A Succinct Introduction to Bayesian Analysis for Public Health

from p-values to Bayes factors, is my interval confident or credible?

Brock Stewart, PhD
jnn6@cdc.gov

DTBE SEOIB Biostats Seminar
September 28, 2016

Bayesian Analysis in Tuberculosis Research – couple of motivating examples

A Bayesian non-linear mixed effects regression model for the characterisation of early bactericidal activity of tuberculosis drugs

Technical report prepared for Global Alliance for TB Drug Development (TB Alliance)

Divan Aristo Burger[1,2] and Robert Schall[1,2]

[1] *Department of Mathematical Statistics and Actuarial Science, University of the Free State, Bloemfontein, South Africa*
[2] *Quintiles Biostatistics, Bloemfontein, South Africa*

Key Words: Bayesian non-linear mixed effects (NLME) regression model, bi-phasic, colony forming unit (CFU) count, early bactericidal activity (EBA), tuberculosis (TB)

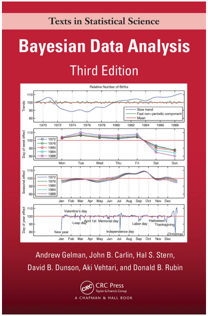
Genetic Structuration, Demography and Evolutionary History of *Mycobacterium tuberculosis* LAM9 Sublineage in the Americas as Two Distinct Subpopulations Revealed by Bayesian Analyses

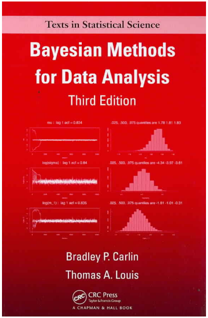
Yann Reynaud*, Julie Millet, Nalin Rastogi*

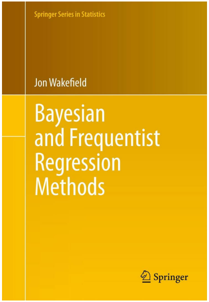
WHO Supranational TB Reference Laboratory, Tuberculosis and Mycobacteria Unit, Institut Pasteur de la Guadeloupe, Abymes, Guadeloupe, France

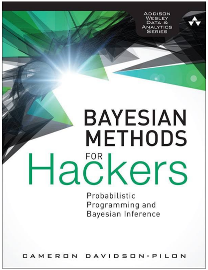
Limitations of this presentation

- This is not a "how-to"
- A proper intro to Bayesian methods requires a 2-semester PhD-level course ...
- ... which is not to say you can't learn on your own ... In fact, I encourage you to read about Bayesian analysis, it teaches a lot about scientific reasoning





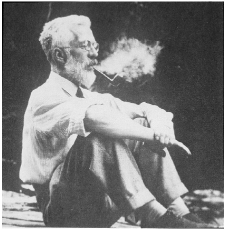




Goals of this presentation

- Explain difference between "regular" and "Bayesian" statistics
- Describe main tenants of the Bayesian approach
- Describe scenarios where a public health scientist trained and entrenched in "regular" statistics might reasonably choose to use Bayesian methods





An absurdly crude history of “modern” Statistics [wikipedia]

- **mid 1700’s:** The term “statistics” for demographic, economic data by states, mostly for tax, war purposes – Latin *statisticum collegium*, “council of state”
- **Late 1700’s:** Bayes, Laplace – Establish “Bayes’ theorem”
- **1800’s – early 1900’s:** Gauss, Pearson, Fisher, Neyman, Wald – Large-sample theory, least squares, use of probability theory, experimental design, hypothesis testing
- **Most of 1900’s:** Bayesian-style ideas and proponents largely dismissed
- **1950** The term “Bayesian” used by those dissatisfied with limitations of frequentist statistics
- **1990:** Bayesian approach sees revival, less acrimony between frequentist and Bayesian, advancements in computing and algorithms

Which hoe is the best?

No need to choose only one and stick with it forever, acquire and learn to use both

When you need a hoe, use whichever you believe is best for the situation

How are “Bayesian Statistics” different from “Statistics”

- Bayesian approach inherently promotes thought, description, and reasoning rather than cookbook or black-box solutions
  - Note, this implies more work for the analyst, but in a good way
- All stat's you've learned may be called “frequentist” statistics

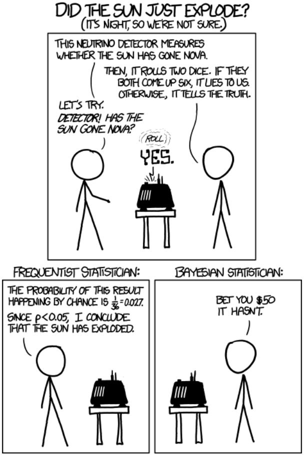
Frequentist

- $Pr(Y|H_0)$
- Parameters are fixed, data are random
- Uncertainty: frequency in hypothetical repeated samples
- Conclusions based on point estimates, CI, p-values

