DEPARTMENT OF BIOSTATISTICS
UNIVERSITY of WASHINGTON
School of Public Health

Bayesian Methods for Clinical Research: Introduction

Outline

- Introduction
- Basic concepts of Bayesian inference
- Introduction to Bayesian computing
- Bayesian GLM
- Comparison of approaches to inference
- Interim monitoring of clinical trials

American Journal of Epidemiology

Vol. 141, No. 3
Printed in U.S.A.

Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Tests in the Absence of a Gold Standard

Lawrence Joseph,[1–3] Theresa W. Gyorkos,[1,2,4] and Louis Coupal[2,3]

It is common in population screening surveys or in the investigation of new diagnostic tests to have results from one or more tests investigating the same condition or disease, none of which can be considered a gold standard. For example, two methods often used in population-based surveys for estimating the prevalence of a parasitic or other infection are stool examinations and serologic testing. However, it is known that results from stool examinations generally underestimate the prevalence, while serology generally results in overestimation. Using a Bayesian approach, simultaneous inferences about the population prevalence and the sensitivity, specificity, and positive and negative predictive values of each diagnostic test are possible. The methods presented here can be applied to each test separately or to two or more tests combined. Marginal posterior densities of all parameters are estimated using the Gibbs sampler. The techniques are applied to the estimation of the prevalence of *Strongyloides* infection and to the investigation of the diagnostic test properties of stool examinations and serologic testing, using data from a survey of all Cambodian refugees who arrived in Montreal, Canada, during an 8-month period. *Am J Epidemiol* 1995;141:263–72.

Bayes theorem; diagnostic tests, routine; epidemiologic methods; models, statistical; Monte Carlo method; prevalence; sensitivity and specificity

CASE STUDY

Applications of the Bayesian approach in biomedical research

Medical Decision Making
http://mdm.sagepub.com/

Practice of Epidemiology

Replication of Breast Cancer Susceptibility Loci in Whites and African Americans Using a Bayesian Approach

Katia M. O'Brien*, Stephen R. Cole, Charles Poole, Jeannette T. Bensen, Amy H. Herring, Lawrence S. Engel, and Robert C. Millikan†

SNPs. Our results demonstrate the utility of Bayesian methods in genetic epidemiology and provide support for their application in small, etiologically driven investigations.

A Bayesian Approach to Aid in Formulary Decision Making: Incorporating Institution-Specific Cost-Effectiveness Data with Clinical Trial Results

Shelby D. Reed, Peter W. Dillingham, Andrew H. Briggs, David L. Veenstra and Sean D. Sullivan
Med Decis Making 2003 23: 252
DOI: 10.1177/0272989X03023003007

Then, we adopted a Bayesian, hierarchical, random-effects model to integrate site-specific and clinical trial data. We ap-

J Community Health (2011) 36:819–830
DOI 10.1007/s10900-011-9380-6

ORIGINAL PAPER

Comparing Child Health, Access to Care, and Utilization of Health Services Between Ohio Appalachia's River and Non-River Bordering Counties

Lauren H. Smith · Christopher Holloman

do not border the river. A secondary analysis of the 28 Appalachian counties from Ohio's 88 counties included in the 2008 Ohio Family Health Survey was conducted using a Bayesian Hierarchical Modeling strategy. Descriptive

Environ Resource Econ (2011) 49:597–624
DOI 10.1007/s10640-011-9456-z

The Effect of Risk Context on the Value of a Statistical Life: a Bayesian Meta-model

Thijs Dekker · Roy Brouwer · Marjan Hofkes · Klaus Moeltner

ACADEMIC EMERGENCY MEDICINE 2008; 15:466–475 © 2008 by the Society for Academic Emergency Medicine

Bayesian Logistic Injury Severity Score: A Method for Predicting Mortality Using International Classification of Disease-9 Codes

Randall S. Byrd, MD, PhD, Ming Ouyang, PhD, David Mutlipatti, PhD

Methods: The authors used Bayesian logistic regression to train and test models for predicting mortality based on injury ICD-9 codes (2,230 codes) and injury codes with two-way interactions (943,037 codes and interactions) using data from the National Trauma Data Bank (NTDB). They evaluated discrimination

---

Causal Agent

The nematode (roundworm) *Strongyloides stercoralis*. Other Strongyloides include *S. fülleborni*, which infects chimpanzees and baboons and may produce limited infections in humans.

Life Cycle

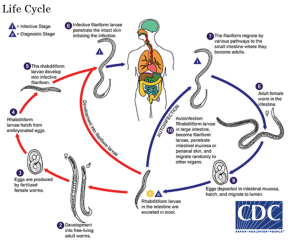


TABLE 1. Results of serologic and stool testing for *Strongyloides* infection on 162 Cambodian refugees arriving in Montreal, Canada, between July 1982 and February 1983

		Stool examination +	Stool examination −	
Serology	+	38	87	125
	−	2	35	37
		40	122	162

- **Goals:**
  - Estimate disease prevalence
  - Estimate sensitivity and specificity of each individual test
  - Estimate sensitivity and specificity of the combined tests
- **Challenge:**
  - No GOLD STANDARD evaluated in the study!

Additional Information

lack of a gold standard for the detection of most parasitic infections means that the properties of these tests are not known with high accuracy. In consultation with a panel of experts from the McGill Centre for Tropical Diseases, we determined equally tailed 95 percent probability intervals (i.e., 2.5 percent in each tail) for the sensitivity and specificity of each test (see table 5). These were derived from a review of the relevant literature and clinical opinion (21–28).

Bayesian Methods

"the explicit quantitative use of external evidence in the design, monitoring, analysis, interpretation and reporting of a health-care evaluation"

(Spiegelhalter, Abrams, Myles, 2004)

Statistical Science
1990, Vol. 5, No. 3, 269–298

Biostatistics and Bayes

Norman Breslow

Abstract. Attitudes of biostatisticians toward implementation of the Bayesian paradigm have changed during the past decade due to the increased availability of computational tools for realistic problems. Empirical Bayes' methods, already widely used in the analysis of longitudinal data, promise to improve cancer incidence maps by accounting for overdispersion and spatial correlation. Hierarchical Bayes' methods offer a natural framework in which to demonstrate the bioequivalence of pharmacologic compounds. Their use for quantitative risk assessment and carcinogenesis bioassay is more controversial, however, due to uncertainty regarding specification of informative priors. Bayesian methods simplify the analysis of data from sequential clinical trials and avoid certain paradoxes of frequentist inference. They offer a natural setting for the synthesis of expert opinion in deciding policy matters. Both frequentist and Bayes' methods have a place in biostatistical practice.

Message is out!

- Science
- The Economist
- The New York Times
  Adding Art to the
  Rigor of Statistical Science
  (Arts & Ideas Section)


Bayes Offers a 'New' Way to Make Sense of Numbers


In praise of Bayes


Arts&Ideas
Adding Art to the Rigor Of Statistical Science

Application Areas include
- Medicine
- Genetics
- Pharmacology
- Epidemiology
- Health services
- Environmental sciences

…

And increasing due to modeling flexibility, computational resources, etc…

Bayesian Software

(Disclaimer: Not intended to provide a complete list of available Bayesian software)

- BUGS/Winbugs/Openbugs/JAGS/Nimble (complex models using MCMC methods)
- INLA (latent Gaussian models; uses Laplace methods)
- BOA/CODA (convergence diagnostics and output analysis)
- BRCAPRO (genetic counseling of women at high risk for breast and ovarian cancer)

…

R-Packages:

- http://cran.r-project.org/web/views/Bayesian.html

  - Download Rstudio: https://www.rstudio.com/products/Rstudio/
  - Download and install R in your computer: http://cran.fhcrc.org/

  Within R session:
  - Install packages with
    - install.packages("mypackage")
  - Load library with
    - library(mypackage)

Primary packages we will use

- LearnBayes
- arm

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY of WASHINGTON
School of Public Health

Basic Concepts/Review

- Probability & Interpretation
- Random Variables
- Likelihood Function
- Traditional Approach to Inference

Diagnostic Testing

- In the presence of a "gold standard"
  - Consider a new diagnostic test

Test	Disease	No Disease
Positive	a (true positives)	b (false positives)
Negative	c (false negatives)	d (true negatives)

- Events:
  - A: {test positive}
  - B: {disease} ⟶ P(B): disease prevalence

Diagnostic Testing

Test	Disease	No Disease
Positive	a	b
Negative	c	d

- **Sensitivity:** the ability of the test to identify correctly those who have the disease among all individuals with the disease

$$Sensitivity: P(A \mid B) = \frac{a}{a+c}$$

- **Specificity**: the ability of the test to identify correctly those who do not have the disease among those free from the disease

$$Specificity: P(A^c \mid B^c) = \frac{d}{b+d}$$

- These are test characteristics.

Diagnostic Testing

Test	Disease	No Disease
Positive	a	b
Negative	c	d

- **Positive predictive value (PPV):** The proportion of patients have the disease among those who tested positive

$$PPV: P(B \mid A) = \frac{a}{a+b}$$

- **Negative predictive value (NPV)**: The proportion of patients are actually free of the disease among those who tested negative

$$NPV: P(B^c \mid A^c) = \frac{d}{c+d}$$

Interpretations of Probability

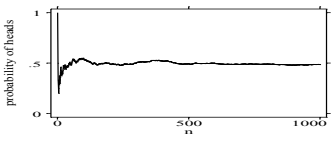
- Classical: If an event can occur in N mutually exclusive and equally likely ways, and if m of these possess a characteristic of interest, E, the probability of the occurrence of E is P(E) = m/N.

  Example: Flip a coin.

  What is the probability of getting a head?

- Frequentist: If some experiment is repeated a large number of times n and if some resulting event with the characteristic E occurs m times, the relative frequency of occurrence of E is approximately equal to the probability of E, that is, P(E) = m/n

Example: Around 1900, Karl Pearson tossed a coin 24,000 times and recorded 12,012 heads, giving a proportion of 0.5005.



Interpretations of Probability: Subjective

- Your degree of uncertainty.

