STA 365: Applied Bayesian Statistics

Boris Babic
Assistant Professor, University of Toronto

Week 1A: Introduction

Introduction

- Boris Babic, Assistant Professor
- Research interests: Bayesian models and theory; law and ethics of machine learning and AI
- E-mail: boris.babic@utoronto.ca
- Office: 700 Univ Rm 9086 & JHB Rm 433
- Office hour: Mondays 2-3pm.
- Class time: Wednesday 3-5pm, Friday 3-4pm.
- Classroom: KP 108
- TAs: Colin Decker, Morris Greenberg (e-mails on syllabus)

Objectives

- Understand foundations of the Bayesian approach and how it differs from classical statistics.
- Understand methods of inference using the Bayesian approach.
- Develop Bayesian models for data analysis.
- Understand computational Bayesian approaches.

Prerequisities

- Adequate background in basic probability and statistics, calculus, linear algebra.
- Familiarity with R
- R, `https://cran.r-project.org`
- We will make use of JAGS and/or Stan: `http://mcmc-jags.sourceforge.net`, `http://mc-stan.org`.
- Reference books (not required): Gelman et al., Bayesian Data Analysis; Hoff, A First Course in Bayesian Statistics; and Robert, The Bayesian Choice: From Decision Theoretic Foundations to Computational Implementation
- The highlighted text above would be a good summary of this course!

Course requirements

- 2 Homeworks (20% each) (due dates will be posted on the course website)
- 2 Assignments (30% each) (due dates will be posted on the course website)
- Regular class participation will be essential to completing the assignments and homeworks successfully.

Topics

- Subjective Probability and Betting, Bayes' Rule
- Exchangeability
- Single Parameter Inference Estimator Optimality
- Hypothesis Tests
- The Normal Model
- The Multivariate Normal Model
- Hierarchical Models
- Mixture Models
- Linear Regression
- Generalized Linear Models Mixed Effects
- Missing Data
- Monte Carlo Approximations
- Markov chain and Hamiltonian Monte Carlo
- Introductions to JAGS and STAN

Bayesian statistics: life changing and life saving

The New York Times

Bayesian statistics can help solve the Monty Hall problem of winning a car. Fred Westbrook Collection/GSN

By F. D. Flam
Sept. 29, 2014

Statistics may not sound like the most heroic of pursuits. But if not for statisticians, a Long Island fisherman might have died in the Atlantic Ocean after falling off his boat early one morning last summer.

The man owes his life to a once obscure field known as Bayesian statistics — a set of mathematical rules for using new data to continuously update beliefs or existing knowledge.

The method was invented in the 18th century by an English Presbyterian minister named Thomas Bayes — by some accounts to calculate the probability of God's existence. In this century, Bayesian statistics has grown vastly more useful because of the kind of advanced computing power that did not exist even 20 years ago.

Bayesian statistics: life changing and life saving

Bayesian statistics are rippling through everything from physics to cancer research, ecology to psychology. Enthusiasts say they are allowing scientists to solve problems that would have been considered impossible just 20 years ago. And lately, they have been thrust into an intense debate over the reliability of research results. – New York Times (Sept 29, 2014)

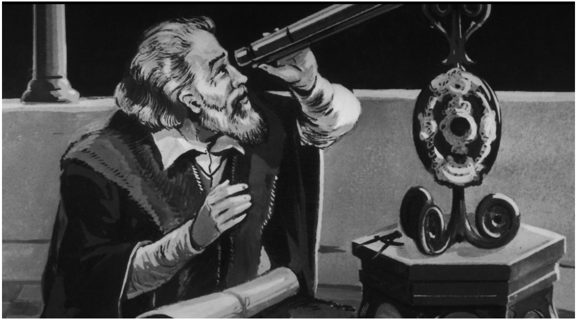
An Example

- Suppose that $X_1, ...X_n \stackrel{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$ with σ^2 known.
- We are interested in learning about μ.
- Construct a 95% confidence interval for μ:

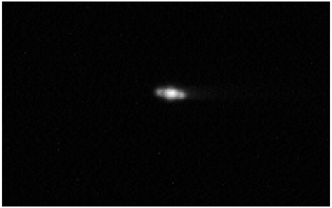
$$\overline{X} - 1.96\frac{\sigma}{n} < \mu < \overline{X} + 1.96\frac{\sigma}{n}.$$

- What is the random variable in this expression?
- What is $f(\mu)$? Or $\Pr(\mu = k)$? Or $\Pr(a < \mu < b)$?
- $\Pr(\overline{X} - 1.96\frac{\sigma}{n} < \mu < \overline{X} + 1.96\frac{\sigma}{n}) = 95\%$ is a statement about $\mathbf{X}$; in particular, it is a statement about how you think the X's will behave across many future samples, and which proportion of those samples will cover the parameter μ.
- But in frequentist statistics, there is no such thing as a probability regarding μ; or a confidence interval about which μ values you deem most likely.
- In frequentist statistics, we evaluate the quality of procedures in terms of certain long run guarantees.
- For example, bias, which is computed with respect to the sampling distribution.
- Or, consistency, which is an asymptotic concept.

