Economics 403A

Bayesian Inference

Dr. Randall R. Rojas

Today's Class

- Introduction to Bayesian Inference
  - Bayes Theorem
  - Historical Background
  - Bayesians vs Frequentists
  - Priors, Likelihood and Posterior
  - Credible Intervals
  - Predictive Inference
  - Hypothesis Testing

Introduction to Bayesian Inference

Bayes' Theorem

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)} = \frac{P(X|A)P(A)}{P(X|A)P(\sim A) + P(X|A)P(\sim A)}$$

Posterior = Prior + Data

Given some phenomenon A (e.g., cancer) that we want to investigate, and an observation X (e.g., mammography test outcome) that is evidence about A, we can update the original probability of A, given the new evidence X.

The **posterior** captures all the information that X can provide about A

Introduction to Bayesian Inference

Goals of Inference

The two main goals of inference:

1. What is a good guess of the population model (the true parameters)?

2. How do I quantify my uncertainty in the guess?

Bayesian inference answers both questions directly through the posterior.

Introduction to Bayesian Inference

Motivating Example

- Assume the following information:
  - 1% of women aged 40 have breast cancer
  - A mammography test has an 80% success rate
  - A mammography test has a 10% false alarm rate.

- Given the above, a women aged 40 receives a positive mammography test. What is the probability that she actually has cancer?

- Q: Do you think this probability is high or low?

Introduction to Bayesian Inference

Motivating Example

- Solution:
  1. Identify A and X:
     X = +mamm. Test, ~X = - mamm. Test
     A = Cancer present, ~A = Cancer not present

  2. Identify the relevant probabilities
     P(A) = Prob. of having breast cancer = 1%
     P(~A) = Prob. of not having breast cancer = 99%
     P(X|A) = Prob. that she receives a +mamm. test, given that she has cancer = 80%
     P(X|~A) = Prob. that she receives a +mamm. test but does not have cancer = 10%
     P(A|X) = Prob. That she has cancer given that she received a positive mamm. test.

  3. Apply Bayes' Theorem:

$$P(A|X) = \frac{P(X|A)P(A)}{P(X|A)P(\sim A) + P(X|A)P(\sim A)} = \frac{80 \times 1}{80 \times 1 + 10 \times 99} = 7.5\%$$

Introduction to Bayesian Inference

Bayes' Theorem (in General)

Prior: contains all the available information about the parameter values "prior" to observing the data.

Likelihood: assuming you know the parameters exactly, how "likely" are the data. Represents the data that are now available.

$$\underbrace{P(A|B_k)} = \frac{\overbrace{P(A_k)}\overbrace{P(B|A_k)}}{\underbrace{\sum_{i=1}^{m} P(A_i)P(B|A_i)}}$$

Posterior: expresses our knowledge about the parameters after seeing the data.

Normalizing Constant

Introduction to Bayesian Inference

Background

James Bernoulli (1713) asked whether the rules of probability (then being worked out to calculate odds in gambling) could also be used in the problem of inference or inductive logic.

Rev. Thomas Bayes answered this in the affirmative in an essay published posthumously in 1763.

Pierre-Simon Laplace reformulated Bayes' theorem in 1812. He expressed it with far greater clarity than Bayes and used it to solve problems in celestial mechanics and medical statistics.

- Laplace combined the available astronomical data to provide an estimate (and uncertainty) on the mass of Saturn. He stated "...it is a bet of 11,000 to 1 that the error in this result is not 1/100th of its value". Note: The modern estimate differs from Laplace's by 0.63%)

Introduction to Bayesian Inference

Background

- Bayes and Laplace thought of probability as measuring a degree of belief in something. They used the language of odds and betting in their arguments.
  - For about 100 years, the Bayesian interpretation reigned supreme.
- However, with the push in the 19th and early 20th centuries towards more rigorous arguments in mathematics, this view was considered too subjective and a new definition of probability was developed.
  - In the beginning of the 20th century, the so-called frequentist interpretation emerged (Ronald Fisher, Jerzy Neyman, Egon Pearson)
  - Probability was defined solely in terms of the long-term frequency of events.
  - Note: This definition works fine for rolling dice and tossing coins but where does it leave Laplace's measurement of the mass of Saturn ?