Example: Will you pass a class?

You will take the class (hopefully!) only once; even if you retake the class next year, you won't be taking it under the same conditions! You'll have a different instructor, a different set of courses, and possibly different working conditions!

Diagnostic Testing

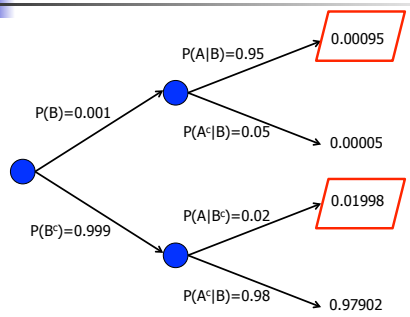
- A new HIV test is known to have 95% sensitivity and 98% specificity. In a population with HIV prevalence of 1/1000, what is the probability that someone testing positive (event *A*) actually has HIV (event *B*)?

  Prevalence = 1/1000

  Sensitivity = $P(A|B) = 0.95$

  Specificity = $P(A^c|B^c) = 0.98 = 1-P(A|B^c) = 1-\text{False Positive}$

Diagnostic Testing



Conditional on positive test result:

PPV= 0.00095/(0.0095+0.01998)=0.045

Diagnostic Testing

- A new HIV test is known to have 95% sensitivity and 98% specificity. In a population with HIV prevalence of 1/1000, what is the probability that someone testing positive (event *A*) actually has HIV (event *B*)?

  Prevalence = 1/1000
  Sensitivity = P(A|B) = 0.95
  Specificity = P(A^c|B^c) = 0.98 = 1-P(A|B^c) = 1-False Positive

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}$$

Bayes Rule!

$$= \frac{0.95 \times 0.001}{0.95 \times 0.001 + 0.02 \times 0.999} = \frac{0.00095}{0.02093} = 0.045$$

Positive Predictive Value

Diagnostic Testing

- A new HIV test is known to have 95% sensitivity and 98% specificity. In a population with HIV prevalence of 1/100, what is the probability that someone testing positive (event *A*) actually has HIV (event *B*)?

  Prevalence = 1/100
  Sensitivity = P(A|B) = 0.95
  Specificity = P(A^c|B^c) = 0.98 = 1-P(A|B^c)

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}$$

Bayes Rule!

$$= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.02 \times 0.99} = 0.324$$

Positive Predictive Value

Diagnostic Testing

- Question: *How should the test result change our belief about the probability of disease?*
  - Our intuition is poor when processing probabilistic evidence, i.e., when updating our probability in the presence of new evidence. Bayes rule shows exactly how to do this!
  - **The disease prevalence** (0.001) can be thought of as our **prior** probability that the individual has the disease.
  - Observing a positive result (i.e. data) changes this probability to 0.045 for the tested individual. This is our updated or ***posterior*** *probability* that the individual has the disease.
  - The posterior probability depends on the test's operating characteristics (e.g. sensitivity/specificity, test results and prevalence).

Diagnostic Testing

- Questions:
  - *Having observed a positive test result for a subject, what is the probability that the next subject also has a positive test result?*
  - *How would the new test result change the current belief about the probability of disease?*

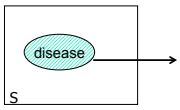
Guiding principle*: Today's posterior is tomorrow's prior!*

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}$$
$$= \frac{0.95 \times 0.045}{0.95 \times 0.045 + 0.02 \times (1 - 0.045)} = \frac{0.04275}{0.06185} = 0.691$$

What is a probability model?

- Random variable:
  - "Rule" that assigns a "value" to each point of the sample space

Example



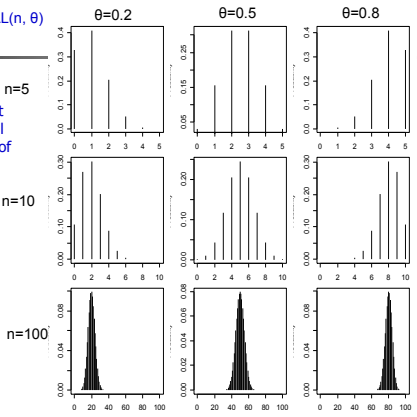
$$X = \begin{cases} 1, & \text{if subject has disease} \\ 0, & \text{otherwise} \end{cases}$$

- Probability model (of a random variable):
  - Defines what values the variable can take and how to assign probabilities to those values.

    Example: X ~ Bernoulli(p); p is the probability of disease

BINOMIAL(n, θ)

Number of times that a particular event will occur in a sequence of n independent observations with the same probability of occurrence

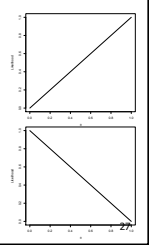


What is a likelihood function?

- A **likelihood function** (or simply the likelihood) is a function of the parameters of a probability model given the outcomes.
  - The *likelihood* of θ, given outcome y, is equal to the *probability* of that observed outcome given θ.

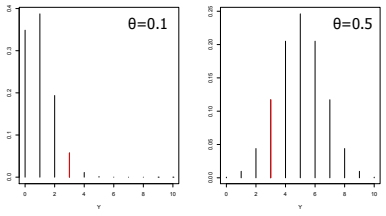
What is a likelihood function?

- Bernoulli model:
  - Random variable Y takes on two possible values: 0 or 1
    - P(Y=1|θ) = θ,
    - P(Y=0|θ) = 1-θ, where θ is a number in [0,1]
  - Likelihood function based on a Bernoulli observation:
    - Given that y=1, the likelihood function of θ is:
      - L(θ|y=1) = P(Y=1|θ)=θ
    - Given that y=0, the likelihood function of θ is:
      - L(θ|y=0) = P(Y=1|θ)=1-θ



What is a likelihood function?

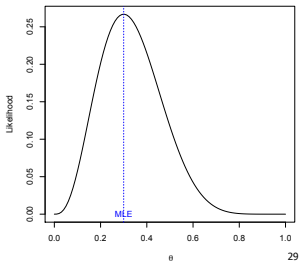
- Binomial Model
  - Test results in a random sample of 10 disease subjects: (0, 1, 0, 0, 0, 1, 0, 0, 0, 1)
  - Probability model for number of positive tests:
    - Y ~ Binomial(10, θ)



What is a likelihood function?

- Likelihood function: $L(\theta \mid Y) = \binom{10}{3}\theta^3(1-\theta)^7$

What is the value of θ that maximizes the likelihood?



Traditional Approach to Inference

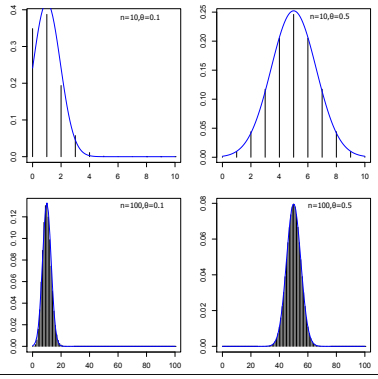
Under certain regularity conditions and for large samples:

$$\hat{\theta}_{MLE} \dot{\sim} N\left(\theta, I^{-1}(\theta)\right), \quad \text{where } I(\theta) = E_{Y|\theta}\left[-\frac{\partial^2 \log L(\theta \mid y)}{\partial \theta^2}\right]$$

Traditional Approach to Inference

Sampling Distribution

(Binomial and Normal Approximation)



Traditional Approach to Inference

H_0: θ = 0.5

H_1: θ ≠ 0.5

```
> binom.test(3,10,p=0.5)                    Uses Exact Method
        Exact binomial test
data:  3 and 10
number of successes = 3, number of trials = 10, p-value = 0.3438
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.06673951 0.65245285
sample estimates:
probability of success
                   0.3

                                            Uses Normal Approximation
> prop.test(3,10,p=0.5)
        1-sample proportions test with continuity correction
data:  3 out of 10, null probability 0.5
X-squared = 0.9, df = 1, p-value = 0.3428
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.08094782 0.64632928
sample estimates:
  p
0.3
```


Traditional Approach to Inference

- P-value interpretation?
  - Under the null hypothesis, the probability of observing an equal or more extreme number of test results is 34%.
  - *It is not the probability of the null hypothesis!*
- Confidence interval interpretation?
  - The confidence interval gives values of the population parameter for which the observed sample proportion is not statistically significant at the 5% level
  - *It does not give us the probability that the true parameter lies between the boundaries of the interval!*

Traditional Approach to Inference

- Often inferences rely on asymptotic results
  - Valid inferences only with large samples

```
> prop.test(0,10,p=0.5)
        1-sample proportions test with continuity correction
data:  0 out of 10, null probability 0.5
X-squared = 10, df = 1, p-value = 0.001565
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.0000000 0.2775328
sample estimates:
  p
0
```

Truncated at zero

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY of WASHINGTON
School of Public Health

Bayesian Approach to Inference

- Overview
- Prior Elicitation
- Prior Distributions
- Introduction to Bayesian Computation

Overview of the Bayesian approach

- Began with the work by Thomas Bayes who, in 1763, formalized what is now called Bayes Theorem.

$$P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A)}$$

where: $P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$

Example: Diagnostic testing

- Data → Result of test
- Parameter → True disease status

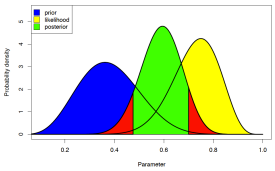
Model
- Prevalence → PRIOR PROB. OF DISEASE
- Sensitivity → LIKELIHOOD of disease given positive test
- Specificity → LIKELIHOOD of no disease given negative test

Bayes Theorem $P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A)}$

- Positive Predictive Value → POSTERIOR PROB. OF DISEASE GIVEN POSITIVE TEST
- Negative Predictive Value → POSTERIOR PROB. OF NO DISEASE GIVEN NEGATIVE TEST

