Bayes Rule and Bayesian Probability

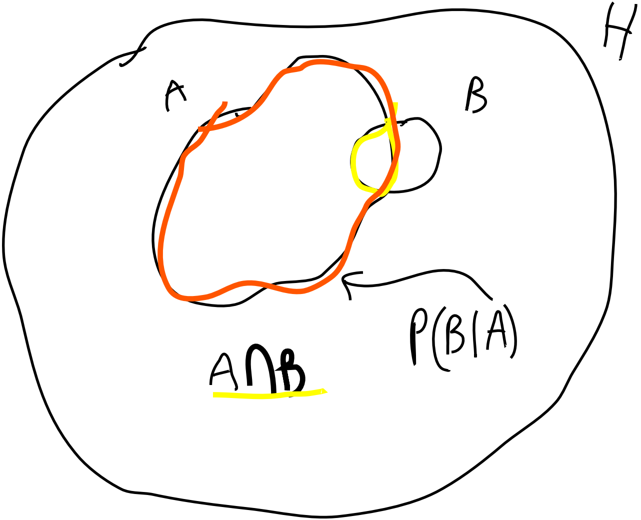
PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

01/20/2021

Basic probability

- Think of "Probability" as a function that assigns a real number to an *event*
- For events E and F in probability space H, a *probability function* P satisfies
  - P1 $0 \leq P(E) \leq 1$ for all E.
  - P2 $P(H) = 1$.
  - P3 $P(E \cup F) = P(E) + P(F)$ if $E \cap F = \phi$
- Objectively:
  - H represents all possible outcomes for a given situation
  - ϕ represents no outcomes
  - E and F each represent a subset of outcomes
  - $\cup$ represents the union of events (E or F)
  - $\cap$ represents the intersection of events (E and F)



H
A
B
P(B|A)
A∩B

Basic probability

- If $E \cap F = \phi$, the events are *disjoint*
  - Also called *mutually exclusive*.
- The complement of an event E ("NOT" E) is $\bar{E}$ or E^c.
  - $E \cap \bar{E} = \phi$, $E \cup \bar{E} = H$
  - From P3, $P(\bar{E}) = 1 - P(E)$.

$1 = P(\bar{E} \cup E) = P(\bar{E}) + P(E)$

$\rightarrow P(\bar{E}) = 1 - P(E)$

- Conditional Probability of E given F:

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

Sometimes we write EF or E, F for $E \cap F$.

  - $P(E \mid F)P(F) = P(EF)$
  - In general $P(E \mid F) \neq P(F \mid E)$

Basic probability

- E and F *independent* if $P(E \mid F) = P(E)$.
  - Occurrence of F does not influence probability of E.
  - The *multiplication rule* for independent events:

$$P(EF) = P(E)P(F).$$

  - If $P(E \mid F) = P(E)$, then $P(F \mid E) = P(F)$.

- Marginalization: Express $P(E)$ by "marginalizing" over F:

$$\begin{aligned} P(E) &= P(EF) + P(E\bar{F}) \\ &= P(F)P(E \mid F) + P(\bar{F})P(E \mid \bar{F}). \end{aligned}$$

Bayes Theorem

- Bayes' Theorem (Bayes' rule):

$$P(F \mid E) = \frac{P(F)P(E \mid F)}{P(E)}.$$

$$P(F \mid E) = \frac{P(F \cap E)}{P(E)} =$$

- Named after Rev. Thomas Bayes (c. 1750)

- Often written as:

$$P(F \mid E) = \frac{P(F)P(E \mid F)}{P(F)P(E \mid F) + P(\bar{F})P(E \mid \bar{F})}.$$

Example: Breast Cancer Screening

C: Cancer

M^+: + Mammography

- Consider the following:
  - 1% of women aged 40 who get a mammography have breast cancer $P(C) = 0.01$
  - 80% of women with breast cancer get positive mammographies $P(M^+|C) = 0.8$
  - 9.6% of women without breast cancer get positive mammographies $P(M^+|\bar{C}) = 0.096$
- What is the probability that a 40 year old women with positive mammography has cancer?

$$P(C|M^+) = \frac{P(C)P(M^+|C)}{P(C)P(M^+|C) + P(\bar{C})P(M^+|\bar{C})}$$

$$= \frac{0.01 \cdot 0.8}{0.01 \cdot 0.8 + 0.99 \cdot 0.096} = 0.078$$

Bayesian vs. frequentist probability

- The *frequentist probability* of an event is the limit of its relative frequency as its number of trials approaches ∞.
  - "What happens if random process repeated many times"

- In frequentist inference, we condition on unknown parameters, and find the probability of our data given these parameters.

$$P(\mathbf{y} \mid \theta)$$

  - $\mathbf{y}$ representa data, θ model parameters.
  - Parameters considered fixed and data considered random.