Introduction to Bayesian Inference

Background

- During the second half of the 20th century, Bayesian statistics made a strong comeback.
- The frequentist approach is plagued by inconsistencies and limitations.
- Bayesian models are often analytically intractable and thus require methods based on simulation. Cheap and fast computers, and general-purpose software such as BUGS, resolved this issue.
- Hierarchical graphical models, such as hidden Markov models and Bayesian networks, provided a unified framework for Bayesian computation.
  - Bayesian approaches are very common in machine learning.
  - Recent work in cognitive science indicates that our minds may well operate by Bayesian inference
  - Bonus Fact: Alan Turing cracked the U-boats encrypted communications using Bayesian methods

Introduction to Bayesian Inference

Background

- Frequentists treat the parameters as fixed (deterministic).
  - Considers the training data to be a random draw from the population model.
  - Uncertainty in estimates is quantified through the sampling distribution: what is seen if the estimation procedure is repeated over and over again, over many sets of training data.
- Bayesians treat the parameters as random.
  - Key element is a prior distribution on the parameters.
  - Using Bayes' theorem, combine prior with data to obtain a posterior distribution on the parameters.
  - Uncertainty in estimates is quantified through the posterior distribution.

Introduction to Bayesian Inference

Key differences between Bayesian and Classical (Frequentist) Approaches

- Unknown parameter θ:
  - Frequentist: a fixed number
  - Bayesian: a random variable
- Inference about θ:
  - Frequentist: ad hoc (different types of estimators/tests are "best" for different problems, no unique algorithm)
  - Bayesian: given 3 choices (likelihood, prior, loss), there is a unique inferential procedure

Introduction to Bayesian Inference

Key differences between Bayesian and Classical (Frequentist) Approaches

- Interpretation of the conclusions: e.g., interval estimation of θ:
  - Frequentist: $(1 - \alpha)100\%$ confidence interval of θ: among all such data sets y, in $(1 - \alpha)100\%$ of them, θ belongs to this interval
  - Bayesian: $(1 - \alpha)100\%$ credible interval of θ: for given data y, θ belongs to this interval with probability $1 - \alpha$
- Subjectivity/objectivity of conclusions about θ:
  - Frequentist: objective (some argue that subjectivity is often hidden in this approach)
  - Bayesian: subjective: use a priori information about θ

Introduction to Bayesian Inference

Posterior Predictive Distribution

If we wish to predict a new observation $\tilde{x}$ on the basis of the sample $x = (x_1, \ldots, x_n)$, we may use its posterior predictive distribution. This is defined to be the conditional distribution of $\tilde{x}$ given x:

$$p(\tilde{x}|x) = \int p(\tilde{x}, \theta|x)d\theta = \int p(\tilde{x}|x, \theta)p(\theta|x)d\theta$$

where p($\tilde{x}$|x, θ) is the density of the predictive distribution

It is easy to simulate the posterior predictive distribution. First, draw simulations $\theta_1, \ldots, \theta_L$ from the posterior p(θ|x), then, for each i, draw ỹ from the predictive distribution p($\tilde{x}$|x, θ).

Introduction to Bayesian Inference

Posterior Predictive Distribution

$p(x\|\theta)$	likelihood
$\pi(\theta)$	prior
$p(x) = \int p(x\|\theta)\pi(\theta)\ d\theta$	marginal likelihood
$p(\theta\|x) = \dfrac{p(x\|\theta)\pi(\theta)}{p(x)}$	posterior probability
$p(x_{new}\|x) = \int p(x_{new}\|\theta)\pi(\theta\|x)\ d\theta$	predictive probability

