Introduction to Bayesian Inference and Modeling

Edpsy 590BAY

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

Overview

- Simulation: https://chi-feng.github.io/mcmc-demo/app.html
- What is Bayes theorem
- Why Bayesian analysis
- What is probability?
- Basic Steps
- An little example
- Brief History
- In class work with probabilities

Depending on the book that you select for this course, read either Gelman et al. Chapter 1 or Kruschke Chapters 1 & 2.

Main References for Course

Throughout the coures, I will take material from

- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (20114). *Bayesian Data Analysis*, 3rd Edition. Boco Raton, FL, CRC/Taylor & Francis.**
- Hoff, P.D., (2009). *A First Course in Bayesian Statistical Methods*. NY: Sringer.**
- McElreath, R.M. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boco Raton, FL, CRC/Taylor & Francis.
- Kruschke, J.K. (2015). *Doing Bayesian Data Analysis: A Tutorial with JAGS and Stan*. NY: Academic Press.**

** There are e-versions these of from the UofI library. There is a verson of McElreath, but I couldn't get if from UofI e-collection.

Bayes Theorem

A whole semester on this?

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

where

- y is data, sample from some population.
- θ is unknown parameter.
- $p(y|\theta)$ is sample model, the data model, or the likelihood function.
- $p(\theta)$ is the prior distribution of the parameter θ.
- $p(y)$ is the probability of data or evidence.
- $p(\theta|y)$ is the posterior distribution of the parameter given data.

Why Use Bayesian Methods?

- Fundamentally sound
- Very flexible.
- Produces clear and direct inferences.
- Make use of all available information.
- Natural and principled way of combing prior information and Data.
- Solid decision theoretical framework.
- Can estimate very complex models (e.g., when MLE fails or is not practical or possible).
- Mixed effects (random & fixed) fit naturally into Bayesian framework.
- There are many computational tools that make it easy to implement.
- Can monitor probability as new information comes in.
- Doesn't depend on assumptions of sampling distributions and unobserved quantities...in Bayesian all relevant information necessary to make inference is contained in the data.

Disadvantages of Bayesian

- How to select prior – it's subjective
  Thoughtful choice, assumptions are clear
- Can produce posterior distributions that are heavily influenced by the priors.
  Requires thinking and allows for sensitivity analyses
- Often comes with high computational cost, especially in models with large numbers of parameters.
  Faster computers and cluster computing.
- Overfitting models is easy to do.
  This can be a problem, but there are ways to do model comparisons.
- Many were trained in frequentist methods and are unfamiliar with Bayesian approach.

Major Problem of Frequentist Hypothesis Testing

- Typically, $H_o : \theta = 0$, which is never true and is generally note the research question of interest.
- Researchers typically test a "no-effect" hypothesis that a straw-man designed to be rejected.
- There is nothing in the null hypothesis significance test (NHST) framework that requires testing a null hypothesis of no effect.
- Any theoretical justifiable hypothesis can serve as null hypothesis.

p-value Problem of NHST

- The calculation of the p-vale is based on data that were never observed.
- If we observed an effect, say $x = 8$, then the significance calculations involve not just $x = 9$ but also from more extreme values, $x > 8$.
- But $x > 8$ was not observed and it might not be possible to observed it in reality.

Jeffrey's (1081, pg. 385)
"I have always considered the arguments for the use of p absurd. They amount to saying that a hypothesis that may or may not be true is rejected because a greater departure from the trial value was improbable; that is, that it has not predicted something that has not happend. . . This seems a remarkable procedure."

p-value problem in NHST

(from Kaplan notes)

- Misunderstanding the Fisher or Neyman-Pearson framework to hypothesis testing and/or poor methodological practice is not a criticism of the approach per se.
  - Fisher: Set up null hypothesis and then compute the p-value which is used as a measure of evidence againist the hypothsis. The p-value is probability of seeing something more extreme.
  - Neyman-Pearson: define p-value as a function of data and then work out it's distribution under the null hypothesis. This is a formalized decision making process, "significant" or not.
  - Combining Fisher & Neyman-Person leads to inconsistencies.
- It is more than a misunderstanding. It seems to be fundamentally flawed.
- What can a Bayesian alternative provide?

Bayeisan Approach Provide. . .

(mostly from Kaplan notes)

- Bayesian hypothesis testing starts by obtaining summaries of relevant distributions
- Bayesian statistics aim to get summaries the "posterior" distribution.
- The formulas for the mean (expected value of posterior), variance, and mode of the posterior distribution come from expressions for the mean and variance of conditional distributions.
- Interval estimates have desirable interpretations.
- Can assess research questions.