Overview of the Bayesian approach

- Moving towards a generic formulation:
  - Goal: learning about an unknown parameter θ (possibly a vector)
    - θ = true disease status
    - θ = hazard ratio
    - θ = probability that experimental treatment is better
    - θ = vector of regression coefficients
    - θ = missing data
    - etc...
  - Data: y (e.g. test result)
  - Input of analysis:
    - Prior distribution: P(θ)
    - Probability Model: P(y|θ)
    - Likelihood Function: L(θ|y) ∝ P(y|θ)
  - Output of analysis:
    - Posterior distribution:

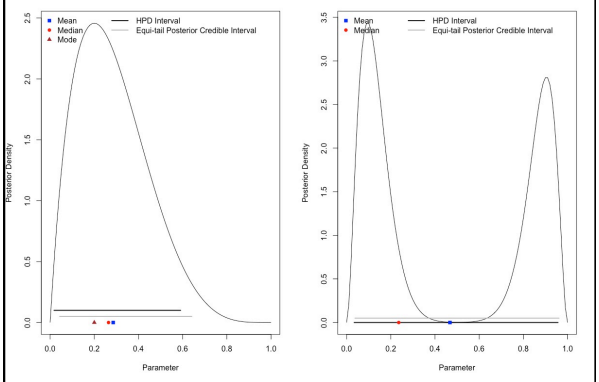


$$P(\theta \mid Y) = \frac{P(\theta)L(\theta \mid Y)}{\int P(\theta)L(\theta \mid Y)d\theta}$$

Overview of the Bayesian approach

- Inferences based on summaries of the <u>posterior distribution</u>
  - Point estimates:
    - Mean/Median/Mode
  - Interval estimates:
    - One-sided credible intervals
    - Two-sided credible intervals
      - Equi-tail area
      - Narrowest interval
        [HPD: highest posterior density intervals]

Choices of summary measures justified with loss functions [decision theory].



Prior Distributions

- Quantifiable (prior) beliefs exist in medicine
  - "… it is generally unrealistic to hope for large treatment effects…"
  - "… it might be reasonable to hope that a new treatment for acute stroke or acute myocardial infarction could reduce recurrent stroke or death rates in hospital from 10% to 9% or 8%, but not to hope that it could halve in-hospital mortality"

(Peto and Baigent, 1998, BMJ)

Prior Distributions

- Key role in Bayesian analysis
- Choice of priors is based on judgments and a degree of subjectivity cannot be avoided
  - Prior is not unique!
  - Sensitivity analysis is crucial in assessing the impact of particular distributions on the conclusions.
- Can we turn informal prior knowledge into a mathematical prior distribution? How?

CASE STUDY

Childhood Polyarteritis nodosa

PLoS One. 2015 Mar 30;10(3):e0120981. doi: 10.1371/journal.pone.0120981. eCollection 2015.

Elicitation of expert prior opinion: application to the MYPAN trial in childhood polyarteritis nodosa.

Hampson LV[1], Whitehead J[1], Eleftheriou D[2], Tudur-Smith C[3], Jones R[4], Jayne D[5], Hickey H[6], Beresford MW[7], Bracaglia C[8], Caldas A[9], Cimaz R[10], Dehoorne J[11], Dolezalova P[12], Friswell M[13], Jelusic M[14], Marks SD[15], Martin N[16], McMahon AM[17], Peitz J[18], van Royen-Kerkhof A[19], Soylemezoglu O[20], Brogan PA[2].

Author information

Abstract

OBJECTIVES: Definitive sample sizes for clinical trials in rare diseases are usually infeasible. Bayesian methodology can be used to maximise what is learnt from clinical trials in these circumstances. We elicited expert prior opinion for a future Bayesian randomised controlled trial for a rare inflammatory paediatric disease, polyarteritis nodosa (MYPAN, Mycophenolate mofetil for polyarteritis nodosa).

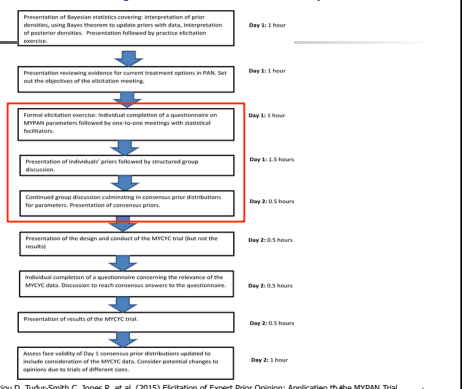
METHODS: A Bayesian prior elicitation meeting was convened. Opinion was sought on the probability that a patient in the MYPAN trial treated with cyclophosphamide would achieve disease remission within 6-months, and on the relative efficacies of mycophenolate mofetil and cyclophosphamide. Expert opinion was combined with previously unseen data from a recently completed randomised controlled trial in ANCA associated vasculitis.

RESULTS: A pan-European group of fifteen experts participated in the elicitation meeting. Consensus expert prior opinion was that the most likely rates of disease remission within 6 months on cyclophosphamide or mycophenolate mofetil were 74% and 71%, respectively. This prior opinion will now be taken forward and will be modified to formulate a Bayesian posterior opinion once the MYPAN trial data from 40 patients randomised 1:1 to either CYC or MMF become available.

CONCLUSIONS: We suggest that the methodological template we propose could be applied to trial design for other rare diseases.

CASE STUDY

Fig 2. Flow diagram illustrating the sequence of activities undertaken during the MYPAN prior elicitation meeting and the time allocated to each activity.



Hampson LV, Whitehead J, Eleftheriou D, Tudur-Smith C, Jones R, et al. (2015) Elicitation of Expert Prior Opinion: Application to the MYPAN Trial in Childhood Polyarteritis Nodosa. PLoS ONE 10(3): e0120981. doi:10.1371/journal.pone.0120981
http://journals.plos.org/plosone/article?id=info:doi/10.1371/journal.pone.0120981

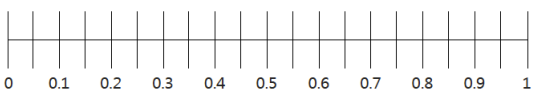
CASE STUDY

S1 File: Structured questionnaire designed to systematically ascertain prior opinion regarding outcomes for treatment with CYC and MMF

NAME:

Before any data are observed, please answer the following questions to specify your prior distributions.

Mark on the scales below your answers to the following questions (to the nearest 0.05).



Questionnaire

- Q1: What do you think the 6-month remission rate for children with PAN treated with cyclophosphamide (CYC) in combination with corticosteroids (steroids) is?
- Q2: Provide a proportion such that you are 75% sure that the true 6-month remission rate on CYC/steroids exceeds this value.

Because of the unpleasant side-effects of CYC, mycophenolate mofetil (MMF) might be considered the preferable treatment even if it is associated with a somewhat lower 6-month remission rate:

- Q3: What is the chance that the 6-month remission rate on MMF/steroids is higher than that on CYC/steroids?
- Q4: What is the chance that the 6-month remission rate on CYC/steroids exceeds that on MMF/steroids by more than 10%?

Please answer the following questions which will allow us to check the adequacy of your fitted prior distributions.

- Q5: What do you think the 6-month remission rate on MMF/steroids is?
- Q6: Provide a proportion such that you are 75% sure that the true 6-month remission rate on MMF/steroids exceeds this value.

S1 Table: Individual experts' final answers to Q1-Q4 and consensus answers agreed by the group before results from the MYCYC trial were revealed

Expert	Q1	Q2	Q3	Q4
1	0.65	0.45	0.63	0.05
2	0.85	0.60	0.35	0.20
3	0.80	0.55	0.10	0.50
4	0.85	0.65	0.20	0.40
5	0.70	0.60	0.20	0.20
6	0.80	0.80	0.15	0.10
7	0.75	0.50	0.10	0.15
8	0.75	0.55	0.30	0.20
9	0.70	0.60	0.20	0.10
10	0.70	0.60	0.25	0.25
11	0.75	0.55	0.30	0.20
12	0.70	0.50	0.10	0.30
13	0.75	0.40	0.20	0.15
14	0.80	0.55	0.20	0.35
15	0.80	0.60	0.20	0.30
Mean	**0.76**	**0.57**	**0.23**	**0.23**
Median	**0.75**	**0.55**	**0.20**	**0.20**
Consensus values†	**0.70**	**0.50**	**0.30**	**0.30**

Q1: What do you think the 6-month remission rate for children with PAN treated with cyclophosphamide (CYC) in combination with corticosteroids (steroids) is?
Q2: Provide a proportion such that you are 75% sure that the true 6-month remission rate on CYC/steroids exceeds this value.
Q3: What is the chance that the 6-month remission rate on MMF/steroids is higher than that on CYC/steroids?
Q4: What is the chance that the 6-month remission rate on CYC/steroids exceeds that on MMF/steroids by more than 10%?

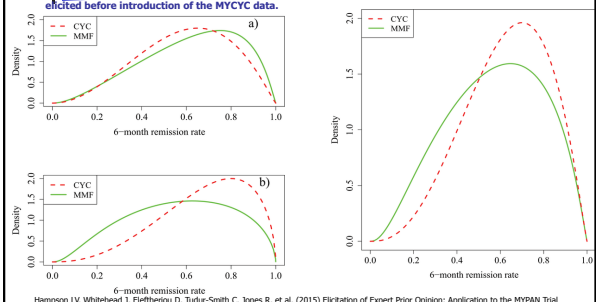
Consensus Prior

- Consensus to questions determined by vote.
  - Experts voted for the pair of answers to (Q1, Q2) which they thought best reflected their prior opinion for p_C.
  - Votes cast between pairs of answers (0.7, 0.5) and (0.75, 0.55), received 10 (67%) and 4 (27%) votes, respectively; one expert abstained.
  - Consensus answers were those voted for by the majority as reflecting their opinion.
- Consensus to (Q3, Q4) determined similarly
  - Experts votes between the following pairs of answers: (0.3, 0.3) and (0.3, 0.35) received 12 (80%) and 3 (20%) votes, respectively.

CASE STUDY

Fig 3. Expert prior opinion before introduction of the MYCYC data regarding 6-month remission rates using treatment with CYC or MMF for children with PAN.

Fig 1. Range of prior opinions elicited before introduction of the MYCYC data.