Introduction to Bayesian Inference

Posterior Predictive Distribution

- Example: Assume that we have a coin with unknown probability θ of a head. If there occurs x heads among the first n tosses what is the probability of a head on the next throw?
- Let $\tilde{x}$ = 1 ($\tilde{x}$= 0) indicate the event that the next throw is a head (tail). If the prior of θ is a Beta(α, β), then

$$p(\tilde{x}|x) = \int_0^1 p(\tilde{x}|x,\theta)p(\theta|x)d\theta \quad = \int_0^1 \theta^{\tilde{x}}(1-\theta)^{1-\tilde{x}} \frac{\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}}{B(\alpha+x,\beta+n-x)}d\theta$$

$$= \frac{(\alpha+x)^{\tilde{x}}(\beta+n-x)^{1-\tilde{x}}}{\alpha+\beta+n}$$

Therefore, for example p($\tilde{x}$ =1|x)=(α+x)/(α+β+n). This tends to the sample proportion x/n as $n \to \infty$, so that the role of the prior information vanishes.

Introduction to Bayesian Inference

Prior

Where did the prior come from?

→ There are two schools of thought on the prior:

- Subjective Bayesian
  - The prior is a summary of our subjective beliefs about the data.
  - E.g., in the coin flipping example: the prior for p should be strongly peaked around 1/2.
- Objective Bayesian
  - The prior should be chosen in a way that is "uninformed".
  - E.g., in the coin flipping example: the prior should be uniform on [0, 1].

Introduction to Bayesian Inference

Non-informative Priors

- We often do not have any prior information, although true Bayesian's would argue we always have some prior information!
- We would hope to have good agreement between the frequentist approach and the Bayesian approach with a non-informative prior.
- Diffuse or flat priors are often better terms to use as no prior is strictly non-informative!
- For our example of an unknown mean, candidate priors are a Uniform distribution over a large range or a Normal distribution with a huge variance.

Introduction to Bayesian Inference

Improper Priors

- The limiting prior of both the Uniform and Normal is a Uniform prior on the whole real line.
- Such a prior is defined as **improper** as it is not strictly a probability distribution and does not integrate to 1.
- Some care has to be taken with improper priors however in many cases they are acceptable provided they result in a proper posterior distribution.
- Uniform priors are often used as non-informative priors however it is worth noting that a uniform prior on one scale can be very informative on another.
- **For example**: If we have an unknown variance we may put a uniform prior on the variance, standard deviation or log(variance) which will all have different effects.

Introduction to Bayesian Inference

Conjugate Priors

- Computations can often be facilitated using **conjugate prior** distributions. We say that a prior is conjugate for the likelihood if the prior and posterior distributions belong to the same family.

Note: See e.g., https://en.wikipedia.org/wiki/Conjugate_prior
for a list of common conjugate priors.

Introduction to Bayesian Inference

Posterior

The posterior can be used in many ways to estimate the parameters. For example, we can compute:

- The mean
- The median
- The mode

Depending on context, these are all potentially useful ways to estimate model parameters from a posterior.

Introduction to Bayesian Inference

Example: Normal Distribution

- Consider a single-parameter model and assume $x = (x_1,\ldots x_n)$ is a sample from a normal distribution with unknown mean θ and variance σ^2.
- Likelihood: $p(x|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\theta)^2} \propto e^{-\frac{n}{2\sigma^2}(\theta-\overline{x})^2}$
- We can find a conjugate prior by replacing σ^2/n by $\tau_0{}^2$ and with μ_0

$$\rightarrow p(\theta) \propto e^{-\frac{1}{2\tau_o^2}(\theta-\mu_0)^2} \sim N(\mu_0, \tau_0^2)$$

Introduction to Bayesian Inference

Example: Normal Distribution

- Now we can easily compute the posterior p(θ|x):

$$p(\theta|x) \propto p(\theta)p(x|\theta) \propto e^{-\frac{1}{2\tau_o^2}(\theta-\mu_0)^2} e^{-\frac{n}{2\sigma_o^2}(\theta-\overline{x})^2}$$

$$\propto e^{-\frac{1}{2\tau_n^2}(\theta-\mu_n)^2} \sim N(\mu_n, \tau_n^2)$$

where $\mu_n = \dfrac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\overline{x}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$ and $\tau_n^2 = \left(\dfrac{1}{\tau_0^2} + \dfrac{n}{\sigma^2}\right)^{-1}$