Bayesian

- $Pr(H_0|Y)$
- Data are fixed, parameters are random
- Uncertainty: random parameters
- Conclusions based on the complete *posterior* probability

Frequentist vs. Bayesian – XKCD



Not all experiments are repeatable

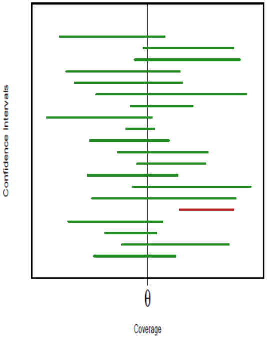
Frequentist vs. Bayesian 90% Intervals

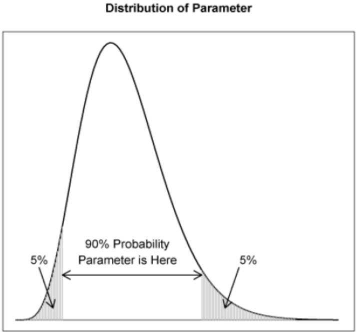
Frequentist

- If we repeat the experiment many times, 90% of the CI will encompass the true value *and* 10% will *not*

Bayesian

- Given this data, this is the interval that has 90% chance of containing the true value





Baye's Rule

To make probability statements about θ given y, we start with a model. The model is a joint probability of θ and y. Specify a *prior* distribution on the paraemter(s):

$$p(\theta)$$

Called *prior* to reflect our belief before considering data. The data are generated from a *sampling distribution*:

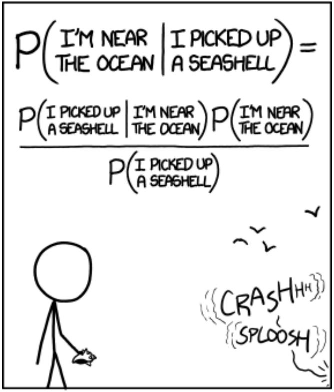
$$p(y|\theta)$$

From the probability axioms (laws), we can get

$$p(\theta, y) = p(\theta)p(y|\theta) = p(y)p(\theta|y)$$

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}$$

Baye's Rule – XKCD



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

$\theta =$ "I'm Near The Ocean"
$y =$ "I Picked Up a Seashell"

In this case,

$$p(y|\theta) \approx p(y)$$

General layout of a Bayesian model

Data is generated from (e.g. Normal, Binomial, Poisson):

$$p(y|\theta)$$

But we don't know the value of θ, and we reflect our uncertainty by

$$p(\theta)$$

What we really want is to update our knowledge of θ using the data

$$p(\theta|y)$$

Which we can obtain from Bayes' Rule

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}$$

The Three Steps of Bayesian Analysis

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}$$

- Probability model: specify the sampling distribution and prior
- Conditioning: obtain the posterior distribution given the data
- Evaluate: does the model fit the data well; how sensitive are the conclusions to the model assumptions, including the prior?

Generally, will repeat these three steps by expanding the model, trying different priors

Conjugate Priors

For a given $p(y|\theta)$, some choices of $p(\theta)$ may lead to the posterior $p(\theta|y)$ being in the same family as the prior.

E.g. if the sampling distribution is

$$p(y|\pi) = Bin(y|n, \pi)$$

(for known, given n) and the prior is

$$p(\pi) = Beta(\alpha, \beta)$$

(for some α, β) then the posterior is

$$p(\pi|y) = Beta(\alpha + y, \beta + n)$$

Conjugate Priors – In Words

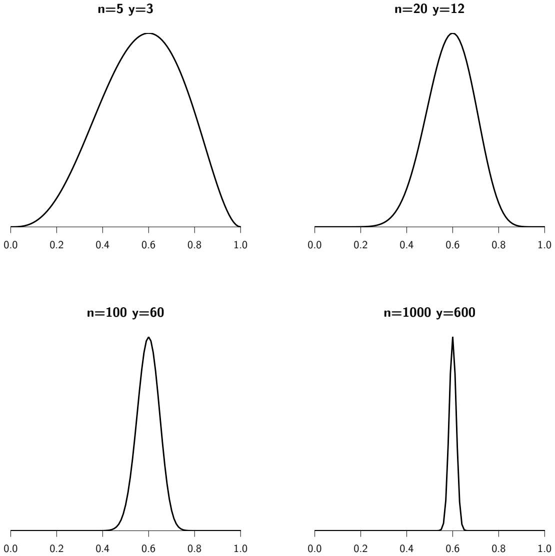
If we have a binomial random variable (e.g. logistic regression), and we the Beta distribution – whose support is $(0, 1)$ – as the prior on the success probability, then the posterior is also Beta.

