Introduction

Bayesian Methodology in Biostatistics (BST 249)

Jeffrey W. Miller

Department of Biostatistics
Harvard T.H. Chan School of Public Health

Outline

Applications of Bayesian statistics

What is Bayesian statistics?

Difference between Bayesian and frequentist perspectives

Outline

Applications of Bayesian statistics

What is Bayesian statistics?

Difference between Bayesian and frequentist perspectives

Applications of Bayesian statistics

- Bayesian statistics is used in many real-world applications.
- Below are a few interesting examples.
- One could just as easily come up with a list of applications where non-Bayesian methods are used, but the focus here is on Bayesian statistics.

Tracking and guidance

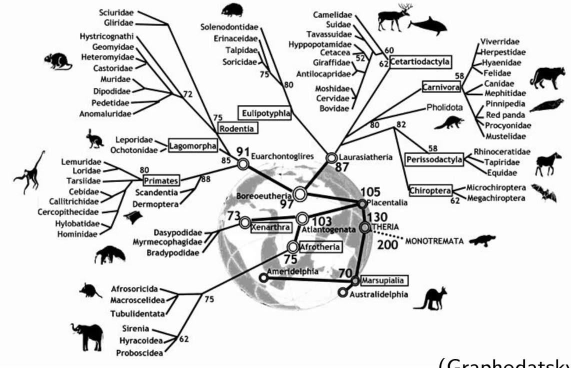
SpaceX rockets

Google self-driving car

- The Kalman filter (KF) is a Bayesian time-series model.
- KFs are widely used in guidance systems of aircraft, spacecraft, and robots.
- The KF models the evolution of vehicle pose (location, velocity, acceleration) using a physical model.
- Measurements from sensors such as accelerometers, GPS, compass, gyroscope are modeled given pose.
- The Bayesian posterior is used to infer pose given sensors.

Phylogenetics

Inferred phylogenetic tree of mammals

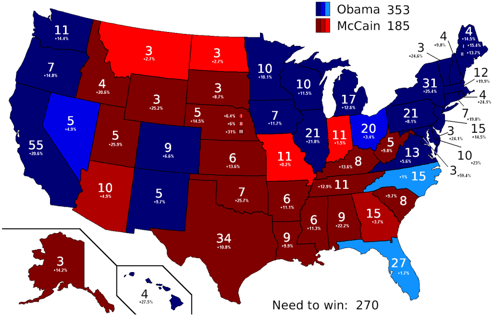


(Graphodatsky et al. 2011)

- Evolution is fundamental in nearly all of biological research.
- Researchers use statistical models to infer the evolutionary "family tree" (a.k.a. phylogeny) of species, given genetic data.
- Some of the most common methods use Bayesian models.

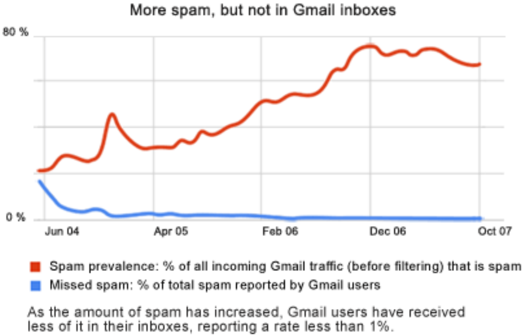
Political science

Nate Silver's predictions for the 2008 presidential election



- Bayesian models are often used for election predictions.
- Nate Silver is reknowned for accurate election predictions.
- Although the details are secret, it appears that Nate Silver uses Bayesian hierarchical regression models.
- These models adjust for systematic biases in polling data.

Computer science

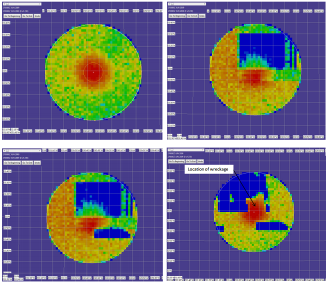


- Email providers use sophisticated spam filtering algorithms.
- Bayesian models are the most prominent methods for spam detection and filtering.
- A Google researcher who used to work at Microsoft said: *"Google uses Bayesian filtering the way Microsoft uses the if statement."*

(`http://www.joelonsoftware.com/items/2005/10/17.html`)

Search and rescue

Prior and posteriors for the location of Air France 447 wreckage



- On June 1, 2009, Air France Flight 447 crashed into the Atlantic Ocean. Even after three intensive searches, the wreckage had not been found a year later.
- French authorities then commissioned a Bayesian search analysis, and AF 447 was finally found.
- This method has also been used to find ships lost at sea.