Why Bayes?

- Probabilities can numerically represent a set of rational beliefs (i.e., fundamentally sound and based on rational rules).
- Explicit relationship between probability and information.
- Quantifies change in beliefs of a rational person when given new information (i.e., uses all available information–past & present).
- Very flexible
- Common sense interpretations.

In other words, if $p(\theta)$ approximates our beliefs, then $p(\theta|y)$ is optimal to what our posterior (after we have new information) beliefs about θ should be. Bayes can be used to explore how beliefs should be up-dated given data by someone with no information (e.g., who will win next election?) or with some information (e.g., will it in Los Vegas in 2022?).

Examples Where Bayes Used

Why Bayes? Because it worked

- Directed Allied artillery fire during WWII
- Alan Turning broke the German Enigma code
- Locate German U-boats
- Locate earthquake epicenters
- Find missing H-bombs during the cold ward (i.e., US and Russian subs)
- Investigate nuclear power safety
- Predict the Challenger tragedy

Examples Where Bayes Used

- Used in creation of worker's compensation insurance in the US
- Saved Bell telephone system financial panic of 1907
- Demonstrated that smoking causes lung cancer
- Show that high cholesterol causes heart attacks
- Probability of cancer given positive mammogram
- Spam filters
- Internet searches
- Spell checkers
- Who wrote the Federalist papers
- Predict winners of elections or sports contests (e.g., see https://fivethirtyeight.com/)

Modern Uses

- Coast Guard to locate survives of shipwrecks
- How genes control and regulate
- Wall Street (financial markets)
- Astronomy & Physics
- AI & machine learning
- Homeland security
- Microsoft
- Google
- Language translation
- Students in Edpsy 590BAY

General Uses of a Bayesian Approach

- Parameter estimates with good statistical properties
- Parsimonious descriptions of observed data.
- Predictions for missing data.
- Predictions of future data.
- Computational frame-work for model estimation and validation.
- Provides a solution to complicated statistical problems that have no obvious (non-Bayesian) method of estimation and inference (e.g., complex statistical model, estimation of rare events).

Psychological Methods 2017, vol 22 Issue 2

- Bayesian hypothesis testing: Editorial to the Special Issue on Bayesian data analysis (Hoijtink, Herbert; Chow, Sy-Miin)
- A systematic review of Bayesian articles in psychology: The last 25 year (van de Schoot, Rens; Winter, Sonja D; Ryan, Oisín; Zondervan-Zwijnenburg, Mariëlle; Depaoli, Sarah.)
- Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist (Depaoli, Sarah; van de Schoot, Rens)
- Bayesian evaluation of constrained hypotheses on variances of multiple independent groups (Böing-Messing, Florian; van Assen, Marcel A. L. M; Hofman, Abe D; Hoijtink, Herbert; Mulder, Joris)
- Bayesian analyses of cognitive architecture (Houpt, Joseph W; Heathcote, Andrew; Eidels, Ami)
- Bayesian analysis of factorial designs (Rouder, Jeffrey N; Morey, Richard D; Verhagen, Josine; Swagman, April R; Wagenmakers, Eric-Jan)

Psychological Methods 2017, vol 22 Issue 2

- Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences (Schönbrodt, Felix D; Wagenmakers, Eric-Jan; Zehetleitner, Michael; Perugini, Marco)
- Decision qualities of Bayes factor and p value-based hypothesis testing (Jeon, Minjeong; De Boeck, Paul)
- A comparison of Bayesian and frequentist model selection methods for factor analysis models (Lu, Zhao-Hua; Chow, Sy-Miin; Loken, Eric)
- Posterior calibration of posterior predictive p values (van Kollenburg, Geert H; Mulder, Joris; Vermunt, Jeroen K.)
- Assessing fit of alternative unidimensional polytomous IRT models using posterior predictive model checking (Li, Tongyun; Xie, Chao; Jiao, Hong.)