Introduction to Bayesian Inference

Example: Normal Distribution

- The inverse of variance is called the precision. Therefore, we see that:

  → *posterior precision = prior precision + data precision*

- For example, for our previous case, the prior precision is $1/\tau_0^2$ and data precision n/σ^2 (i.e., the inverse of the variance of the sample mean)
- The **posterior mean** is a weighted average of the prior mean μ and sample mean $\overline{x}$ where the weights are the corresponding precisions. When $n \to \infty$ (or when $\tau_0 \to \infty$), the *role of the prior information vanishes*. Thus, for large values of n, approximately $\theta | y \sim N(\overline{x}, \sigma^2/n)$.

Introduction to Bayesian Inference

Example: Normal Distribution

- Next, we determine the posterior predictive distribution of a new observation $\tilde{x}$. The joint posterior distribution of θ and $\tilde{x}$ is:

$$\rightarrow p(\theta, \tilde{x}|x) = p(\theta|y)p(\tilde{x}|x, \theta)$$

$$\propto e^{-\left(\frac{1}{2\tau_n^2}(\theta-\mu_n)^2+\frac{1}{2\sigma^2}(\tilde{x}-\theta)^2\right)}$$

$$\rightarrow E[\tilde{x}|x] = E[\theta|x] = \mu_n$$

$$\mathrm{Var}[\tilde{x}|x] = E[\sigma^2|x] + \mathrm{Var}[\theta|x] = \sigma^2 + \tau_n^2$$

- Therefore, the posterior predictive distribution is:

$$p(\tilde{x}|x) = N(\tilde{x}|\mu_n, \tau_n^2)$$

Introduction to Bayesian Inference

Point and Interval Estimation

- In Bayesian inference the outcome of interest for a parameter is its full posterior distribution however we may be interested in summaries of this distribution.

- A simple **point estimate** would be the mean of the posterior (although the median and mode are alternatives)

- Interval estimates are also easy to obtain from the posterior distribution and are given several names, for example **credible intervals**, Bayesian confidence intervals and Highest density regions (HDR). All of these refer to the same quantity.

Introduction to Bayesian Inference

Credible Intervals

- Credible intervals can be interpreted in the more natural way that there is a probability of 0.95 that the interval contains μ rather than the frequentist conclusion that 95% of such intervals contain μ.
- Def: A credible interval, for a real-valued parameter $\psi(\theta)$, is an interval $C(s) = [l(s), u(s)]$ that we believe will contain the true value of ψ. As with the sampling theory approach, we specify a probability γ and then find an interval $C(s)$ satisfying

$$\Pi(\psi(\theta) \in C(s) \mid s) = \Pi(\{\theta : l(s) \leq \psi(\theta) \leq u(s)\} \mid s) \geq \gamma$$

We then refer to C(s) as a γ-credible interval for ψ.

Introduction to Bayesian Inference

Credible Intervals

- A Bayesian credible interval of size *1 − α* is an interval *(a, b)* such that

$$P\,(a \le \theta \le b|x) = 1 - \alpha$$

$$\int_a^b p(\theta|x)\, d\theta = 1 - \alpha$$

- Note: When you are calculating credible intervals, you willl find the values of *a* and *b* by several means. You could be asked do the following:
  - Find the *a, b* using means of calculus to determine the credible interval or set.
  - Use a Standard Normal (Z-scores) when appropriate.
  - Use R to approximate the values of *a* and *b*.

Introduction to Bayesian Inference

Credible Intervals

- Our definition for the credible interval could lead to many choices of *(a, b)* for particular problems. Suppose that we required our credible interval to have equal probability *α/2* in each tail. That is, we will assume:

  $P(\theta < a|x) = \alpha/2$ and $P(\theta > b|x) = \alpha/2$

- Note: Credible intervals are not necessarily unique.

