The Bayesian Approach to Data Analysis

The Bayesian Approach to Data Analysis

Fridtjof Thomas, PhD

Associate Professor, Division of Biostatistics

TN-CTSI seminar on statistical reasoning
in biomedical research
https://tnctsi.uthsc.edu/

Additional Seminars in the Series:

- May 21st Multiple Testing and the False Discovery Rate (Saunak Sen, PhD)
- May 28th The Perfect Doctor: An introduction to Causal Inference (Fridtjof Thomas, PhD)
- June 4th Enhancing Statistical Methods in Grants and Papers (Saunak Sen, PhD)

Somehow…

- More data should give more information but there shouldn't be a "threshold" for "too little data"?
- Every observation gives some information – no?
- Learning from data is a gradual affair – no?
- Some data should not override everything else I believe – wouldn't you agree?
- Sequential learning/"updating" should be possible – no?

Bayesian Data Analysis…

What is it?

Why should I do it?

How can I do it?

Outline

- How can we learn from data?
- What are the elements in Bayesian modeling?
  - What is a posterior distribution?
  - What is a prior distribution?
  - What relates the posterior to the prior?
  - Does the posterior have to come after the prior?
- What makes a Bayesian analysis "Bayesian"?
- What are the advantages of Bayesian modeling?
- What are the difficulties in Bayesian modeling?

Schools of statistical thought

- Likelihood based approach
- Bayesian approach
- Fiducial approach
- Various *ad hoc* approaches

Schools of statistical thought

- Fiducial approach:

Fiducia (lat.) = trust/faith. Proposed by Fisher, "inverse probability without prior distributions". generally not coherent and "fiducial probabilities" lack the property of additivity, thus, are not a probability measure following Kolmogorov's axioms for probability measures. Mostly of historical interest.

- Various *ad hoc* approaches

E.g., "3+3 design" for dose escalation methods in Phase I clinical trials.

Schools of statistical thought (cont.)

Frequentist/Fisherian approach (often likelihood based)

Sir Ronald A. Fisher (1890 – 1962): English statistician, evolutionary biologist, and geneticist

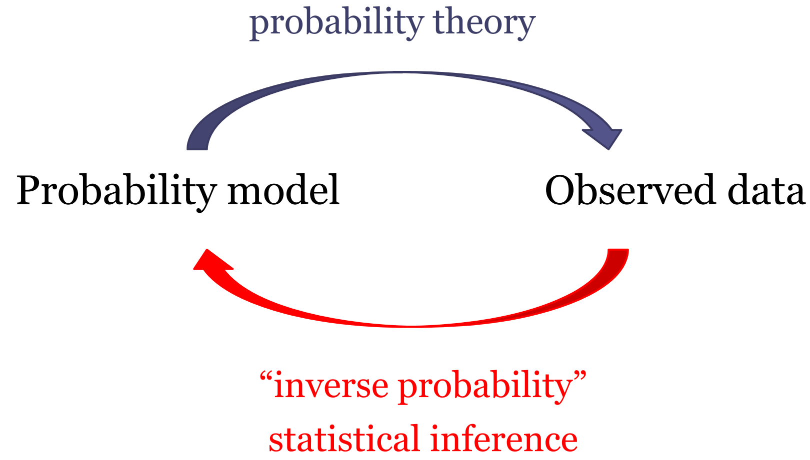
Probability: A frequency that results from an infinite sequence of independent repetitions of the same statistical experiment.

- Hypothesis testing (Fisher)
- Confidence intervals (Neyman-Pearson)

Bayesian

Probability: A "degree of belief".

Statistics: The basic problem



How can we learn from data?

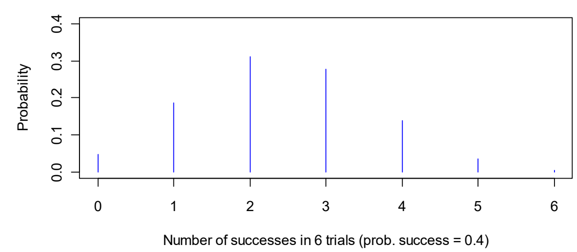
Let's do it!