Psychological Methods 2017, vol 22 Issue 4

- Using phantom variables in structural equation modeling to assess model sensitivity to external misspecification (Harring, Jeffrey R; McNeish, Daniel M; Hancock, Gregory R.)
- Distinguishing outcomes from indicators via Bayesian modeling (Levy, Roy)
- Moderation analysis with missing data in the predictors (Zhang, Qian; Wang, Lijuan)
- Bayesian dynamic mediation analysis (Huang, Jing; Yuan, Ying)
- An alternative to post hoc model modification in confirmatory factor analysis: The Bayesian lasso (Pan, Junhao; Ip, Edward Haksing; Dubé, Laurette)
- Using expert knowledge for test linking (Bolsinova, Maria; Hoijtink, Herbert; Vermeulen, Jorine Adinda; Béguin, Anton)

Psychological Methods 2017, vol 22 Issue 4

- Bayesian models for semicontinuous outcomes in rolling admission therapy groups (Burgette, Lane F; Paddock, Susan M)
- Bayesian unknown change-point models to investigate immediacy in single case designs (Natesan, Prathiba; Hedges, Larry V)
- Multilevel modeling of single-case data: A comparison of maximum likelihood and Bayesian estimation (Moeyaert, Mariola; Rindskopf, David; Onghena, Patrick; Van den Noortgate, Wim)
- Developing constraint in bayesian mixed models (Haaf, Julia M; Rouder, Jeffrey N.)
- A Bayesian "fill-in" method for correcting for publication bias in meta-analysis (Du, Han; Liu, Fang; Wang, Lijuan)

Major Problems using Bayesian Approach

- Specifying prior knowledge; that is, choosing a prior.
- Sample from

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

- Computationally intensive

What do we Mean by "Probability"

Different authors of Bayesian texts use different terms for probability, which reflect different conceptualizations.

- Beliefs
- Credibility
- Plausibilities
- Subjective

There are multiple specific definitions:

- Frequentist: long run relative frequency of an event.
- Bayesian: a fundamental measure of uncertainty that follow rules probability theory.
  - What is the probability of thunder snow tomorrow?
  - What is the probability that Clinton nuclear power plant has a melt down?
  - What is the probability that a coin tossed lands on head?

Probabilities as We've Known Them

Probabilities are foundational concept!
Probabilities are numerical quantities that measures of uncertainty.

Justification for the statement "The probability that an even number comes up on a toss of a dice equals 1/2."

Symmetry or exchangeability argument:

$$p(\text{even}) = \frac{\text{number of evens rolled}}{\text{number of possible results}}$$

The justification is based on the physical process of rolling a dice where we assume each side of a 6 sided die are equal likely, three sides have even numbers, the other three have odd numbers.

$y_1 =$even or odd on first roll should be the same as y_2 on 2nd, etc.

Probabilities as We've Known Them

Alternative justification for the statement "The probability that an even number comes up on a toss of a dice equals $1/2$."

Frequency argument:

$$p(\text{even}) = \text{Long run relative frequency}$$

Long (infinite) sequence of physically independent rolls of the dice.

Are these justifications subjective? These involve hypotheticals: physical independence, infinite sequence of rolls, equally likely, mathematical idealizations.

How would either of these justifications apply to

- If we only roll dice once?
- What's the probability that USA womens soccer team wins the next World Cup?

Probabilities as measure of Uncertainty

- Randomness creates uncertainty and we already do it in common speech...what are synonyms for "probability"?
- Coherence: probabilities principles of basic axioms of probability theory, which have a consequences things such as :
  - $0 \leq \text{p(X)} \leq 1$
  - if X is subset or equal to Y, then $p(X) \leq p(Y)$, e.g., Standard deck of 52 cards where $Y =$ all the red cards and X are the red heart cards.

$$P(\text{red}) = 26/52 = .5 \qquad \text{and} \qquad P(\text{red heart}) = 2/26 = .0759$$

  - $\sum p(X) = 1$ or $\int p(X) = 1$, e.g., X are the cards with hearts, so $\sum_{\text{facevalue}} P(\text{heart} = \text{facevalue}) = 1$
  - $p(X, Y) = p(X) + p(Y) - p(X \bigcap Y)$,

Expected Value

- Expected value is a mean of some statistic or quantity based on a random event or outcome.
- For discrete random variables, say X with "probability mass function" $p(x)$, the mean of X is

$$E(X) = \mu_X = \sum_{i=1}^{I} x_i Pr(X = x_i),$$

where I is the number of possible values for x, and $Pr(X = x_i)$ is the probability that $X = x_i$.
e.g., A bet where P(win $\$100$)=.1 and P(loose $\$10$)=.9, the expected value is

$$E(\text{outcome}) = 100 * .1 - 10 * 0.9 = 1$$

What is the "rational" price to pay to play?