Why is that helpful? It's a mathematical convenience. It is trivial to obtain means, modes, standard errors, quantiles, etc. from a familiar, closed-form probability distribution.

Still, non-conjugate priors usually present little difficulty today, since posteriors can be obtained via simulation algorithms.

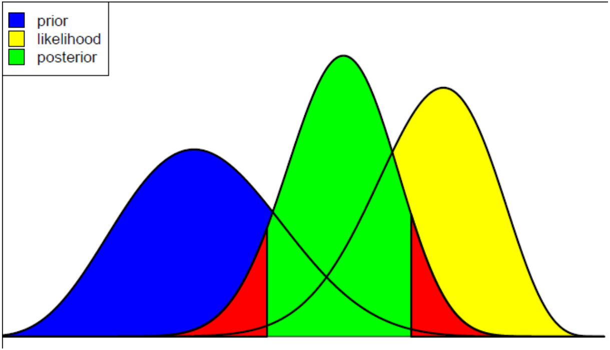
Assuming a conjugate prior as a rough model may often be of no more consequence than, say, assuming normality with moderate sample sizes. In fact, in many practical situations, the data may dominate the posterior (regardless whether the prior is conjugate).

Posterior range decreases with sample size



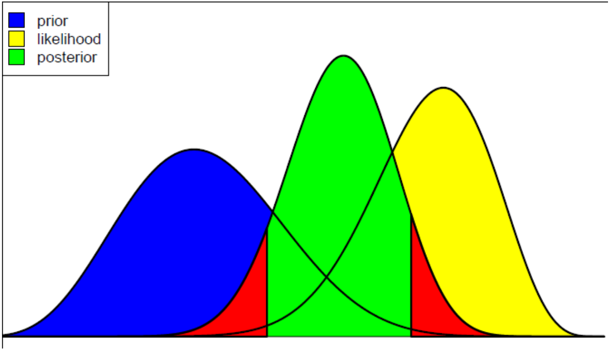
Posterior

"...the posterior distribution is centered at a point that represents a compromise between the prior information and the data, and the compromise is controlled to a greater extent by the data as the sample size increases." BDA p.37



Posterior

"...the posterior distribution is centered at a point that represents a compromise between the prior information and the data, and the compromise is controlled to a greater extent by the data as the sample size increases." BDA p.37



"Bayesian: One who, vaguely expecting a horse and catching a glimpse of a donkey, strongly concludes that he has seen a mule." Stephen Senn, Stat. Issues in Drug Development, 2nd Ed. p.46

But where do priors come from?

There's a whole sub-field of literture on *prior elicitation*.

- Population interpretation: the prior distribution represents a population of possible parameter values
- State of Knowledge interpretation: express our knowledge and uncertainty about θ as if its value could be thought of as a random realization from the prior distribution
- Past experiments, subject-domain knowledge, game theory trains of thought ...
- Sometimes there aren't any perfectly relevant populations to draw from, e.g. failure of an industrial process

The idea is that: Our data are generated from some random process. This random process depends on 1+ parameters. The value of these parameters are unknown to us, and can be viewed as being drawn from a probability distribution. E.g. essentially the same random process except for differences due to location (site), time of year, weather, etc.

The Bayesian approach forces one to think, understand the model

When you hear “linear regression model”, which do you envision?

Maybe, this?

```
proc reg;
  model y=x;
run;
```

Or maybe even this?

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Neither is a statistical model in and of itself, because they don't address probability or uncertainty.

Writing out models – Frequentist

Linear Regression

$y_i \sim N(\mu_i, \sigma^2)$

$\mu_i = \beta_0 + \beta_1 x_i$

Logistic Regression

$y_i \sim Bin(n, p_i)$

$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i$

Writing out models – Bayesian

Linear Regression

$y_i \sim N(\mu_i, \sigma^2)$

$\mu_i = \beta_0 + \beta_1 x_i$

$\beta_j \sim N(\mu_j, \sigma_j^2) \quad , \; j = 0, 1$

$\sigma^2 \sim InvGamma(\nu)$

Logistic Regression

$y_i \sim Bin(n, p_i)$

$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i$

$\beta_j \sim N(\mu_j, \sigma_j^2) \quad , \; j = 0, 1$

Probability as a measure of uncertainty (BDA 1.5)

Why the emphasis on probabilities?