Galileo's telescope



When Galileo first observed Saturn through a telescope, it looked something like this:



How confident are you that those are rings around Saturn?

Unique events

What is the probability of naval conflict in the South China Sea?

Unique events

What is the probability of US dollar collapse in the next 12 months?

Unique events

What is the probability of war in Ukraine?

Probability: Mathematical Axioms

$p = p(E)$ is defined on some set S such that

(1) $p(E) \geq 0, \ \ \forall E \subset S$

(2) If $E_1 \cap E_2 = \emptyset$ then $p(E_1 \cup E_2) = p(E_1) + p(E_2)$

(2*) If $E_1, E_2, ..., E_n$ are disjoint, then

$$p(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} p(E_i)$$

(3) $p(S) = 1$

Interpretations

- Classical, based on the notion of symmetry:

$$p(E) = \frac{\text{\# of outcomes in E}}{\text{Total \# of outcomes in S}}$$

- Relative frequency, based on the idea of identical repeated trials:

$$p(E) = \frac{\text{\# of times E occurs}}{\text{Total \# of trials}}$$

- Bayesian, based on the notion of subjective (epistemic) uncertainty

Betting

Lottery A: $\begin{cases} \text{win } \$X, \text{ if } E \\ \text{win } \$0, \text{ if } \bar{E} \end{cases}$

Lottery B: $\begin{cases} \text{win } \$X, \text{ with } p \\ \text{win } \$0, \text{ with } 1-p \end{cases}$

canonical prob'y

$P(E) = p$ that makes you indifferent between A + B.

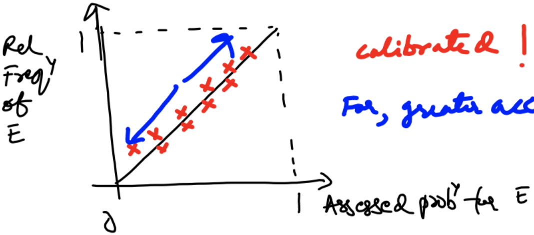
Dutch Books: $p(E) \geq 0$

	Buyer	Seller
E	$(1-p)S$	$-(1-p)S$
$\bar{E}$	$-pS$	pS

$$\text{lottery}: \begin{cases} \$1, & E \\ \$0, & \bar{E} \end{cases}$$

- Seller assesses $p(E) = p$
- Buyer decides on S, the number of lottery tickets to buy or sell at p
- Suppose seller assesses $p(E) < 0$
- Then buyer is able to create a dutch book argument against the seller (i.e., the seller will lose money whether or not E occurs)

Calibration



calibrated!

For, greater accuracy!

Incentive schemes & evaluation measures for p (scoring rules)

r = assessor's <u>stated</u> prob' for E

p = assessor's "<u>real</u>" prob' for E

$$s(r) = \begin{cases} s_1(r), & E \\ s_2(r), & \bar{E} \end{cases}$$

$$E_p[s(r)] = p\, s_1(r) + (1-p)\, s_2(r)$$

$$E_p[s(p)] = p\, s_1(p) + (1-p)\, s_2(p)$$

honest reporting !

$$E_p[s(p)] > E_p[s(r)], \ \forall r \neq p$$

strictly proper scoring rule !