"Likelihood"

Probability distribution: Probabilities that a certain value is observed given a parameter (or many...)

Example: Binomial distribution

$$P(s) = \binom{n}{s} \theta^s (1-\theta)^{n-s}$$



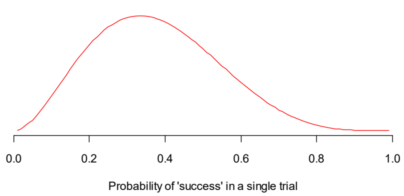
$$P(s) = \binom{6}{s} 0.4^s (1-0.4)^{6-s}$$

"Likelihood" (cont.)

$$P(s) = \binom{n}{s} \theta^{s} (1-\theta)^{n-s}$$

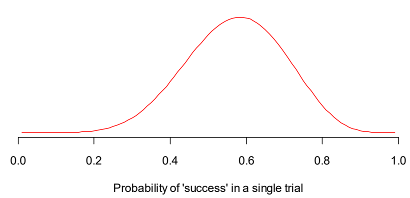
Observed s = 2 in n = 6 trials

$$\binom{6}{2} \theta^{2} (1-\theta)^{6-2}$$



Observed s = 7 in n = 12 trials

$$\binom{12}{7} \theta^{7} (1-\theta)^{12-7}$$

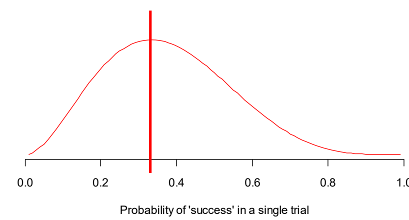


"Likelihood" (cont.)

$$P(s) = \binom{n}{s} \theta^{s} (1-\theta)^{n-s}$$

Observed s = 2 in n = 6 trials

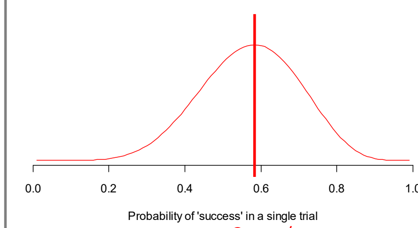
$$\binom{6}{2} \theta^{2} (1-\theta)^{6-2}$$



0.33 = 2/6

Observed s = 7 in n = 12 trials

$$\binom{12}{7} \theta^{7} (1-\theta)^{12-7}$$

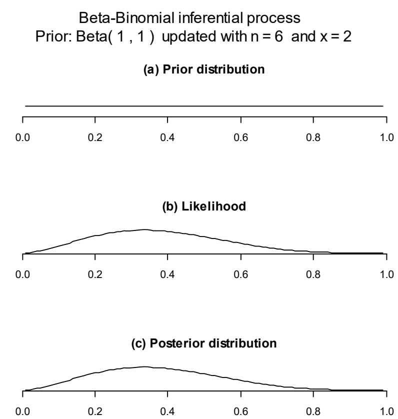


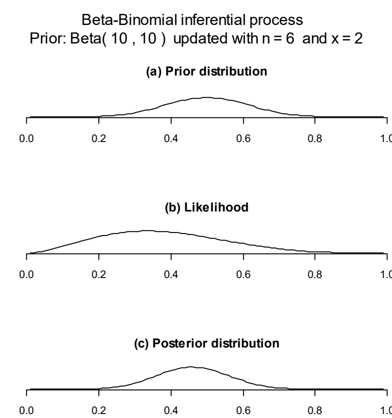
0.58 = 7/12

What are the elements in Bayesian modeling?

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

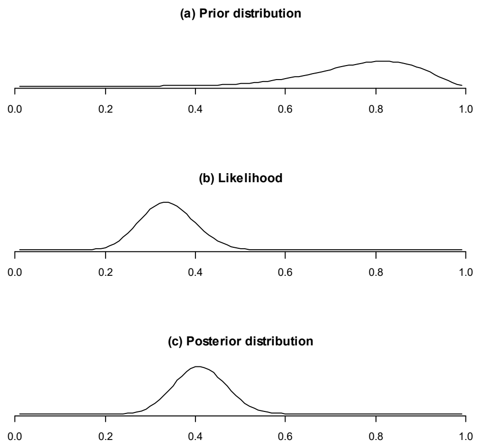
What are the elements in Bayesian modeling? (cont.)