Hampson LV, Whitehead J, Eleftheriou D, Tudur-Smith C, Jones R, et al. (2015) Elicitation of Expert Prior Opinion: Application to the MYPAN Trial in Childhood Polyarteritis Nodosa. PLoS ONE 10(3): e0120981. doi:10.1371/journal.pone.0120981
http://journals.plos.org/plosone/article?id=info:doi/10.1371/journal.pone.0120981

PLOS ONE

Prior elicitation

- Elicitation of prior distributions can be made from a number of people (for example, clinicians and patients)
  - Combined group (hierarchical) prior distribution
  - Consensus
  - Multiple prior distributions
    - Clinical prior: averages prior distributions elicited from experts
    - Vague prior: leads to a posterior distribution proportional to the likelihood
    - Skeptical prior: represents no treatment effect
    - Enthusiastic prior: represents large treatment effect

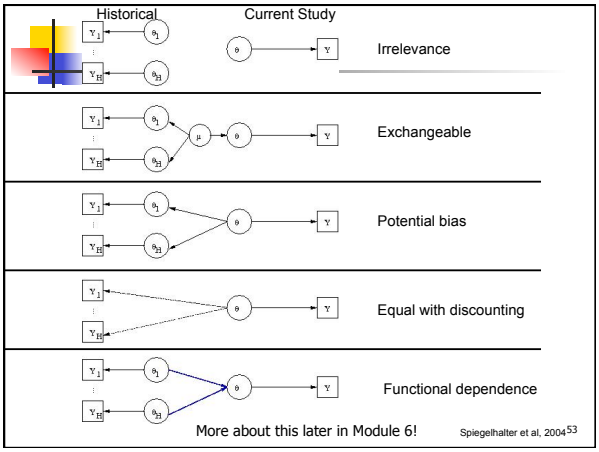
Prior elicitation

- General recommendations:
  - Interactive feedback: helps formulate probabilistic ideas and to reconcile inconsistencies
  - Scripted interview: uniformity in the elicitation process across experts
  - Review: the expert should have access to literature review
  - Percentile: Useful to consider 2.5th and 97.5th percentiles (95% probability intervals)

Prior elicitation

- Problem: how to turn informal opinions into a mathem prior distribution?
  - Summarizing historical evidence
    - Previous similar studies/trials can be used as the basis of a prior distribution
    - Several modeling approaches
      - Degrees of "similarity" between studies/trials
      - Possibility of bias

Note: These approaches are also used when considering historical controls in randomized trials, modeling for potential biases in observational studies and in pooling data for evidence synthesis (meta-analysis)



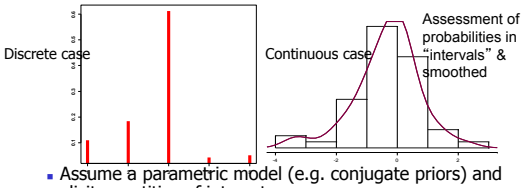
More about this later in Module 6!

Spiegelhalter et al, 2004

Prior elicitation

- Problem: how to turn informal opinions into a mathematical prior distribution?
  - Elicitation of subjective opinion

    - 'Histogram' approach



  - Assume a parametric model (e.g. conjugate priors) and elicit quantities of interest

Prior Distributions

- Conjugate priors
- Non-informative
- Hierarchical priors
- Mixture priors

Prior Distributions

- Conjugate priors:
  - Let F denote a class of sampling distributions p(y|θ) and P a class of prior distributions for θ. Then P is conjugate for F
    p(θ|y) ∈ P for all p(.|θ) ∈ F and p(.) ∈ P
    [prior and the posterior distribution are of the same family].
  - Interpreted as "prior data"
  - Computational convenience

Likelihood	Prior	Posterior
$X\|\theta \sim \mathcal{N}(\theta, \sigma^2)$	$\theta \sim \mathcal{N}(\mu, \tau^2)$	$\theta\|X \sim \mathcal{N}(\frac{\tau^2}{\sigma^2+\tau^2}X + \frac{\sigma^2}{\sigma^2+\tau^2}\mu, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2})$
$X\|\theta \sim \mathcal{B}(n, \theta)$	$\theta \sim \mathcal{B}e(\alpha, \beta)$	$\theta\|X \sim \mathcal{B}e(\alpha + x, n - x + \beta)$
$X_1, \ldots, X_n\|\theta \sim \mathcal{P}(\theta)$	$\theta \sim \mathcal{G}a(\alpha, \beta)$	$\theta\|X_1, \ldots, X_n \sim \mathcal{G}a(\sum_i X_i + \alpha, n + \beta).$
$X_1, \ldots, X_n\|\theta \sim \mathcal{NB}(m, \theta)$	$\theta \sim \mathcal{B}e(\alpha, \beta)$	$\theta\|X_1, \ldots, X_n \sim \mathcal{B}e(\alpha + mn, \beta + \sum_{i=1}^n x_i)$
$X \sim \mathcal{G}(n/2, 2\theta)$	$\theta \sim \mathcal{IG}(\alpha, \beta)$	$\theta\|X \sim \mathcal{IG}(n/2 + \alpha, (x/2 + \beta^{-1})^{-1})$
$X_1, \ldots, X_n\|\theta \sim \mathcal{U}(0, \theta)$	$\theta \sim \mathcal{P}a(\theta_0, \alpha)$	$\theta\|X_1, \ldots, X_n \sim \mathcal{P}a(\max\{\theta_0, x_1, \ldots, x_n\}\alpha + n)$
$X\|\theta \sim \mathcal{N}(\mu, \theta)$	$\theta \sim \mathcal{IG}(\alpha, \beta)$	$\theta\|X \sim \mathcal{IG}(\alpha + 1/2, \beta + (\mu - X)^2/2)$
$X\|\theta \sim \mathcal{G}a(\nu, \theta)$	$\theta \sim \mathcal{G}a(\alpha, \beta)$	$\theta\|X \sim \mathcal{G}a(\alpha + \nu, \beta + x)$

Prior Distributions

- Non-informative:
  (reference prior, vague prior or flat prior)
  - Intended to provide "objective" analysis
    - Connections to Frequentist Inference!
  - Prior is "flat" relative to the likelihood function
    - Minimal impact on the posterior distribution of θ.
  - May be improper (does not "sum up" to 1)
    - DANGER: may lead to improper posteriors!!

Prior Distributions

- Discrete parameter:
  - Discrete uniform prior
    - Example:
      - Parameter = true hypothesis (null or alternative)
      - Prior: $P(H_0)=P(H_1)=0.5$
- Continuous parameter:
  - Jeffreys' prior

$$P(\theta) = |I(\theta)|^{1/2}, \text{where } I(\theta) = E\left[-\frac{\partial^2 \log P(Y|\theta)}{\partial\theta_i \partial\theta_j}\right] \text{(Fisher information)}$$

    - Idea: Fisher information measures the curvature of the log-likelihood. High curvature occurs whenever small changes in the parameter values are associated with large changes in the likelihood. Jeffreys' prior gives more weight to those parameter values, ensuring that the influence of the data and the prior essentially coincide
    - Invariant to transformations of θ

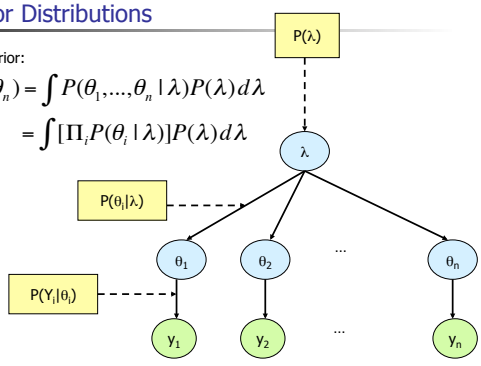
Prior Distributions

- Hierarchical priors:
  - Prior specification in phases
    - Structural division into stages
    - Quantitative (subjective) specification at each stage
- Borrowing strength:
  - improves precision for each parameter
- Nothing prevent us from going further into the hierarchy and adding stages.
  - Harder to interpret parameters in higher levels of the hierarchy
  - Common practice: non-informative priors at the higher levels (of course, "caveats" to such choices)

Prior Distributions

Hierarchical Prior:

$$P(\theta_1,\ldots,\theta_n) = \int P(\theta_1,\ldots,\theta_n \mid \lambda)P(\lambda)d\lambda$$
$$= \int [\Pi_i P(\theta_i \mid \lambda)]P(\lambda)d\lambda$$



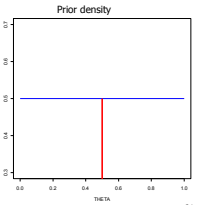
Prior Distributions

- Mixture Prior:
  - Example:

    Test results among 10 disease subjects:
    - (0, 1, 0, 0, 0, 1, 0, 0, 0, 1) : 'successes'=3, 'failures'=7

H_0: $\theta = 0.5$ versus H_1: $\theta \neq 0.5$

Priors for hypotheses:
- $P(H_0)=P(H_1)=0.5$
- Under alternative: $\theta \sim$ Beta(1,1)



- Prior can be re-written as a mixture:

$$P(\theta) = 0.5 \times I_{\{\theta=0.5\}} + 0.5 \times U(0,1)$$

American Journal of Epidemiology
Copyright © 1995 by The Johns Hopkins University School of Hygiene and Public Health
All rights reserved

Vol. 141, No. 3
Printed in U.S.A.

CASE STUDY

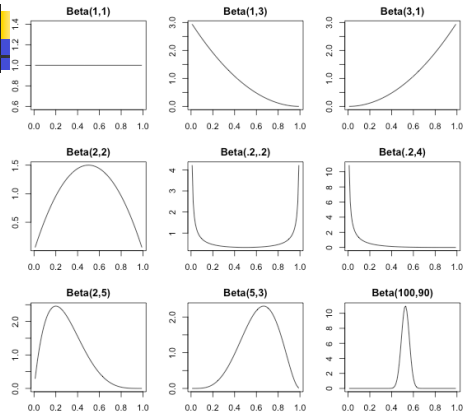
Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Tests in the Absence of a Gold Standard

lack of a gold standard for the detection of most parasitic infections means that the properties of these tests are not known with high accuracy. In consultation with a panel of experts from the McGill Centre for Tropical Diseases, we determined equally tailed 95 percent probability intervals (i.e., 2.5 percent in each tail) for the sensitivity and specificity of each test (see table 5). These were derived from a review of the relevant literature and clinical opinion (21–28).