Expected Value

For continuous random variables, say X with a probability density function $f(x)$, the mean of X is

$$E(X) = \mu_X = \int_x x f(x) d(x),$$

where integration is over all possible values of x.

e.g., If $f(x)$ is a normal distribution and $-\infty < x < \infty$.

Basic Steps of Bayesian Analysis

(from Gelman et al.)

I assume that you have research questions, collected relavent data, and know the nature of the data.

- Set up full probability model (a joint probability distribution for all observed and unobserved variables that reflect knowledge and how data were collected):

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

  This can be the hard part
- Condition on data to obtain the posterior distribution:

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

  Tools: analytic, grid approximation, Markov chain Monte Carlo (i.e., Metropolis(-Hastings), Gibbs sampling, Hamiltonian).
- Evaluate model convergence and fit.

A Closer Look at Bayes Rule

I am being vagues in terms of what y and θ are (e.g., continuous, discrete, number of parameters, what the data are).

$$\begin{aligned} p(\theta|y) &= \frac{p(y|\theta)p(\theta)}{p(y)} \\ &\propto p(y|\theta)p(\theta) \end{aligned}$$

where $p(y)$

- ensures probability sums to 1.
- is constant (for a given problem).
- "average" of numerator or the evidence.
- For discrete y: $p(y) = \sum_{\theta \in \Theta} p(y|\theta)p(\theta)$
- For continuous y: $p(y) = \int_{\theta} p(y|\theta)p(\theta)d(\theta)$

More on the Uses of a Bayesian Approach

- If $p(\theta)$ is wrong and doesn't represent our prior beliefs, the posterior is still useful. The posterior, $p(\theta|y)$, is optimal under $p(\theta)$ which means that $p(\theta|y)$ will generally serve as a good approximation of what our beliefs should be once we have data.
- Can use Bayesian approach to investigate how data would be updated using (prior) beliefs from different people. You can look at how opinions may changes for someone with *weak prior information* (vs someone with strong prior beliefs). Often diffuse or flat priors are used.
- Can handle complicated problems.

Example: Spell Checker

(from Gelman et al.)

Suppose the word "radom" is entered and we want to know the probability that this is the word intended, but there 2 other similar words that differ by one letter.

θ	frequency from Google database	Prior $p(\theta)$	Google's model $p(\text{'radom'}\|\theta)$	numerator $p(\theta)p(\text{'radom'}\|\theta)$
random	7.60×10^{-5}	0.9227556	0.001930	1.47×10^{-7}
radon	6.05×10^{-6}	0.0734562	0.000143	8.65×10^{-10}
radom	3.12×10^{-7}	0.0037881	0.975000	3.04×10^{-7}
total	8.2362×10^{-5}	1.00	1.00	4.51867×10^{-7}

Spell Checker (continued)

θ	Prior p('radom')	Google's model p('radom'$\|\theta$)	numerator of Bayes $p(\theta)p$('radom'$\|\theta$)	Posterior $p(\theta\|$'radoM')
random	0.9227556	0.001930	1.47×10^{-7}	0.325
radon	0.0734562	0.000143	8.65×10^{-10}	0.002
radom	0.0037881	0.975000	3.04×10^{-7}	0.673
total	1.00	1.00	4.51867×10^{-7}	1.000

Note:

$$p(\text{'radom'}|\text{'random'}) = \frac{1.47 \times 10^{-7}}{4.51867 \times 10^{7}} = .325$$

Spell Checker (continued)

θ	Prior $p(\theta)$	Data model $p(\text{'radom'}\|\theta)$	Posterior $p(\theta\|\text{'radom'})$
random	0.9227556	0.001930	0.325
radon	0.0734562	0.000143	0.002
radom	0.0037881	0.975000	0.673
total	1.00	1.00	1.000

What is "radom"?

Some averaging of prior and data going on...most in this next lecture.

What could be some criticisms of this example or how might it be improved?

History: Rev Thomas Bayes

Leonard (2014), Fienberg, S. (2006), but mostly S.B. McGrayne (2011). *The Theory That Would Not Die*.

David Humes

Before talking about Bayes. . .