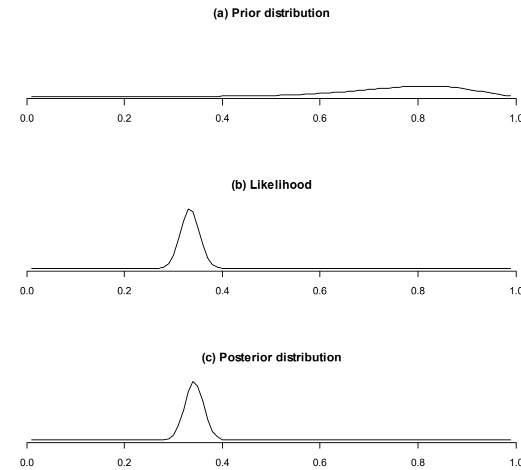
What are the elements in Bayesian modeling? (cont.)

Beta-Binomial inferential process
Prior: Beta(10 , 3) updated with n = 60 and x = 20



What are the elements in Bayesian modeling? (cont.)

Beta-Binomial inferential process
Prior: Beta(10 , 3) updated with n = 600 and x = 200



Bayes' theorem

Essay Towards Solving a Problem in the Doctrine of Chances (1764)

Published posthumously in the Philosophical Transactions of the Royal Society of London

Source: http://www.bayesian.org/

Thomas Bayes (c. 1702 – 1761)
Presbyterian minister
British mathematician

Born in London
Studied logic and theology in Edinburgh
Elected as a Fellow of the Royal Society in 1742

Bayes' theorem (cont.)

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

- $P(A|B)$: conditional probability
- $P(B|A)$: conditional probability
- $P(A)$: marginal probability
- $P(B)$: marginal probability

Bayes' theorem (cont.)

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

Example in diagnostic testing
(Spiegelhalter et al. (2004): Bayesian Approaches to Clinical Trials and Health-Care Evaluation)

Suppose new home HIV test claims to have "95% sensitivity and 98% specificity."
Test is to be used in a population with HIV prevalence of 1/1000.

"Event A": Person is HIV+
"Event B": Person *tests* positive (home HIV test)

$P(A) = {}^{1}\!/_{1000}$ (prevalence)

$P(B|A) =$ sensitivity of test $= 0.95$

$$P(B) = P(\underbrace{\text{positive test if HIV+}}_{\text{sensitivity of test}}) + P(\underbrace{\text{positive test if HIV-}}_{1\ -\ \text{specificity of test}})$$

$$= 0.95 \times 0.001 + (1 - 0.98) \times 0.999 = 0.02093$$

Bayes' theorem (cont.)

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

Q: What is the probability that a person who *tests* positive actually *is* HIV+? (Positive predictive value)

$$P(A|B) = \frac{0.95 \times 0.001}{0.02093} = 0.045$$

A: - Less than 5 out of 100 who test positive are HIV+.
- Over 95% of those who test positive are in fact *not* HIV+.

(End of example.)

What are the elements in Bayesian modeling? (Cont.)

A somewhat wider interpretation of Bayes' theorem:

$$p(b|a) = \frac{p(a|b)}{p(a)}p(b)$$

$$p(\text{Model}|\text{Data}) = \frac{p(\text{Data}|\text{Model})}{p(\text{Data})}p(\text{Model})$$

What are the elements in Bayesian modeling? (Cont.)

$$p(\text{Model}|\text{Data}) = \frac{p(\text{Data}|\text{Model})}{p(\text{Data})}p(\text{Model})$$

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

"Model" typically means parameters | model form

Examples: - unknown mean and stddev in a Normal distribution
- unknown intensity parameter in the Poisson distribution

What are the elements in Bayesian modeling? (Cont.)

$$p(\theta \mid y) = \frac{p(y \mid \theta)}{p(y)} p(\theta) \propto p(y \mid \theta) p(\theta)$$

$$p(y) = \begin{cases} \sum_{\Theta} p(y \mid \theta) p(\theta) \\ \int_{\Theta} p(y \mid \theta) p(\theta)\, d\theta \end{cases}$$