TABLE 5. Equally tailed 95% probability ranges and coefficients of the beta prior densities for the test parameters in the diagnosis of *Strongyloides* infection*

	Stool examination			Serology		
	Range (%)	Beta coefficients		Range (%)	Beta coefficients	
		α	β		α	β
Sensitivity	5–45	4.44	13.31	65–95	21.96	5.49
Specificity	90–100	71.25	3.75	35–100	4.1	1.76

* A uniform density over the range [0,1] (α=1, β=1) was used for the prior distribution for the prevalence of *Strongyloides* in the refugee population.



Translating the information into a prior distribution

The particular beta prior density for each test parameter was selected by matching the center of the range with the mean of the beta distribution, given by $\alpha/(\alpha+\beta)$, and matching the standard deviation of the beta distribution, given by

$$\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}},$$

with one quarter of the total range. These two conditions uniquely define α and β. An alternative approach is to match the end points of the given ranges to beta distributions with similar 95 percent probability intervals. The coefficients obtained from these two approaches usually give very similar prior distributions. One way to consider a beta(α,β) distribution is to equate it with the information contained in a prior sample of ($\alpha + \beta$) subjects, α of whom were positive. The sum ($\alpha + \beta$) is often referred to as the "sample size equivalent" of the prior information (18).

	Range (%)	Beta coefficients	
		α	β
Sensitivity	5–45	4.44	13.31

Beta distribution obtained by solving these equations:

$$\frac{\alpha}{\alpha+\beta} = \frac{(.45+.05)}{2} = .25$$

$$\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} = \frac{1}{4}(.45-.05) = .10$$

CASE Study

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY of WASHINGTON
School of Public Health

Comparison of approaches to inference

Comparison of approaches to inference

- MLE
- Intervals based on values of θ with large likelihood
- Evidence against null hypothesis via p-values

Traditional Approach

Estimation/Testing satisfying long-run properties (repeated sampling)
- Unbiased estimation
- Confidence intervals
- Type I/II error rates
- Minimax
- Admissibility…

TABLE 4 A taxonomy of six possible 'philosophical' approaches to health technology assessment, depending on their objective and their quantitative use of prior information

	Objective		
	Inference (estimation)	**Hypothesis testing**	**Decision (loss function)**
No prior	Fisherian	Neyman–Pearson	Classical decision theory
Prior	Proper Bayesian	'Bayes's factors'	Full decision-theoretic Bayesian

Comparison of approaches to inference

- Sequential Analysis
  - Data periodically analyzed and study stops if there are sufficiently convincing results
  - Traditional Approach:
    - Identifies "stopping boundaries" with fixed overall Type I error and chooses designs with minimum type II error for particular alternative hypotheses
    - At the end of the study, p-values and confidence intervals are adjusted for the sequential nature of the design
  - Bayesian Approach:
    - Posterior distribution following each observation becomes the prior for the next
    - Posterior distribution does not depend on the stated stopping procedure (data influence the posterior only through the likelihood)

Comparison of approaches to inference

- Sequential use of Bayes Theorem:

$$1.\ p(\theta \mid y_1) \propto p(\theta)p(y_1 \mid \theta)$$

$$2.\ p(\theta \mid y_1, y_2) \propto p(\theta)p(y_1, y_2 \mid \theta)$$
$$\propto p(\theta)p(y_1 \mid \theta)p(y_2 \mid \theta)$$
$$\propto p(\theta \mid y_1)p(y_2 \mid \theta)$$

  - Posterior distribution using initial prior p(θ) given all the data is the same as that obtained sequentially where posterior for the current observation becomes the prior for the next observation.

Comparison of approaches to inference

- P-values and Bayes factors (BF)
  - Example:
    - Model
      - Y ~ Binomial(n, θ)
    - Parameter
      - θ=True unknown population proportion of preference for A
    - Hypotheses
      - H_0: θ = 0.5 versus H_1: θ ≠ 0.5
      - Under alternative θ ~ U(0,1)=Beta(1,1)

$$P(Y = y \mid \theta) = \binom{n}{y} \theta^y (1-\theta)^{(n-y)}$$

Recall:

$$\frac{P(H_0 \mid Data)}{P(H_1 \mid Data)} = \frac{P(Data \mid H_0)}{P(Data \mid H_1)} \times \frac{P(H_0)}{P(H_1)}$$

Bayes Factor

Comparison of approaches to inference

- Bayes Factor (BF):

$$P(Data \mid H_0) = P(Y = y \mid \theta = 0.5) = \binom{n}{y}\left(\frac{1}{2}\right)^y\left(\frac{1}{2}\right)^{n-y} = \binom{n}{y}2^{-n}$$

$$P(Data \mid H_1) = P(Y = y \mid \theta \neq 0.5) = \int P(Y = y \mid \theta)p(\theta)d\theta = \ldots = \frac{1}{n+1}$$

$$BF = \frac{P(Data \mid H_0)}{P(Data \mid H_1)} = \binom{n}{y}\frac{n+1}{2^n}$$

- Alternative: Likelihood-based Bayes Factor (Minimum BF)

$$P(Data \mid H_1) = P(Y = y \mid \theta = \hat{\theta}_{MLE}) = \binom{n}{y}\left(\frac{y}{n}\right)^y\left(1-\frac{y}{n}\right)^{n-y}$$

$$BF_{\min} = \frac{P(Data \mid H_0)}{P(Data \mid H_1)} = \frac{\frac{1}{2^n}}{\left(\frac{y}{n}\right)^y\left(1-\frac{y}{n}\right)^{n-y}}$$

Comparison of approaches to inference

Sample Size	Preference for A	Estimate	P-value (One-sided)	Min. BF	BF
20	15	0.750	0.02	0.07	0.31
200	115	0.575	0.02	0.10	1.20
2000	1046	0.523	0.02	0.12	4.30
2000000	1001445	0.500	0.02	0.12	139.8

- Interpretation of p-values is dependent on sample size!
- Minimum BFs obey the Likelihood Principle, but have similar qualitative behavior to P-values
- Proper BFs can, for large samples relative to the prior precision, support the null hypothesis when a classical analysis would lead to its rejection.
  - This is known as Lindley's paradox
    - Explanation: For large sample sizes, a p-value can be small even if the data support parameter values very close to the null hypothesis. Such data may be unlikely under the null, but even more unlikely under the alternative that spreads the prior over a wide range of values. Thus, the BF can support the null when the significance test would reject it.

Large Sample Properties

$$y = (y_1, \ldots, y_n) \text{ where } y_i \sim p(y_i \mid \theta) \text{ and } p(y \mid \theta) = \prod_{i=1}^{n} p(y_i \mid \theta)$$

$$\text{Let: } I(\theta) = E\left[-\frac{\partial^2 \log P(Y \mid \theta)}{\partial\theta_i \partial\theta_j}\right] \text{ (Fisher information)}$$

Large Sample Properties

- Likelihood-based Inference (MLE)

$$\hat{\theta} \sim N\left(\theta, I^{-1}(\hat{\theta})\right)$$

- Bayesian Inference

$$\theta \sim N\left(\hat{\theta}, I^{-1}(\hat{\theta})\right)$$

  - Thus, the posterior distribution will give essentially the same asymptotic estimates and intervals as the maximum likelihood estimator. However, note that the posterior distribution is a distribution of θ given $\hat{\theta}$ whereas the previous result gives the sampling distribution of $\hat{\theta}$ given θ.
  - This is a nice result as it connects Bayesian and Frequentist analyses. But it is important to note that Bayesian inference does not need to rely on asymptotic result! You get 'exact' inference.

Large Sample Properties

To convince you of the previous result, suppose the parameter is uni-dimensional. Note that we get the same density functions:

$$\hat{\theta} \sim N\left(\theta, I^{-1}(\hat{\theta})\right)$$

$$p(\hat{\theta} \mid \theta) = \frac{1}{\sqrt{2\pi I^{-1}(\hat{\theta})}} \exp\left[-\frac{1}{2I^{-1}(\hat{\theta})}(\hat{\theta}-\theta)^2\right]$$

$$\theta \sim N\left(\hat{\theta}, I^{-1}(\hat{\theta})\right)$$

$$p(\theta \mid \hat{\theta}) = \frac{1}{\sqrt{2\pi I^{-1}(\hat{\theta})}} \exp\left[-\frac{1}{2I^{-1}(\hat{\theta})}(\theta-\hat{\theta})^2\right]$$

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY of WASHINGTON
School of Public Health

Introduction to Bayesian Computation

Introduction to Bayesian Computation: Grid approach (discrete prior for a continuous valued parameter)

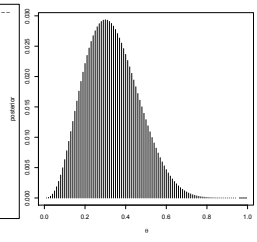
- Bayesian inference can be achieved by approximating the continuous θ with a (dense) grid of discrete values.
- A disadvantage of this approach is that the approximation is only as good as the grid is.
- An advantage of this approach is that it provides flexibility in the choice of prior distributions.
- We will illustrate this approach using
  - "brute-force" method (simple application of Bayes rule) or,
  - R package (LearnBayes)

Introduction to Bayesian Computation

- Test results of 10 disease subjects:
  - (0, 1, 0, 0, 0, 1, 0, 0, 0, 1)
    ('successes'=3, 'failures'=7)
- Parameter of interest:
  - Probability of disease