- Probability is the fundamental measure of uncertainty
- Reasonable to quantify uncertainty in terms of probability: 'rain tomorrow', 'UGA wins national championship', 'coin toss = H', 'LTBI becomes active'
- After acquiring data: probability an unknown quantity lies with a certain range
- *Before* acquiring data: probability the mean of a sample lies within a certain range

Bayesian methods enable statements about partial knowledge concerning some situation or 'state of nature'

Probability as a measure of uncertainty (BDA 1.5)

Two justifications for numerical measures of uncertainty:

- An unknown, or unobservable (often physical) process
- Frequency obtained over a long sequence of repetitions

Consider both in terms of a coin toss.

Although a coin toss may be reasonably modeled either way, some situations are not easily nested under the frequency viewpoint: probability UGA wins; probability it rains; probability UGA wins given that it rains; probability the next shuttle launch will explode?

Probability as a measure of uncertainty (BDA 1.5)

Some reasons advanced in support of probability as a measure of uncertainty:

- Physical randomness induces uncertainty
- Statistical inference placed in terms of decision theory
- Coherence of bets – for what value of p would you exchange (i.e. bet) \$$p$ for a return of \$1?

Subjectivity vs. Objectivity

- Bayesian methods – especially those using informative priors – may sometimes thought to be by some ‘too’ subjective
- Actually, all scientific experiments carry a certain level of subjectivity
- Also, as pointed out above, a major part of proper Bayesian analysis is examining sensitivity to assumptions, including priors
- Bayesian analyses performed with weak or uninformative priors often results in nearly identical outcome measures as frequentist analysis

Strong examination of the Posterior

- Conclusions from frequentist analyses are often made from numbers representing a single aspect of the inferential model

$$Pr(Y|H_0)$$

  e.g. p-value, confidence interval, mean

- There are no generally-agreed upon similar measures in Bayesian analysis, although credible intervals (mentioned above), and Bayes Factors (mentioned below) have analogues in the frequentist approach
- Instead, the posterior is examined in-depth
- Especially when simulations are used to obtain the posterior, certain features may appear from the histogram representing the posterior that are difficult to detect from basic summary measures, e.g., heavy tails, multi-modality

Model selection, building, comparison

A nice intro to Bayes Factors

- The Bayesian paradigm carries a complete suite of tools for checking model accuracy, building/comparing models, etc.
- In particular, when we wish to chose between two competing hypotheses/models, we may use Bayes Factors

We wish to compare H_A vs. H_B, which we may initially express as

$$\frac{Pr(H_A)}{Pr(H_B)}$$

or, after observing data, as

$$\frac{Pr(H_A|y)}{Pr(H_B|y)} = \frac{Pr(y|H_A)}{Pr(y|H_B)} \times \frac{Pr(H_A)}{Pr(H_B)}$$

The middle term, the *Bayes Factor*, is multiplied against the prior odds to obtain the posterior odds, and represents the relative evidence in teh data.

Benefits of simulation-generated posteriors

I haven't covered the simulation methods for obtaining the posterior, because I considered this beyond the scope of this introductory talk. However, it is worth noting:

- Model complexity, large number of parameters don't pose the same difficulties they do with frequentist approaches
- It is straightforward to obtain the posterior probability distribution of complex functions of parameters, e.g.

$$\ln\left(\frac{\beta_1+\beta_5}{\beta_2^{\beta_3}}\right)$$

When might *you* favor Bayesian methods?

- When you have a prior and want to use it
- For models too complex to be reasonable under frequentist methods
- Heirarchical models (e.g. dose within person within site)

Summary

Bayesian methods ...

- Ask "what is the probability of my hypothesis given that I've observed these data" rather than "what is the probability of observing these data if my hypothesis is correct"; many find the former more intuitive
- Were once disregarded by mainstream science, but are now generally accepted
- Have seen great advancements in past 3 decades – theory, application, computational power

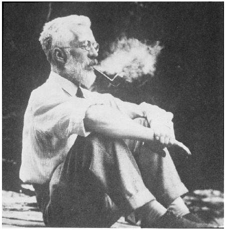
Summary (cont.)

Bayesian methods ...

- Generally require the analyst to engage in a higher level of scientific reasoning, to take a more profound look at underlying model assumptions and sensitivity of the conclusions to said assumptions.
- Generally require more forethought and more work, in particular in choosing priors; but a lot of the heavy lifting computations are still things for which you don't necessarily need to be familiar.
- Are an alternative to the frequentist approach; neither can be proven superior to the other; there exists more overlap between the two approaches than one might expect.

Thank You

Questions?



It might not look like it, but Fisher isn't pleased that you're considering Bayesian methods for your next research project.