What are the elements in Bayesian modeling? (Cont.)

$$p(\theta \mid y) \propto p(y \mid \theta) p(\theta)$$

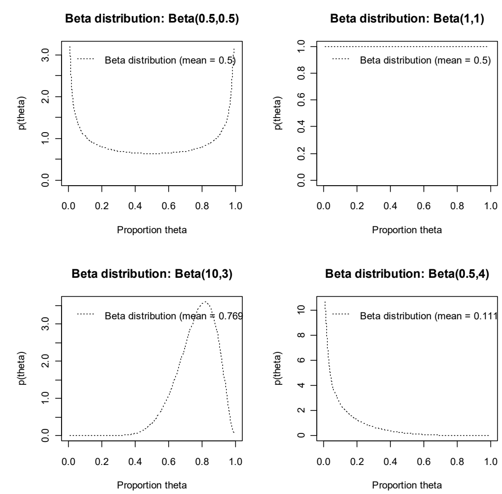
$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Thumbtack example

Prior: Beta distribution

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Thumbtack example (cont.)



Thumbtack example (cont.)

Likelihood: Binomial

$$p(\#success \mid \theta) = \binom{n}{s} \theta^{s} (1-\theta)^{n-s}$$

Thumbtack example (cont.)

posterior $\propto$ likelihood $\times$ prior

$$p(\theta \mid \#success) \propto p(\#success \mid \theta)\, p(\theta)$$

$$= \binom{n}{s} \theta^{s} (1-\theta)^{n-s} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$\propto \theta^{s} (1-\theta)^{n-s} \times \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$= \theta^{s+(\alpha-1)} (1-\theta)^{n-s+(\beta-1)}$$

$$= \theta^{\alpha+s-1} (1-\theta)^{\beta+n-s-1}$$

Thumbtack example (cont.)

$$p(\theta \mid \#success) \propto p(\#success \mid \theta)\, p(\theta)$$

$$\propto \theta^{\overbrace{\alpha+s}^{\tau}-1} (1-\theta)^{\overbrace{\beta+n-s}^{\eta}-1}$$

$$\propto \theta^{\tau-1} (1-\theta)^{\eta-1} \qquad \frac{\Gamma(\tau+\eta)}{\Gamma(\tau)\Gamma(\eta)} \theta^{\tau-1} (1-\theta)^{\eta-1}$$

Conclusion: $p(\theta \mid \#success)$ is Beta(τ, η)

Beta$(\tau = \alpha + s,\ \eta = \beta + n - s)$

Thumbtack example (cont.)

Posterior distribution: Beta$(\tau = \alpha + s,\ \eta = \beta + n - s)$

This distribution is the answer to our inference problem about the (still) unknown probability of success.

We can (and should) summarize this posterior distribution in any way that is meaningful to our problem at hand.

Summaries include (but are not limited) to:
- Point estimates such as means, medians, modes, quartiles
- Intervals such as Highest Posterior Density (HPD) interval or intervals with equal tail probabilities

Thumbtack example (cont.)

Take notice!

Once we have the posterior distribution, we can readily predict what we are likely to see in future experiments!

Such a predictive distribution reflects the remaining uncertainty about the probability of "success".

In sharp contrast:
"Traditional" or "frequentist" confidence intervals cannot be interpreted as probability intervals and the predictive distribution for future experiments remains unclear.
The unknown parameter remains a fixed but unknown quantity that does not have a probability distribution.

Thumbtack example (cont.)

Predictive distribution
(*s* is the number of "successes" to be observed):

$$\int_0^1 \text{Bin}(s\,|\,\theta,n)\,\text{Beta}(\alpha,\beta)\,d\theta = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n)}\binom{n}{s}\Gamma(\alpha+s)\Gamma(\beta+n-s)$$

$$E[s] = n\frac{\alpha}{\alpha+\beta} \qquad V[s] = \frac{n\alpha\beta}{(\alpha+\beta)^2}\frac{(\alpha+\beta+n)}{(\alpha+\beta+1)}$$

Mode: greatest integer that does not exceed $s_m = \dfrac{(n+1)(\alpha-1)}{(\alpha+\beta-1)}$

(If s_m is an integer, s_m and $s_m - 1$ are both modes.)