Introduction to Bayesian Computation: Grid approach (discrete prior for a continuous valued parameter)

```
# Method 1: Brute-force ------------------------------------------
#-- Prior 1:
prior <- rep(1/99, 99)

#-- likelihood times prior
product <- dbinom(x=data[1], size=sum(data), prob=theta)*prior

#-- posterior is the normalized likelihood times prior
posterior <- product/sum(product)

#-- plot posterior distribution
plot(theta, posterior, type='h', xlab=expression(~theta))

#-- posterior mean
mean.post <- sum(theta*posterior)

#-- cumulative posterior distribution
cumulative.post <- cumsum(posterior)

#-- median (approximate)
median.post <- theta[max(which(cumulative.post <=0.50))]
```

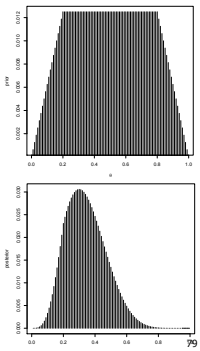


$$P(\theta_i \mid Y) = \frac{P(\theta_i)P(Y \mid \theta_i)}{\sum_j P(\theta_j)P(Y \mid \theta_j)}$$

```
> mean.post
[1] 0.3333333

> median.post
[1] 0.31
```

Introduction to Bayesian Computation: Grid approach (discrete prior for a continuous valued parameter)

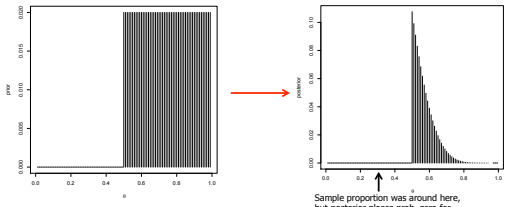
```
# Method 1: Brute-force ----------------------------------------

#-- Prior 2:
f <- theta*(theta<=0.2) +(1-theta)*(theta>=0.8) + 0.2*(theta>0.2 & theta < 0.8)
prior <- f/sum(f)

#-- likelihood times prior
product <- dbinom(x=data[1], size=sum(data), prob=theta)*prior

#-- posterior is the normalized likelihood times prior
posterior <- product/sum(product)

#-- plot of prior distribution
plot(theta, prior, type='h', xlab=expression(~theta))

#-- posterior is the normalized likelihood times prior
posterior <- product/sum(product)

#-- plot posterior distribution
plot(theta, posterior, type="h", xlab=expression(~theta))

#-- posterior mean
mean.post <- sum(theta*posterior)

#-- cumulative posterior distribution
cumulative.post <- cumsum(posterior)

#-- median (approximate)
median.post <- theta[max(which(cumulative.post <=0.50))]
```



```
> mean.post
[1] 0.342133

> median.post
[1] 0.32
```


Introduction to Bayesian Computation: Grid approach (discrete prior for a continuous valued parameter)

Suppose a prior which places probability zero for θ < 0.5 and uniform otherwise



Sample proportion was around here, but posterior places prob. zero for values < .5!

Prior distributions

- Be careful!
  - **Cromwell's rule**:
    - "If a coherent Bayesian attaches a prior probability of zero to the hypothesis that the Moon is made of green cheese, then even whole armies of astronauts coming back bearing green cheese cannot convince him otherwise" (Lindley, 1985)
    - In other words, by placing a prior probability of zero, then there is no learning with data!

Overview of the Bayesian approach

- Likelihood function: $L(\theta \mid Y) = \binom{n}{y} \theta^{y}(1-\theta)^{n-y}$

  where y: number of successes
  n: sample size

- Prior?
  - Let's consider a prior with a functional form that resembles that of the likelihood function
    - Prior should be of the form $\theta^{a'}(1-\theta)^{b'}$
    - It turns out that such a prior for θ is a Beta

Cool fact: multiply likelihood and the prior and you'll again get a function of the same form as the prior…

Overview of the Bayesian approach

- Likelihood function: $L(\theta \mid Y) = \binom{n}{y} \theta^{y}(1-\theta)^{n-y}$
- Prior: $\theta \sim Beta(a,b)$ and $P(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}$
  - a: "prior" successes
  - b: "prior" failures
- Posterior (via Bayes Theorem):

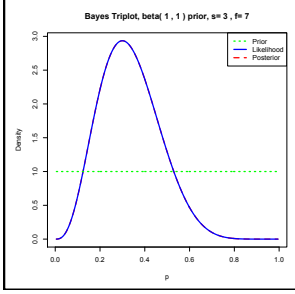
$$P(\theta \mid Y) \propto \theta^{y}(1-\theta)^{n-y}\theta^{a-1}(1-\theta)^{b-1}$$
$$\propto \theta^{a+y-1}(1-\theta)^{b+n-y-1}$$

$$(\theta \mid Y) \sim Beta(a+y, b+n-y)$$

Introduction to Bayesian Computation: conjugate models

Bayesian Inference for a Proportion Using R:

```
library(LearnBayes)
triplot(prior=c(1,1),data=c(3,7))
```

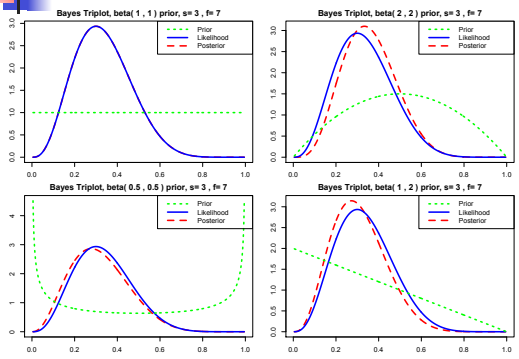


- Point estimation:
  - Mean = 0.333
  - Median = 0.324
  - Mode = 0.300
- Interval estimation:
  - Equal tail 95% credible interval: [0.109, 0.610]
  - 95% HPD: [0.101,0.581]

Interpretation: there is a 95% probability that the test sensitivity lies between [0.101, 0.581]

[Note: we obtain probability statements about θ]

Introduction to Bayesian Computation: conjugate models



Overview of the Bayesian approach

- Hypothesis testing:
  - Hypotheses: H_0 vs. H_1 [simple vs. simple]
  - Prior probabilities: $Pr(H_0)$ & $P(H_1)$
  - Likelihood: $P(Data|H_0)$ & $P(Data|H_1)$
  - Posterior probabilities:
    $P(H_0|Data) = P(H_0)\, P(Data|H_0) / P(Data)$
    where $P(Data) = P(Data|H_0)\, P(H_0) + P(Data|H_1)\, P(H_1)$
  - Odds:

$$\frac{P(H_0 \mid Data)}{P(H_1 \mid Data)} = \frac{P(Data \mid H_0)}{P(Data \mid H_1)} \times \frac{P(H_0)}{P(H_1)}$$

Posterior Odds = Likelihood Ratio x Prior Odds
(a.k.a. Bayes Factor)

Overview of the Bayesian approach

- Strength of evidence provided by Bayes Factor

BF will partially eliminate the influence of the prior and emphasizes the role of data

Bayes Factor	Evidence in favor of H_0 versus H_1
1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 32	Strong
32 to 100	Very strong
>100	Decisive

Overview of the Bayesian approach

Back to example:

- Test results among 10 disease subjects:
  - (0, 1, 0, 0, 0, 1, 0, 0, 0, 1)
    ('successes'=3, 'failures'=7)

Introduction to Bayesian Computation: conjugate models

Back to example:

- Test results among 10 disease subjects:
  - (0, 1, 0, 0, 0, 1, 0, 0, 0, 1)
    ('successes'=3, 'failures'=7)

H_0: $\theta = 0.5$ versus H_1: $\theta \neq 0.5$ Priors for hypotheses: • $P(H_0)=P(H_1)=0.5$ • Under alternative: $\theta \sim$ Beta(1,1)	> pbetat(p0=0.5, prob=0.5, ab=c(1,1), data=c(3,7)) $bf [1] 1.289063 $post [1] 0.5631399

- The posterior probability of the null hypothesis is 0.56

Overview of the Bayesian approach

- Prediction:
  - Prior predictive distribution:

$$P(Y) = \int P(Y \mid \theta) P(\theta) d\theta$$

  - Posterior Predictive Distribution of Y_{NEW}

$$P(Y_{NEW} \mid Data) = \int P(Y_{NEW} \mid Data, \theta) P(\theta \mid Data) d\theta$$
$$= \int P(Y_{NEW} \mid \theta) P(\theta \mid Data) d\theta$$

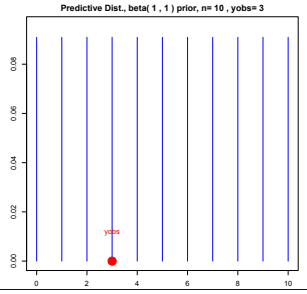
  - Uses:
    - Design and (predictive) power calculations
    - Sequential monitoring
    - Model checking
    - Decision making
    - …

Introduction to Bayesian Computation: conjugate models

Prior Predictive Distribution

```
> pbetap(ab=c(1,1), n=10, s=0:10)
 [1] 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909
 [7] 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909
```

```
> predplot(prior=c(1,1), n=10, yobs=3)
```



Posterior Predictive Distribution

```
> pbetap(ab=c(4,8), n=1, s=0:1)
[1] 0.6666667 0.3333333
```


DEPARTMENT OF BIOSTATISTICS
UNIVERSITY of WASHINGTON
School of Public Health

Bayesian GLM

Generalized Linear Regression Models

- Mean: $E[Y_i \mid X_{i1}, X_{i2}, \ldots, X_{ip}] = \mu_i = g^{-1}(\eta_i)$ where g is a link function
- Regression Model: $g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$
  - Linear regression model
    $$g(\mu_i) = \mu_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$$
  - Logistic regression model
    $$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$$
  - Probit regression model
    $$g(\mu_i) = \Phi^{-1}(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$$
  - Poisson regression model
    $$g(\mu_i) = \log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$$

Bayesian GLM

- Mean: $E[Y_i \mid X_{i1},\ldots,X_{ip}] = \mu_i = g^{-1}(\eta_i)$ where g is a link function
- Regression Model: $g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$
- **Priors**:
  - Regression parameters: $(\beta_0, \beta_1, \beta_2, \ldots \beta_p)$
  - "Nuisance" parameters (e.g. in linear regression σ^2)
- Note:
  - Regression coefficients have the same interpretation (e.g. difference in means; log-odds ratio; etc)
  - Interpretation of inferential results are different (e.g. posterior mean; probability that the regression parameter lies in some interval; etc)

Bayesian GLM in R

- We will use the arm package
- Different approaches to estimation of GLMs
  - Approximate posterior inference (Bayesian CLT)
- Advantages:
  - Syntax very similar to traditional GLMs
  - No need for heavy programming (e.g. MCMC methods)
- Disadvantages:
  - Approximate method under small samples
  - Constrained by model formulations handled by the packages

Bayesian GLM in R: arm package

- Builds on a modification of glm()
  - Uses priors on an augmented regression
  - Uses an approximate EM algorithm to update regression coefficients
    - Gelman, Jakulin, Grazia, Pittau, Su, 2008. A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models. The Annals of Applied Statistics, 2,1360-1383.