- In the 1700s, scientists thought that the existence of natural laws proved in the existence of God.
- David Hume (philosophiser) published an essay challenging Christianity's belief that God designed and created the world (i.e., God as the first cause).
- Hume argued that you can't know anything with certainty based on inductive reasoning (i.e., association, correlation); that is, observing the effect doesn't tell you with certainty the cause.
- Ideas of cause and effect were central.
  - Hume: Can't know cause with certainty when only observing the effect.
  - Others: Knowing the effect can prove the cause.
  - Hume's essay was non-mathematical

1740s: Rev Thomas Bayes Contribution

- Bayes was a Presbyterian paster and amateur mathematician. Since Bayes was not a member of Church of England, he would be considered a nonconformist or dissenter. As a mathematician, he would have been considered a "infidel mathematician".
- 1763 Rev Thomas Bayes gave the first description of the theorem in "An essay toward solving a problem in the doctrine of chance".
- Bayes was concerned with an to question of how to get from effect to cause or the "inverse" probability question.b
- Bayes dealt with the problem of drawing inference; that is, concerned with "degree of probaility".
- Wanted to learn the probability of a future event given he knew nothing about it except past and needed to quantify this.
- Bayes introduces uniform prior distribution for binomial proportion.
- He devised a thought experiment about where a ball tossed was located on table behind him. He could narrow down the position and infer where is could land between 2 bounds but never know the precise location. However with could increase confidence in the location.
- Bayes did not give statement of what we call "Bayes Theorem".

Rev Richard Price Contribution

. . . also a Presbyterian clergy

- Found Bayes' essay and answered Humes attack on causation.
- Price edited it, added citations, etc.
- Added an appendix that deals with the problem of prediction; that is, a use for it.
- Called it "probability of casues" or "inverse probability"
- Published
- By today's standard's we would call it the "Bayes-Price Rule" or something like this.

1774 Pierre Simon LaPlace)

- LaPlace was the Einstein of the 1770s.
- 1774 Pierre Simon LaPlace gave more elaborate version of Bayes theorem for the problem of inference for an unknown binomial probability in more modern language. He clearly augured for choosing a uniform prior because he reasoned that the posterior distribution of the probability should be proportional to the prior,

$$f(\theta|x_1, x_2, \ldots x_n) \propto f(x_1, x_2, \ldots, x_n|\theta)$$

- LaPlace introduced the idea of "indifference" as an argument to use uniform prior; that is, you have no information what the parameter should be. (He meet and discussed things with Richard Price, but LaPlace's developed idea without first knowing about Bayes's work).
- Bayes was concerned with the probability something would occur. LaPlace dealt with probability of an amount (as well as occurrence).

Pierre Simon LaPlace)

LaPlace is also known for (among other things)

- Work in the area of astronomy where in the 1700s time measurements were unreliable; hence some uncertain.
- Discovered the Central Limit Theorem
- Developed expansion of integrals (the "LaPlace" method is still used — estimating linear and non-linear mixed models in the nlme R package and SAS NLMIXED (with qpoint=1)).
- As an instructor at a Miliary School, he passed Napoleon in his mathematics course.

Basic/Early Work

- 1749 David Hartley's book describes the "inverse" result and attributes is to a friend. Speculation is that the friend was either Saunderson or Bayes.
- I.J. Bienayme generalized LaPlace's work.
- von Mise gave a rigorous proof of Bayes theorem.
- 1837–1843: at least 6 authors, working independently, made distinctions between probabilities of things (objective) and subjective meaning of probability (i.e., S.D. Poisson, D. Bolzano, R.L Ellis, J.F. Frees, J.S. Mills and A.A. Counot).
- Debate on meaning of probability continued throughout the 1800s.
- Some adopted the inverse probability (i.e, Bayesian) but also argued for a role of experience, including Pearson, Gosset and others.

1900s

Science should be objective and precise.

- 1912–1922: Fisher advocated moving away from inverse methods toward inference based on likelihood.
- Fished moved away from "inverse probability" and argued for a frequentist approach.
  "...the theory of inverse probability is founded upon an error, and must wholly be rejected."
- Fundamental change in thinking.
- Beginnings of formal methodology for significance tests.
- J Neyman & Ego Pearson gave more mathematical detail and extended ("completed") Fisher's work which gave rising to the hypothesis test and confidence intervals

1900s (continued)

- After WWI, frequentist methods usurped inverse probability and Bayesian statistician were marginalized.
- R. von Mises justified the frequentist notion of probability; however, in 1941 he used a Bayesian argument to critique Neyman's method for confidence intervals. He argued that what really is wanted was posterior distribution.
- 1940 Wald showed that Bayesian approach yielded good frequentist properties and helped to rescue Bayes Theorem from obscurity.
- 1950s The term "frequentist" starts to be used. The term "Bayes" or "Bayes solution" was already in use. The term "classical" statistics refers to the frequentist.