Special case: $\alpha = \beta = 1$ gives discrete uniform with mass $(n+1)^{-1}$ for each $s = 0, 1, \ldots, n$.

What is a posterior distribution?

- The result of a mathematical computation.
- The "vehicle" that contains all the information about our parameters of interest (conditional on the model form)

"The answer is the answer." (Adrian F.M. Smith)

What is a prior distribution?

- A summary of what we know about a quantity of interest before we conduct an experiment.
- A summary of what we do *not* know about a quantity of interest before we conduct an experiment.
- A necessary part in our inferential procedure.
- Part of one side of a relationship that must be satisfied.

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

What relates the posterior to the prior?

- The likelihood
- Probability calculus (mathematics)

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Does the posterior have to come "after" the prior? (In time?)

- No

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Our example revisited

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Beta Binomial **Beta**

This property makes life easy!

Conjugate prior

experiment ↓ experiment ↓

prior ⟶ posterior/prior ⟶ posterior

Sequential updating/learning

What makes a Bayesian analysis "Bayesian"?

- Bayes theorem is used by "non-Bayesians" as well.
- Interpretation of what is meaningfully put into this theorem, especially: parameters.
- Bayesian inference: in parameter space (not sample space)

In sharp contrast:
"Probability" in the "traditional" or "frequentist" school of statistics always refers to quantities that can be thought of as resulting from repeated trials/experiments (the frequentist definition of probability). Since parameters are interpreted as having a fixed (but unknown) value it is not meaningful to operate on them using Bayes' theorem.

What are the advantages of Bayesian modeling?

- Flexibility in modeling.
- Predictive distributions for future use (new experiments) is readily available.
- A theory of support for scientific hypotheses (instead of a theory for evidence against scientific hypotheses).
- Inference results lend themselves readily to a decision theoretic analysis.
- Explicit formulation of prior information: greater transparency.

What are the difficulties in Bayesian modeling?

- Prior distributions:
  - Conjugate priors are restrictive
  - General priors come with difficult mathematical/numerical problems to solve
  - Informative and influential
  - Elicitation of prior information
  - “Non-informative” priors/reference priors
  - Consensus priors
- Flexible models are models less well understood
- Greater need for sensitivity analysis
- Computations: Markov Chain Monte Carlo (MCMC) a break through but problems remain

Extensions

- Hierarchical models
- Non-nested models
- Missing data problems
- Model selection
- Model averaging
- Bayesian design of experiments
- …

A Bayesian reading list

Jeffreys, H. (1961). *Theory of Probability (Third ed.)*. Oxford: Clarendon Press. (First published in 1939)

Savage, L. J. (1972). *The Foundations of Statistics (Second ed.)*. New York: Dover.

De Finetti, B. (1974). *Theory of Probability*. New York: Wiley.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. New York: Wiley.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis (Second ed.)*. New York: Springer.

A Bayesian reading list (cont.)

Jeffrey, R. C. (1983). *The Logic of Decision (Second ed.)*. Chicago: University of Chicago Press.

Kaplan, M. (1996). *Decision Theory as Philosophy*. Cambridge: Cambridge University Press.

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. Third ed. Boca Raton: Chapman & Hall/CRC; 2014.

Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester: Wiley.

Lesaffre, E., and Lawson, A. B. (2012). *Bayesian Biostatistics*. Chichester: Wiley.

O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics Vol. 2B: Bayesian Inference*. London: Edward Arnold.

Dey, D. K., & Rao, C. R. (Eds.). (2005). *Handbook of Statistics 25 - Bayesian Thinking: Modeling and Computation*. Amsterdam: Elsevier.

Summary

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Thank you!

Additional Seminars in the Series:

- May 21st Multiple Testing and the False Discovery Rate (Saunak Sen, PhD)
- May 28th The Perfect Doctor: An introduction to Causal Inference (Fridtjof Thomas, PhD)
- June 4th Enhancing Statistical Methods in Grants and Papers (Saunak Sen, PhD)

BERD Clinics: https://tnctsi.uthsc.edu/