Bayesian GLM in R: arm package

- Augmentation Idea (context linear models):

Matrix Formulation:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

In short : $Y = X\beta + \varepsilon$

Prior: $\beta_j \sim N(m_j, v_j^2),\ j = 0,\ldots,p$

Augmented Data: $Y^* = \begin{bmatrix} Y \\ m \end{bmatrix}, X^* = \begin{bmatrix} X \\ I_p \end{bmatrix}$

Bayesian GLM in R: arm package

bayesglm {arm} **R Documentation**

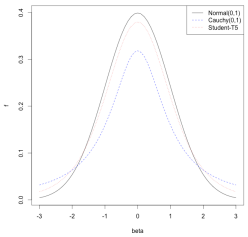
Bayesian generalized linear models.

Description

Bayesian functions for generalized linear modeling with independent normal, t, or Cauchy prior distribution for the coefficients.

Usage

```
bayesglm (formula, family = gaussian, data,
  weights, subset, na.action,
  start = NULL, etastart, mustart,
  offset, control = glm.control(...),
  model = TRUE, method = "glm.fit",
  x = FALSE, y = TRUE, contrasts = NULL,
  drop.unused.levels = TRUE,
  prior.mean = 0,
  prior.scale = NULL,
  prior.df = 1,
  prior.mean.for.intercept = 0,
  prior.scale.for.intercept = NULL,
  prior.df.for.intercept = 1,
  min.prior.scale=1e-12,
  scaled = TRUE, keep.order=TRUE,
  drop.baseline=TRUE, n.iter = 100,
  print.unnormalized.log.posterior=FALSE,
  Warning=TRUE,...)
```



Motivating example: Fracture Intervention Trial

- The Fracture Intervention Trial was an RCT that enrolled women age 55-81 who were at high risk of experiencing a fracture due to low bone mineral density (BMD)
- Women were randomized to receive alendronate or placebo and followed-up to assess the number of osteoporotic fractures they experienced in the subsequent 3 years
- The scientific question of interest is whether alendronate decreases the number of osteoporotic fractures a woman experiences and whether this effect is modified by a woman's baseline fracture risk

Motivating example: Fracture Intervention Trial

- Data for this study are available on the course Github page:
  https://github.com/rhubb/SISCR2017
- Data are for a subset of 344 women and include the following variables

id: participant id

age: age at baseline (years, continuous)

numnosp: number of non-spine osteoporotic fractures (continuous)

trt01: treatment group assignment (0 = placebo, 1 = alendronate)

riskcat4: high risk of fracture (1 = high risk, 0 = low risk)

htotbmd: total BMD (continuous)

FIT: data description and exploration

```
> ## read FIT data set
> fit <- read.csv("https://raw.githubusercontent.com/rhubb/SISCR2017/master/data/FIT.csv", header = T)

> ## examine a few entries of the data set
> head(fit)

  id age numnosp trt01 htotbmd riskcat4
1  1  69       0     0   0.517        1
2  2  76       0     1   0.583        1
3  3  66       0     1   0.709        0
4  4  72       0     0   0.738        0
5  5  58       0     1   0.690        0
6  6  74       0     1   0.480        0

> ## summarize the variables
> summary(fit)

       id              age            numnosp          trt01           htotbmd          riskcat4
 Min.   :  1.00   Min.   :56.00   Min.   :0.0000   Min.   :0.0000   Min.   :0.3990   Min.   :0.0000
 1st Qu.: 86.75   1st Qu.:65.00   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.6060   1st Qu.:0.0000
 Median :268.50   Median :69.00   Median :0.0000   Median :1.0000   Median :0.6675   Median :0.0000
 Mean   :236.46   Mean   :69.31   Mean   :0.1453   Mean   :0.5058   Mean   :0.6622   Mean   :0.1871
 3rd Qu.:360.25   3rd Qu.:74.00   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.7260   3rd Qu.:0.0000
 Max.   :457.00   Max.   :81.00   Max.   :3.0000   Max.   :1.0000   Max.   :0.8740   Max.   :1.0000
                                                                                      NA's   :2
```


FIT: data description and exploration

```
## Summarize number of fractures stratified by treatment group

> by(fit$numnosp,fit$trt01,summary)
fit$trt01: 0
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.0000  0.0000  0.1529  0.0000  3.0000
------------------------------------------------------------------------
fit$trt01: 1
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.0000  0.0000  0.1379  0.0000  3.0000
```

Bayesian GLM in R: arm package

```
> ## -- Normal priors for regression coefficients (with mean=0 and scale=10)
> fit.arm <- bayesglm(numnosp ~ trt01 + htotbmd + riskcat4 + age, data=fit,
          family=poisson, prior.mean=0,    prior.scale=10, prior.df=Inf)

> summary(fit.arm)

Call:
bayesglm(formula = numnosp ~ trt01 * riskcat4, family = poisson,
    data = fit, prior.mean = 0, prior.scale = 10, prior.df = Inf)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6031  -0.5492  -0.5492  -0.5078   3.6251

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.89186    0.21296  -8.883  <2e-16 ***
trt01         -0.15675    0.32242  -0.486   0.627
riskcat4       0.18742    0.54184   0.346   0.729
trt01:riskcat4  0.06999    0.70247   0.100   0.921
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 222.1  on 341  degrees of freedom
Residual deviance: 221.5  on 338  degrees of freedom
  (2 observations deleted due to missingness)
AIC: 315.17

Number of Fisher Scoring iterations: 6
```

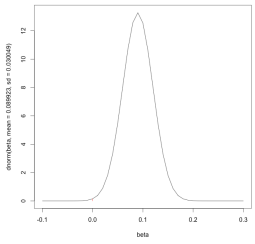
This can be interpreted as posterior mean/median & posterior standard deviations of the regression coefficients

This can be interpreted as two-sided posterior tail probabilities of "no effect"…

Bayesian GLM in R: arm package

- More formally, the posterior probabilities are:

$$2 \times \min(P(\beta_j \leq 0 \mid data), P(\beta_j \geq 0 \mid data))$$



Traditional GLM in R

```
> fit.glm <- glm(numnosp ~ trt01*riskcat4, data=fit, family=poisson)
> summary(fit.glm)

Call:
glm(formula = numnosp ~ trt01 * riskcat4, family = poisson, data = fit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6030  -0.5490  -0.5490  -0.5075   3.6258

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.89256    0.21320  -8.877   <2e-16 ***
trt01         -0.15702    0.32292  -0.486    0.627
riskcat4       0.18782    0.54355   0.346    0.730
trt01:riskcat4  0.07001    0.70508   0.099    0.921
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 222.1  on 341  degrees of freedom
Residual deviance: 221.5  on 338  degrees of freedom
  (2 observations deleted due to missingness)
AIC: 315.17

Number of Fisher Scoring iterations: 6
```

Exercise:
Compare and contrast the Bayesian and traditional GLM results

Bayesian GLM in R: alternative priors

- You can customize choice of prior distribution, mean, and scale
- In this example, results are similar across a wide range of choices
- We will take a closer look at the available options for priors in the lab

```
> ##-- T prior with df = 10 and scale 10
> fit.arm2 <- bayesglm(numnosp ~ trt01*riskcat4, data=fit, family=poisson, prior.mean=0,
+     prior.scale=10, prior.df = 10)
>
> ##-- Cauchy prior with scale 10
> fit.arm3 <- bayesglm(numnosp ~ trt01*riskcat4, data=fit, family=poisson, prior.mean=0,
+     prior.scale=10)
>
> ##-- Normal prior with different prior mean and scale for each coefficient
> fit.arm4 <- bayesglm(numnosp ~ trt01*riskcat4, data=fit, family=poisson, prior.mean=c(log(0.5),0,0),
+     prior.scale=c(1,10,10), prior.df = Inf)
```


A Bayesian perspective on trials of fracture risk

IBMS BoneKEy. 2009 August;6(8):279-294
http://www.bonekey-ibms.org/cgi/content/full/ibmske;6/8/279
doi: 10.1138/20090391

PERSPECTIVES

Interpretation of Randomized Controlled Trials of Fracture Prevention

Tuan V. Nguyen

Osteoporosis and Bone Biology Program, Garvan Institute of Medical Research, Sydney, Australia

Abstract

The question that a reader of a randomized controlled trial (RCT) is interested in is whether therapy is effective. However, prevailing methodology addresses the opposite question: if the therapy is not effective, what is the chance of obtaining the present (or more extreme) data? This current methodology has generated considerable confusion and misinterpretation in the literature. In this *Perspective*, an alternative interpretation of major data from RCTs of fracture prevention is offered in light of Bayesian inference, with the hope that this approach will be adopted more often in future clinical research studies of osteoporosis.
IBMS BoneKEy. 2009 August;6(8):279-294.