Bayesian vs. frequentist probability

- *Bayesian probability* quantifies a current degree of belief
- Need not consider repeating a process
  - e.g., "there is a 30% chance polio will be eradicated by 2025"
- In Bayesian inference we condition on the data, and find the probability of unknown parameters, given the data

$$P(\theta|\mathbf{y})$$

  - Parameters considered random and data considered fixed

Example: tipping pennies

- Claim: If you stand a penny on its side and gently let it fall, it will land heads more often than tails.
  - H_0: head/tails are equally likely
  - H_1: heads is more likely

$H_0: X \sim Binomial(20, 0.5)$

$P(X=k) = \binom{20}{k} 0.5^k (1-0.5)^{20-k}$

- In 20 trials, $X = 15$ land heads.
- Frequentist approach:
  - P-value $= P(X \geq 15 \mid H_0) \approx 0.02$

    $= P(X=15|H_0) + P(X=16|H_0) + \dots + P(X=20|H_0)$

    $= 0.02$
  - "If we repeat 20 trials many times, we would observe 15 or more heads only 2% of the time if H_0 is true"
  - Small p-value $\rightarrow$ evidence against H_0.

Example: tipping pennies

- Bayesian approach:
  - By Bayes' rule:

$$P(H_0|X = 15) = \frac{P(X = 15 \mid H_0)P(H_0)}{P(X = 15 \mid H_0)P(H_0) + P(X = 15 \mid H_1)P(H_1)}$$

  - Must compute $P(X = 15 \mid H_1)$. Requires more info for H_1
  - Must specify $P(H_0)$, our *prior* probability of H_0
  - $P(H_0|X = 15)$ is the *posterior* probability of H_0
  - This "updating" of prior to posterior characterizes Bayesian inference

Historical timeline

- c. 1750: Thomas Bayes first derives Bayes' theorem.
- Early 1800's: Pierre-Simon Laplace develops the notion of Bayesian probability
- Early 1900's: Statistics "founders " K. Pearson, R.A Fisher, J. Neyman and others develop frequentist approaches
  - Shun Bayesian probability and inference
- 1950's and 1960's: H. Jeffries, L.J. Savage, D. Lindley and others advocate Bayesian approaches
  - The "Bayesian revival"
- 1980's-:Advances in computing technology and methodology facilitates adoption of Bayes for wide variety of situations.
  - Still not without controversy.

Quotes

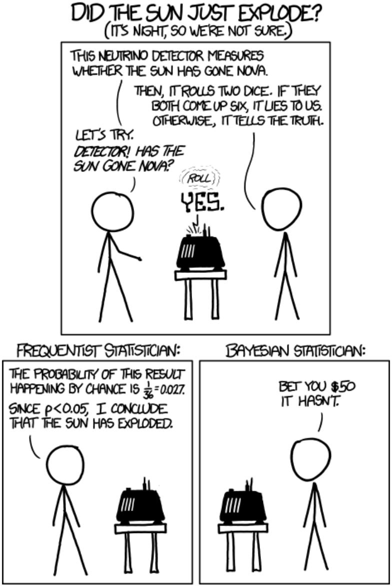
- "My personal conviction, is that the theory of inverse probability [Bayesianism] is founded upon an error and must be rejected." R.A. Fisher, 1925.
- "That question just doesn't make sense. Probability applies to a long sequence of repeatable events, and this is clearly a unique situation.." D. Blackwell, 1950, asked for the probability of a war within the next 5 years.
- "Uncertainty is a personal matter; it is not THE uncertainty but YOUR uncertainty." D. Lindley, 2006.
- "Today's posterior is tomorrows prior." D. Lindley, 1970.
- "Under Bayes' theorem, no theory is perfect. Rather, it is a work in progress, always subject to further refinement and testing." N. Silver, 2011.

Frequentist vs. Bayesian

- Criticism of Bayesian probability/inference:
  - Too subjective, prone to personal biases
  - Choice of "prior" may be unclear or arbitrary
  - Lack of consistency: different analyses of same data can give different results.
- Praise for Bayesian probability/inference
  - Broader conceptual framework, applies to more diverse situations involving chance.
  - Easily allows for incorporation of subjective belief and prior information.
  - More flexible incorporation of uncertainty for complex models.

Frequentist vs. Bayesian

Frequentist vs. Bayesian



Objective vs. subjective Bayesianism

- An *objective Bayesian*
  - minimizes influence of personal beliefs, maximizes influence of data, on inference.
  - espouse *non-informative* priors (e.g., $P(H_0) = 0.5$).
  - generally believe rational people (or machines) should make the same inferences when presented with the same data.
  - "the chance of temperatures below -30° F tomorrow is 20%"
- A *subjective Bayesian*
  - believes in "personal probability" – two people may give different but valid probability statements based on their personal understanding.
  - is more comfortable with strong prior beliefs influencing inference.
  - "I believe there is a 20% chance of temps below -30° F tomorrow".

Looking toward the future

- Bayesian approaches are rapidly gaining acceptance in several fields
  - Articles containing the word "Bayes" or "Bayesian" have steadily increased from near 0% in 1950s to over 50% for several journals in statistics, bioinformatics, econometrics and machine learning.
- Many today are not avidly "frequentist", "subjective Bayesian" or "objective Bayesian".
  - Choose the framework that is best suited for a given application or method.
- Theoretical and philosophical debates still linger...