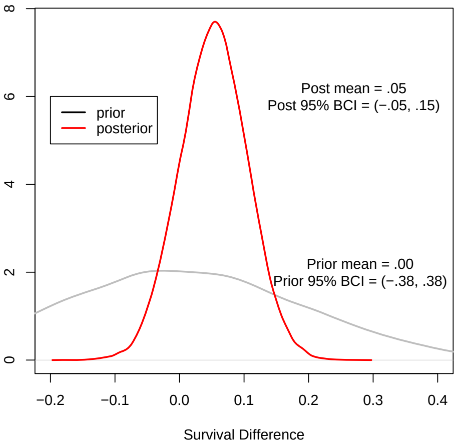
Let δ denote the difference between the Stent and PTCA propensities of survival.

- Density of δ. (δ is now a random variable)
- Density of difference of two beta densities
- Analytically extremely complicated
- Obtain density of difference by simulation

Prior Survival Difference



Posterior Survival Difference



Models of Interest

Recall δ denotes the difference between the Stent and PTCA propensities of survival. Consider the following models:

Superiority: M_S	$.05 < \delta$	
Equivalence: M_E	$-.05 \leq \delta \leq .05$	
Inferiority: M_I	$\delta < -.05$	
Non-inferiority: M_{ES}	$-.05 \leq \delta$	$M_E \cup M_S$
Non-Superiority: M_{EI}	$\delta \leq .05$	$M_E \cup M_I$
Non-Equivalence: M_{SI}	$\delta < -.05$ or $.05 < \delta$	$M_S \cup M_I$

Models of Interest

Recall δ denotes the difference between the Stent and PTCA propensities of survival. Consider the following models:

Superiority: M_S	$.05 < \delta$	
Equivalence: M_E	$-.05 \leq \delta \leq .05$	
Inferiority: M_I	$\delta < -.05$	
Non-inferiority: M_{ES}	$-.05 \leq \delta$	$M_E \cup M_S$
Non-Superiority: M_{EI}	$\delta \leq .05$	$M_E \cup M_I$
Non-Equivalence: M_{SI}	$\delta < -.05$ or $.05 < \delta$	$M_S \cup M_I$

- What are the priors for the Ms?

Models of Interest

Recall δ denotes the difference between the Stent and PTCA propensities of survival. Consider the following models:

Superiority: M_S	$.05 < \delta$			
Equivalence: M_E	$-.05 \leq \delta \leq .05$			
Inferiority: M_I	$\delta < -.05$			
Non-inferiority: M_{ES}	$-.05 \leq \delta$			$M_E \cup M_S$
Non-Superiority: M_{EI}	$\delta \leq .05$			$M_E \cup M_I$
Non-Equivalence: M_{SI}	$\delta < -.05$	or	$.05 < \delta$	$M_S \cup M_I$

- What are the priors for the Ms?
- Uniform priors on all models?

Models of Interest

Recall δ denotes the difference between the Stent and PTCA propensities of survival. Consider the following models:

Superiority: M_S	$.05 < \delta$	
Equivalence: M_E	$-.05 \leq \delta \leq .05$	
Inferiority: M_I	$\delta < -.05$	
Non-inferiority: M_{ES}	$-.05 \leq \delta$	$M_E \cup M_S$
Non-Superiority: M_{EI}	$\delta \leq .05$	$M_E \cup M_I$
Non-Equivalence: M_{SI}	$\delta < -.05$ or $.05 < \delta$	$M_S \cup M_I$

- What are the priors for the Ms?
- Uniform priors on all models?
- Uniform priors on first three models?

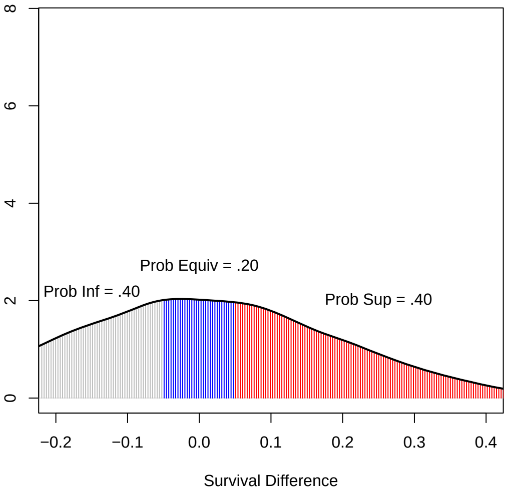
Models of Interest

Recall δ denotes the difference between the Stent and PTCA propensities of survival. Consider the following models:

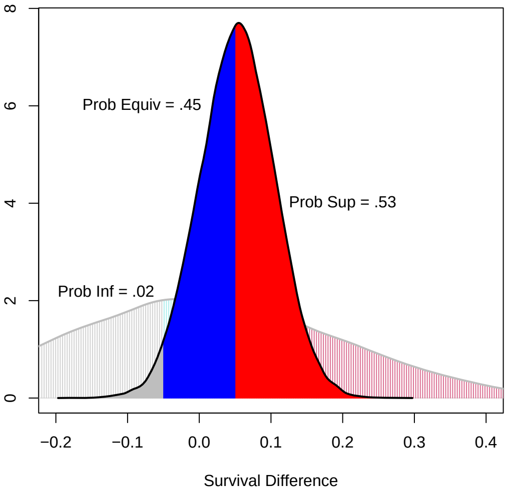
Superiority: M_S	$.05 < \delta$	
Equivalence: M_E	$-.05 \leq \delta \leq .05$	
Inferiority: M_I	$\delta < -.05$	
Non-inferiority: M_{ES}	$-.05 \leq \delta$	$M_E \cup M_S$
Non-Superiority: M_{EI}	$\delta \leq .05$	$M_E \cup M_I$
Non-Equivalence: M_{SI}	$\delta < -.05$ or $.05 < \delta$	$M_S \cup M_I$

- What are the priors for the Ms?
- Uniform priors on all models?
- Uniform priors on first three models?
- Priors are already defined from priors on survival propensities!

Prior Probabilities of Models (δ)



Posterior Probabilities of Models (δ)



Frequentist Hypothesis Testing

- Compare two simple models:
- $M_0 : x \sim N(0, 1)$ versus $M_1 : x \sim N(.5, 1)$
- One sample.
- Select a one-sided test
- Choose the significance level α and power $1 - \beta$.
- Determine the minimum sample size N.
- Obtain a sample of size n, $\{x_1, \ldots, x_n\}$.
- Obtain the test statistic $t = \sqrt{n}\bar{x}$.
- Calculate $p = 1 - \Phi(t)$.

Frequentist Hypothesis Testing

Which study yields the most evidence favoring M_1?

Study	α	$1-\beta$	N	n	p
A	.05	.80	25	9	.06
B	.05	.80	25	25	.05
C	.05	.80	25	50	.025
D	.01	.90	53	53	.01
E	.01	.90	53	106	.005
F	.005	.95	72	72	.005
G	.005	.95	72	144	.0025
H	.001	.99	118	118	.001
I	.001	.99	118	236	.0005

Bayesian Hypothesis Testing

- Compare two simple models:
- $M_0 : x \sim N(0,1)$ versus $M_1 : x \sim N(.5,1)$
- Assume $\Pr(M_0) = \Pr(M_1)$.
- $\ldots\ldots$
- Calculate $p = 1 - \Phi(t)$.
- Treat p as datum.
- Use density of p-statistic.
- Calculate

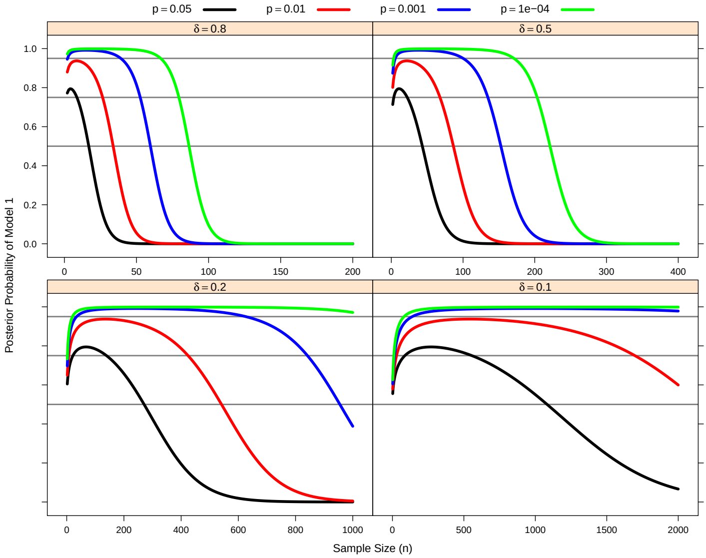
$$\Pr(M_1|p) = \frac{\mathrm{pr}(p|M_1)\Pr(M_1)}{\mathrm{pr}(p|M_1)\Pr(M_1) + \mathrm{pr}(p|M_0)\Pr(M_0)}.$$

Bayesian Hypothesis Testing

Which study yields the most evidence favoring M_1?

Study	α	$1-\beta$	N	n	p	$\Pr(M_1 \mid p)$
A	.05	.80	25	9	.06	.77
B	.05	.80	25	25	.05	.73
C	.05	.80	25	50	.025	.66
D	.01	.90	53	53	.01	.86
E	.01	.90	53	106	.005	.50
F	.005	.95	72	72	.005	.87
G	.005	.95	72	144	.0025	.24
H	.001	.99	118	118	.001	.88
I	.001	.99	118	236	.0005	.014

Posterior Probabilities and Sample Size



Implications for Statistics

1. Statistics is probability theory.
2. Statistics is a logic of inference from data.
3. Parameters can be considered uncertain with a (subjective) probability distribution
4. Uncertainty regarding parameters is updated by observations via Bayes's Theorem.
5. Flow of uncertainty from prior to posterior is objective (deductive), not subjective.

Additional Topics

1. Likelihood Principle: Inference is based solely on data observed, not on data that could have been observed but were not.
2. Data selection mechanisms.
3. Stopping rules.
4. Missing data.
5. Causal modeling.
6. Bayesian model comparison, selection, and averaging.
7. Computation.

Evolution of Statistics