(Stone et al., 2011)

Radiocarbon dating

- Radiocarbon dating is often used to infer the age of archeological objects.
- Often, contextual information is available as well, such as other objects found nearby, where and how deep the object was found, and other dating techniques.
- Bayesian methods are often used to integrate multiple sources of information when dating objects (Ramsey, 2009).

Outline

Applications of Bayesian statistics

What is Bayesian statistics?

Difference between Bayesian and frequentist perspectives

Bayes, in a nutshell

- The Bayesian approach can be summarized as follows:
    1. assume a probability distribution on any unknowns (this is the prior),
    2. assume the distribution of the knowns given the unknowns (this is the likelihood),
    3. then just follow the rules of probability to answer any questions of interest.

- An overarching theme of the Bayesian perspective is that uncertainty is quantified with probability distributions.

- We will formalize these concepts shortly.

Bayesian bubble of knowledge

- Fully Bayesian approaches put distributions on everything.
- This creates a self-contained "bubble of knowledge".
- Any modeled quantity within this bubble can be inferred probabilistically.
- Pro: Coherent framework for inference in complex models.
- Pro: Can exploit prior knowledge.
- Con: May be problematic if part of the model is wrong.
- Con: Excludes frequentist criteria for evaluating performance.

Computation in Bayesian models

- Answering a question of interest usually requires computing some posterior expectation.
- Three main categories of methods:
  - exact solution
  - deterministic approximation
  - stochastic approximation

Computation: Exact solution

- In certain cases, it is computationally feasible to compute the posterior (and posterior expectations) exactly.

- Examples where exact solutions are often possible:
  - Exponential families with conjugate priors
  - Multivariate Gaussian models
  - Dynamic programming for some classes of graphical models

Computation: Deterministic approximation

- Methods include:
  - numerical integration, a.k.a. quadrature/cubature
  - quasi-Monte Carlo (QMC), low discrepancy sequences
  - Laplace approximation
  - variational inference (VI), expectation propagation
- For low-dimensional integrals, numerical integration and QMC are superior to stochastic approximations.
- VI can work well in high-dimensions, but sometimes has insufficient accuracy.

Computation: Stochastic approximation

- For high-dimensional integrals, stochastic approximation is often the only option.

- Basic idea: Use samples from the posterior to approximate posterior expectations.

- Methods include:
  - Monte Carlo approximation, Importance sampling
  - Markov chain Monte Carlo (MCMC) — Gibbs sampling, Metropolis–Hastings, slice sampling, Hamiltonian MC
  - sequential importance sampling, sequential Monte Carlo, population Monte Carlo
  - approximate Bayesian computation (ABC)

Outline

Bayesian and frequentist perspectives

- Let x = observed data, let θ = unknowns.
- Essential difference:
  - The Bayesian considers only the observed value of x, and treats θ as random.
  - The frequentist considers all possible values of x, and treats θ as fixed.
- In decision-theoretic terms:
  - Bayesian: Make best possible decision given observations x, allowing for uncertainty in θ.
  - Frequentist: Use a procedure with guaranteed performance when used repeatedly, no matter what θ turns out to be.

Group activity: Check your understanding

Go to breakout rooms and work together to answer these questions:
https://forms.gle/NMnNDSX7Ax5SvJtr8

(Two people per room, randomly assigned. 15 minutes.)

Example: Diagnosing celiac disease

- Around 1 in 100 people is affected by celiac disease, resulting in sensitivity to gluten.
- Celiac is commonly diagnosed by measuring the level x of a certain antibody in the blood.
- If x is above a cutoff point, then a positive diagnosis is made (i.e., disease is present), otherwise, negative diagnosis is made.
- (This description is simplified a bit, for illustration purposes.)

Example: Diagnosing celiac disease

- How should the cutoff c be chosen?
  - If c is too low, there will be too many false positives.
  - If c is too high, there will be too many false negatives.

- We have the following hypothesis testing problem.

  H_0: $\theta = \theta_0$ (No disease)
  H_1: $\theta = \theta_1$ (Disease)

- Suppose the distributions $p(x|\theta_0)$ and $p(x|\theta_1)$ are known, based on many previous cases.

Celiac example: Bayesian approach

(Whiteboard activity)

- Prior: Since 1 in 100 people is affected, we have

$$p(\theta_1) = ??? \text{ and } p(\theta_0) = ???$$

- Posterior: By Bayes' theorem,

$$p(\theta_1 | x) = ???$$

- The diagnosis ($a = \theta_0$ or $a = \theta_1$) is made to minimize posterior expected loss,

$$\mathrm{E}(\ell(\theta, a) \mid x) = ???$$

where $\ell(\theta, a) =$ loss incurred by picking a when the truth is θ.