Bayesian interpretation of trial results

Table 3. Posterior probability of anti-hip fracture efficacy

Study	Relative risk reduction and 95% CI	Posterior probability of relative risk reduction of hip fracture by at least 25%		
		Vague prior	Skeptical prior	Enthusiastic prior
Alendronate, FIT-1 study (32)	51 (1-77)	0.873	0.687	0.787
Alendronate (5/10 mg), FIT-2 study, T-scores < -2.5 (33)	56 (3-82)	0.893	0.681	0.790
Alendronate (5/10 mg), FIT-2 study, T-scores < -1.6 (33)	21 (+44 to -57)	0.433	0.311	0.413

CLINICAL TRIALS ARTICLE Clinical Trials 2014; 11: 485–493

The utility of Bayesian predictive probabilities for interim monitoring of clinical trials

Benjamin R Saville[a], Jason T Connor[b,c], Gregory D Ayers[a] and JoAnn Alvarez[a]

Background Bayesian predictive probabilities can be used for interim monitoring of clinical trials to estimate the probability of observing a statistically significant treatment effect if the trial were to continue to its predefined maximum sample size.
Purpose We explore settings in which Bayesian predictive probabilities are advantageous for interim monitoring compared to Bayesian posterior probabilities, *p*-values, conditional power, or group sequential methods.
Results For interim analyses that address prediction hypotheses, such as futility monitoring and efficacy monitoring with lagged outcomes, only predictive probabilities properly account for the amount of data remaining to be observed in a clinical trial and have the flexibility to incorporate additional information *via* auxiliary variables.
Limitations Computational burdens limit the feasibility of predictive probabilities in many clinical trial settings. The specification of prior distributions brings additional challenges for regulatory approval.
Conclusions The use of Bayesian predictive probabilities enables the choice of logical interim stopping rules that closely align with the clinical decision-making process. *Clinical Trials* 2014; **11**: 485–493. http://ctj.sagepub.com

CASE STUDY

Background

- Interim analyses for stopping/continuing trials are one form of adaptive trials
- Various metrics for decisions of stopping
  - Frequentist: Multi-stage, group sequential designs, conditional power
  - Bayesian: Posterior distributions, predictive power, Bayes factors
- Question: Why and when should we use Bayesian predictive probabilities for interim monitoring?

Why interim analyses?

- Questions they can address:
  - Is there convincing evidence in favor of the null or alternative hypotheses?
    - Evidence presently shown by data
  - Is the trial likely to show convincing evidence in favor of the alternative hypothesis if additional data are collected?
    - Prediction of what evidence will be available later
- Important factors to consider:
  - ethical imperative to avoid treating patients with ineffective or inferior therapies
  - inefficient allocation of resources

Predictive Probability of Success

- Definition:
  - The probability of achieving a successful (significant) result at a future analysis, given the current interim data
- Computation:
  - Obtained by integrating the data likelihood over the posterior distribution (i.e. we integrate over future possible responses) and predicting the future outcome of the trial
- Decision making:
  - Efficacy rules based either on Bayesian posterior distributions (fully Bayesian) or frequentist p-values (mixed Bayesian-frequentist)

Computation via Simulation

1) At an interim analysis, sample the parameter of interest from the current posterior given current data.
2) Complete the dataset by sampling future samples, observations not yet observed at the interim analysis, from the predictive distribution.
3) Use the complete dataset to calculate success criteria (p-value, posterior probability). If success criteria are met (e.g. p-value < 0.05), the trial is a success.
4) Repeat steps 1-3 a total of B times; the predictive probability (PPoS) is the proportion of simulated trials that achieve success.

Example

- Trial:
  - Single arm Phase II study of 100 patients measuring binary outcome (favorable response to treatment)
  - Goal: compare proportion to a gold standard 50% response rate
- Model: $X \sim Bin(p;N = 100)$ where
  - p = probability of response in the study population
  - N = total number of patients
- Prior: $p \sim Uniform(0,1) = Beta(1,1)$

Example

- Trial Design:
  - Trial is a success if the posterior probability that the proportion exceeds the gold standard is greater than η=0.95, that is,

$$\Pr(p > 0.5|x) > \eta$$

  - Success if 59 or more of 100 patients respond
    - Pr(p > 0.50|x = 58; n = 100) = 0.944
    - Pr(p > 0.50|x = 59; n = 100) = 0.963
  - 3 interim analyses monitoring at 20, 50, and 75 patients

Table 2. Definitions of key measures and methods for illustrative example

Measure/method	Description	Formula
p-value	Probability of observing a proportion equal to or greater than x/n given $H_0: p = p_0$	$\sum_{i=x}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i}$
Posterior probability	Bayesian posterior probability that proportion exceeds the null value p_0	$\Pr(p>p_0|x) = \int_{p_0}^{1} f(x|p)\pi(p)/f(x)\,dp$
Predictive probability	Bayesian predictive probability of statistical significance at N given x/n and $\pi(p)$	$\sum_{y=0}^{m} I\{\Pr(p>p_0|x,y,N)>\eta\} f(y|x)$
Conditional power	Frequentist probability of statistical significance at N given x/n and assumed p_0^*	$\sum_{y=0}^{m}\left[I\left\{\sum_{i=x+y}^{N}\binom{N}{i}p_0^i(1-p_0)^{N-i}<\alpha\right\} f(y|p^*)\right]$
Repeated testing of H_1	Method of monitoring for futility based on *p*-value for test of alternative hypothesis	$p\text{-value} = \sum_{i=0}^{x}\binom{n}{i} p_1^i (1-p_1)^{n-i}$
Group sequential	Frequentist design for interim monitoring that allocates Type I/II errors across interim analyses	Varies by method
Stochastic curtailment	Method that estimates the probability of statistical significance at some future sample size	Varies by method

n and N: number of patients at interim and final sample sizes, respectively; $m = N - n$: number of remaining patients yet to be observed in the study; x: number of successes observed at the interim analysis; y: number of successes yet to be observed in the remaining patients; p_0 and p_1: proportion of successes under the null hypothesis and alternative hypotheses; p^*: estimated or assumed value of p required for conditional power computation; α and η: criteria required to demonstrate 'statistical significance' for *p*-value or posterior probability, respectively, I (): indicator function taking the value 1 if expression is true and 0 if otherwise; $\pi(p)$: beta $(1, 1) = 1$: prior distribution of p, uniform over (0,1); $f(x) = \int_0^1 f(x|p)\pi(p)\,dp$: marginal likelihood or normalizing constant; $f(y|x) = \int_0^1 f(y|p)f(p|x)dp = \int_0^1 f(y|p)f(x|p)\pi(p)/f(x)dp$ = beta-binomial$(m, 1+x, 1+n-x)$: *Bayesian posterior predictive distribution of* y given x; $f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$: data likelihood of x given p for n patients observed by interim; $f(y|p) = \binom{m}{y} p^y (1-p)^{m-y}$: data likelihood of y given p for remaining m patients.

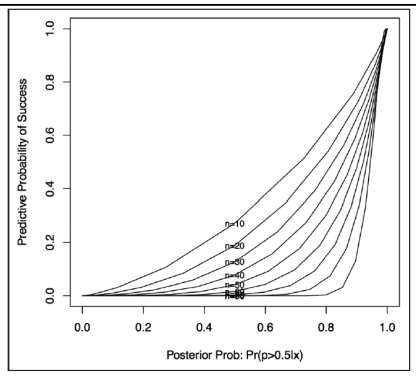
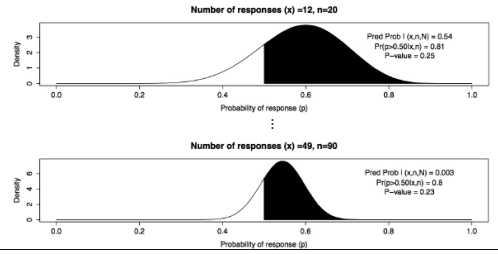


Figure 4. Predictive probability of success versus posterior estimate $\Pr(p > 0.50|x)$ by interim sample size n, with maximum sample size $N = 100$ and posterior threshold $\eta = 0.95$.

Table 1. Illustrative example

n_j	x_j	m_j	y_j^*	p-value	Pr(p>0.5)	CP_{H_a}	CP_{MLE}	PP
20	12	80	47	0.25	0.81	0.90	0.64	0.54
50	28	50	31	0.24	0.80	0.73	0.24	0.30
75	41	25	18	0.24	0.79	0.31	0.060	0.086
90	49	10	10	0.23	0.80	0.013	0.002	0.003

n_j and x_j: the number of patients and successes at interim analysis j; MLE: maximum likelihood estimate; m_j : number of remaining patients at interim analysis j; y_j^*: minimum number of successes required to achieve success; CP_{H_a} and CP_{MLE}: conditional power based on original H_a or MLE; PP: Bayesian predictive probability of success.



R function to compute PP

```
PP <- function(n.total= 100, nullp = 0.5, eta=0.95, data=c(12,8), prior.par=c(1,1), B=1000){
  # posterior
  post.par <- data + prior.par

  # samples from posterior distribution
  post.sample <- rbeta(B, post.par[1], post.par[2])

  # samples new values of x (extending to the maximum sample size)
  x.new <- rbinom(B, size=n.total-sum(data), post.sample)

  # organize data with first column number of 'responses' and second 'non responses'
  data.new <- cbind(x.new, n.total-sum(data)-x.new)

  # posterior parameters given predicted data
  post.pred.par <- cbind(data.new[,1] + post.par[1], data.new[,2]+ post.par[2])

  # posterior probability that P(p > nullp |data)
  post.pred <- pbeta(nullp, post.pred.par[,1], post.pred.par[,2], lower.tail=FALSE)

  # posterior predictive probability of success
  PP <- mean(post.pred > eta)
  return(PP)
}
```

```
> PP(n.total=100, nullp=0.5, eta=0.95, data=c(12,20-12), prior.par=c(1,1), B=1000)
[1] 0.55
> PP(n.total=100, nullp=0.5, eta=0.95, data=c(28,50-28), prior.par=c(1,1), B=1000)
[1] 0.307
> PP(n.total=100, nullp=0.5, eta=0.95, data=c(41,75-41), prior.par=c(1,1), B=1000)
[1] 0.081
> PP(n.total=100, nullp=0.5, eta=0.95, data=c(49,90-49), prior.par=c(1,1), B=1000)
[1] 0.003
```


Final Comments

- There is 'art' in Bayesian Analysis

- Achieving 'mastery' requires practice!