1900s (continued)

- J.M Keynes (1920s) laid out axiomatic formulation and new approach to subjective probabilities via the concept of expect utility. Some quotes that reflect this thinking:
  - "In the long run we are all dead."
  - "It is better to be roughly right than precisely wrong."
  - "When the facts change, I change my mind."
- 1930s: Bruno de Finetti gave a different justification for subject probabilities and introduced the notion of "exchangeability" and implicit role of the prior distribution.
- Savage build on de Finetti's ideas and developed set of axioms for non-frequentist probabilities.

WWII

- Alan Turing and his code breaking work was essentially Bayesian — sequential data analysis using weights of evidence. It is thought that he independently thought of these ideas.
- Decision-theory developments in the 1950s.

1980s and Beyond

Large revival started in the late 1980s and 1990s. This was due to new conceptual approaches and lead to rapid increases in applications. The increase in computing power helped fuel this.

Non-Bayesian approaches will likely remain important because of the high computational demand and expense of Bayeisan methods, even though there are continual developments in computing power and improvements in algorithms. . . and that's what is taught in most statistics courses.

Edwards Bayesian Research Conference

Ward Edwards introduced to Bayesian ideas from Jimmy Savage and applied to Decision research:

- Decision making under risk, uncertainty, and ambiguity
- Intertemporal choice
- Cognitive models of judgment and decision making
- Mathematical and statistical methodology for analyzing behavioral data
- Applications of JDM theory and models to health care and public policy
- Medical, legal, and business decision making
- Expert forecasting
- Wisdom of the crowds

Ward Edwards Bayesian Research Conference ($\sim$ 1985?)

Practice 1: Subjective Probability

Discuss the following statements: "The probability of event E is considered 'subjective' if two rational people A and B can assign unequal probabilities to E, $P_A(E)$ and $P_B(E)$. These probabilities can also be interpreted as 'conditional': $P_A(E) = P(E|I_A)$ and $P_B(E|I_A)$, where I_A and I_B represent the knowledge available to person A and B, respectively." Apply this idea to the following examples.

- The probability that a "6" appears when a fair die is rolled, where A observes the outcome and B does not.
- The probability that USA wins the next mens World Cup, where A is ignorant of soccer and B is a knowledgable sports fan.
- The probability that UofI's football team goes to a bowl game, where A is ignorant of Illini football and B is knowledgable of Illini football.

Practice 2: Cancer & Monograms

What is the probability of breast cancer given a positive result of mammogram?

US women in 40s without symptoms or family history of disease:

- Probability of breast cancer is .01
- Probability of breast cancer patients getting an abnormal results is .80
- Probability of a positive mammogram result if a woman does not have breast cancer is .0996

Practice 2: Cancer & Monograms: SOLUTION

What is the probability of breast cancer given a positive result of mammogram?

$$
\begin{aligned}
P(BC|+) &= \frac{P(+|BC) \times P(BC)}{P(+)} \\
&= \frac{P(+|BC) \times P(BC)}{P(+|BC) \times P(BC) + P(+|\overline{BC})P(\overline{BC})} \\
&= \frac{.80 * .01}{.80 * .01 + .0996 * (1 - .01)} \\
&= \frac{.08}{.9584} = .0750
\end{aligned}
$$

If had 2nd positive test, what is $P(BC|+)$?

$$
\begin{aligned}
P(BC|+) &= \frac{.80 * .0750}{.80 * .0750 + .0996 * (1 - .0750)} \\
&= .408 \longrightarrow \text{ about } 41
\end{aligned}
$$

Practice 3: Conditional Probabilities and a little R

(from Gelman et al.) This is homework. . .

Suppose that $\theta = 1$, then Y has a normal distribution with mean 1 and standard deviation σ, and if $\theta = 2$ then Y is normal with mean 2 and standard deviation σ. Also suppose $Pr(\theta = 1) = 0.5$ and $Pr(\theta = 2) = 0.5$.

- For $\sigma = 2$, write the formula for the marginal probability density for y and sketch/plot it. For the graph, these R commands are sufficient:
  - *seq*
  - *dnorm*
  - *plot*
- What is $Pr(\theta = 1|y = 1)$ and what is $Pr(\theta = 1|y = 2)$. (hint: Definition of conditional probability, Bayes Theorem)
- Describe how the posterior density of θ changes shape as
  - σ increases
  - σ decreases
  - Difference between μ's increase.
  - Different between μ's decrease.