Celiac example: Bayesian approach

- Prior: Since 1 in 100 people is affected, we have

$$p(\theta_1) = 1/100 \text{ and } p(\theta_0) = 99/100.$$

- Posterior: By Bayes' theorem,

$$p(\theta_1|x) = \frac{p(x|\theta_1)p(\theta_1)}{p(x|\theta_0)p(\theta_0) + p(x|\theta_1)p(\theta_1)}.$$

- The diagnosis ($a = \theta_0$ or $a = \theta_1$) is made to minimize posterior expected loss,

$$\mathrm{E}(\ell(\theta, a) \mid x) = \ell(\theta_0, a)p(\theta_0|x) + \ell(\theta_1, a)p(\theta_1|x)$$

where $\ell(\theta, a) =$ loss incurred by picking a when the truth is θ.

Celiac example: Frequentist approach

- The standard frequentist approach is to minimize false negatives subject to a bound on false positives, say, $\alpha = 0.05$.
- By the Neyman–Pearson lemma, this is achieved by choosing $a = \theta_1$ when
$$\frac{p(x|\theta_1)}{p(x|\theta_0)} > \lambda$$
and $a = \theta_0$ otherwise, where $\lambda \geq 0$ is chosen so that the probability of a false positive equals α, i.e.,
$$\mathbb{P}(X \in R_\lambda \mid \theta_0) = \int_{R_\lambda} p(x|\theta_0) dx = \alpha$$
where $R_\lambda = \big\{x : p(x|\theta_1)/p(x|\theta_0) > \lambda\big\}$.

Celiac example: Comparing Bayesian & frequentist

- Bayesian:
  - the unknown θ is treated as a random variable
  - we only consider the observed value of x
- Frequentist:
  - θ is unknown but fixed (non-random)
  - the choice of λ depends on all possible values of x

- Each approach satisfies an optimality criterion:
  - Bayesian is optimal for the assumed prior and loss.
  - Frequentist is optimal for the assumed bound on false positives.

- For binary decisions such as this, it turns out that the two approaches are equivalent:
  - For any prior and loss, there is a λ for which the Bayesian and frequentist procedures coincide, and vice versa.

Frequentist evaluation of Bayesian procedures

- From a purely Bayesian perspective, if the prior and likelihood are chosen properly, then the resulting inferences are correct and optimal, and there is nothing more to be said.

- However, in practice this is not very satisfying, due to:
  - uncertainty about the choice of likelihood or prior, and
  - approximations used for computational reasons.

- The frequentist perspective provides tools to deal with these issues.
  - Empirical tools: cross-validation, test sets, bootstrap, goodness-of-fit tests.
  - Theoretical tools: consistency, rates of convergence, and calibration/coverage.

Overall recommendation: be pragmatic, not dogmatic

- Be pragmatic—that is, use what has been shown to work.
- As a default approach, the following maxim will serve you well:

  Design as a Bayesian, and evaluate as a frequentist.

- In other words, construct models and procedures starting from a Bayesian perspective, and use frequentist tools to evaluate their performance.

References and supplements

- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82(1), 35-45.
- Graphodatsky, A. S., Trifonov, V. A., & Stanyon, R. (2011). The genome diversity and karyotype evolution of mammals. Mol Cytogenet, 4(1), 22.
- Stone, L. D., Keller, C. M., Kratzke, T. M., & Strumpfer, J. P. (2011). Search analysis for the underwater wreckage of Air France Flight 447. In Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on (pp. 1-8). IEEE.
- Ramsey, C. B. (2009). Bayesian analysis of radiocarbon dates. Radiocarbon, 51(1), 337-360.
- Kass, R. E. (2011). Statistical inference: The big picture. Statistical Science, 26(1), 1. `http://projecteuclid.org/euclid.ss/1307626554`
- Jordan, M. I. Are You a Bayesian or a Frequentist? (2009) `http://videolectures.net/mlss09uk_jordan_bfway/`

Your to-do items

- Readings for this week (see syllabus).
  (Note: LN A is on Canvas at
  `Files/Lecture slides/A-Probability-and-Linear-algebra-basics.pdf`)
- Homework #1 is due next Thursday.
- Talk to classmates to form your project groups. (It's easiest to have the same groups for the case study and project.)

Individual activity: Exit ticket survey

Answer these questions individually:

https://forms.gle/1biGyHcZYnNcXsCm8