Incentive schemes & evaluation measures for p (scoring rules)

<u>Linear Scoring Rule:</u>

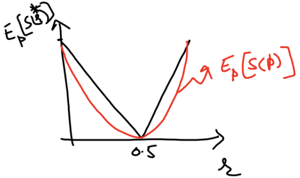
$$s(r) = \begin{cases} r, & \text{if } E \\ 1-r, & \text{if } \bar{E} \end{cases}$$

$$E_p[s(r)] = pr + (1-p)(1-r)$$
$$= (2p-1)r + (1-p)$$

$$E_p[S(p)] = p^2 + (1-p)^2$$

optimal $r = r^*$ that maximizes $E_p[S(r)]$

$$r^* = \begin{cases} 0, & p \leq 0.5 \\ 1, & p \geq 0.5 \end{cases}$$



Incentive schemes & evaluation measures for p (scoring rules)

Quadratic Scoring Rule

$$s(x) = \begin{cases} s_1(x) = -(1-x)^2, & E \\ s_2(x) = -x^2, & \bar{E} \end{cases}$$

could we $1-(1-x)^2$, $1-x^2$ to make it +ve !

$$E_p[s(x)] = -p(1-x)^2 - (1-p)x^2$$

$$= -p - \cancel{px^2} + 2px - x^2 + \cancel{px^2}$$

$$= -p + p^2 - p^2 + 2px - x^2$$

$$= \underbrace{-p(1-p)}_{\text{penalty for lack of sharpness (always)}} - \underbrace{(p-x)^2}_{\text{penalty for dishonesty! penalty for lack of calibration ex post!}}$$

What is a good subjective probability?

- Coherence: follow the 3 mathematical axioms of probability
- Calibration: relative frequency of E should approach p in the long run
- Sharpness: probability should be close to 0 or 1 without losing calibration.

Conditional Probability

Def^n: If $P(I) > 0$, then

$$P(E|I) = \frac{P(E \cap I)}{P(I)}$$

$$\text{Lottery}_1: \begin{cases} \$1, & I \\ \$0, & \bar{I} \end{cases} \qquad \text{Lottery}_2: \begin{cases} \$1, & E \cap I \\ \$0, & \overline{E \cap I} \end{cases}$$

$$\text{Lottery}_3: \begin{cases} \$1, & E|I \\ \$0, & \bar{E}|I \end{cases}$$

Assessor provides $p_1 = P(I)$, $p_2 = P(E \cap I)$, $p_3 = P(E|I)$

Conditional Probability

Buyer decides S_1, S_2, & S_3

If $E \cap I$, $G_1 = (1-p_1)S_1 + (1-p_2)S_2 + (1-p_3)S_3$

If I, but not E, $G_2 = (1-p_1)S_1 - p_2 S_2 - p_3 S_3$

If $\bar{I}$, $G_3 = -p_1 S_1 - p_2 S_2 + 0$

can create a dutch book, unless

$$P(E|I) = \frac{P(E \cap I)}{P(I)}$$

Bayes' Theorem

$$\Pr(E|I) = \frac{\Pr(I|E)\Pr(E)}{\Pr(I|E)\Pr(E) + \Pr(I|\overline{E})\Pr(\overline{E})}$$

- Prior: $\Pr(E), \Pr(\overline{E})$
- Likelihoods: $\Pr(I|E), \Pr(I|\overline{E})$
- Posteriors: $\Pr(E|I), \Pr(\overline{E}|I)$

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

Bayes' Theorem

$$\Pr(E|I) = \frac{\Pr(I|E)\Pr(E)}{\Pr(I)}$$

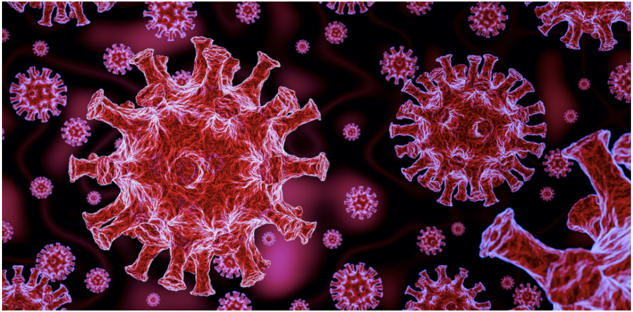
- Simple example: Pr of die landing on 4, conditional on it landing on an even number.
- $\Pr(4|even) = \frac{\Pr(even|4)\Pr(4)}{\Pr(even)}$
- $\Pr(4|even) = \frac{1 \times 1/6}{1/2} = 1/6 \times 1/2 = 1/3$

Total probability

Law of total probability:

$$\mathrm{P}(E) = \sum_{j=1}^{n} \mathrm{P}(E|I_j)\mathrm{P}(I_j)$$

Medical Screening



- Suppose that the base rate (prevalence) of COVID-19 is 1% (ha!).
- Suppose PCR tests have 99% sensitivity. That is, $p(+|C) = 0.99$
- Suppose they also have specificity of 95%. That is, $p(-|\overline{C}) = 0.95$
- Find $p(C|+)$.