Introduction to Bayesian Inference

Confidence Intervals vs. Credible Intervals

- A **confidence interval** is constructed to contain θ a percentage of the time, say 95%.
  - Suppose our confidence level is 95% and our interval is (L,U). Then we are 95% confident that the true value of θ is contained in (L,U) in the long run.
  - In the long run means that this would occur nearly 95% of the time if we repeated our study millions and millions of times.
- Conceptually, probability comes into play in a frequentist confidence interval **before collecting the data**.
  - For example, there is a 95% probability that we will collect data that produces an interval that contains the true parameter value.

Introduction to Bayesian Inference

Confidence Intervals vs. Credible Intervals

- We would like to make statements about the probability that the interval contains the true parameter value given the data that we actually observed.
- Probability comes into play in a Bayesian credible interval **after collecting the data**
  - For example, based on the data, we now think there is a 95% probability that the true parameter value is in the interval.
- This is more natural because we want to make a probability statement regarding that data after we have observed it.

Introduction to Bayesian Inference

Confidence Intervals vs. Credible Intervals

- Example: Assume we are interested in the proportion of the population of American college students that sleep at least eight hours each night (θ).
- The Gamecock, at the USC printed an internet article "College Students Don't Get Enough Sleep" (2004).
  - Most students spend six hours sleeping each night.
- University of Notre Dame's paper, "Fresh Writing" (2003).
  - The article reported took random sample of 100 students: "approximately 70% reported to receiving only five to six hours of sleep on the weekdays
  - 28% receiving seven to eight and only 2% receiving the healthy nine hours for teenagers."

Introduction to Bayesian Inference

Confidence Intervals vs. Credible Intervals

- Assumptions:
  - Suppose we collect a random sample of 27 students from UCLA, where 11 students recorded they slept at least eight hours per night.
- Based on this information, we are interested in estimating θ
  - From USC's and UND's studies, we believe it's probably true that most college students get less than eight hours of sleep.
  - We want our prior to assign most of the probability to values of $\theta < 0.5$.
  - From the information given, we decide that our best guess for θ is 0.3, although we think it is very possible that θ could be any value in [0, 0.5]

Introduction to Bayesian Inference

Confidence Intervals vs. Credible Intervals

- Calculations:
- Lets assume that a Beta-Binomial Distribution is appropriate, such that: $X|\theta \sim \text{Binomial}(n, \theta)$

$$\theta \sim \text{Beta}(a, b)$$

→Need to find: $\pi(\theta|X)$ and $\theta|x$:

$$\begin{aligned}
\pi(\theta|x) &\propto p(x|\theta)p(\theta) \\
&\propto \binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \\
&\propto \theta^x(1-\theta)^{n-x}\theta^{a-1}(1-\theta)^{b-1} \\
&\propto \theta^{x+a-1}(1-\theta)^{n-x+b-1} \implies
\end{aligned}$$

$$\theta|x \sim \text{Beta}(x+a, n-x+b).$$

Introduction to Bayesian Inference

Confidence Intervals vs. Credible Intervals

- Calculations:
- Given that $\theta|x \sim \text{Beta}(x+a, n-x+b)$, we need to find *a* and *b*.
  - If given the studies' information, we believe that the median of θ is 0.3 and the 90th percentile is 0.5, we can estimate the unknown values of *a* and *b* by solving:

$$\left.\begin{aligned} \int_0^{0.3} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}\, d\theta = 0.5 \\ \int_0^{0.5} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}\, d\theta = 0.9 \end{aligned}\right\} \text{Use } \textbf{BBsolve} \text{ in R}$$

Introduction to Bayesian Inference

Confidence Intervals vs. Credible Intervals

- R Code:

```
## install the BBsolve package
install.packages("BB", repos="http://cran.r-project.org")
library(BB)
## using percentiles
myfn <- function(shape){
test <- pbeta(q = c(0.3, 0.5), shape1 = shape[1],
 shape2 = shape[2]) - c(0.5, 0.9)
return(test)
}
BBsolve(c(1,1), myfn)
## using quantiles
fn = function(x){qbeta(c(0.5,0.9),x[1],x[2])-c(0.3,0.5)}
BBsolve(c(1,1),fn)
```

→ $a = 3.3$ and $b = 7.2$

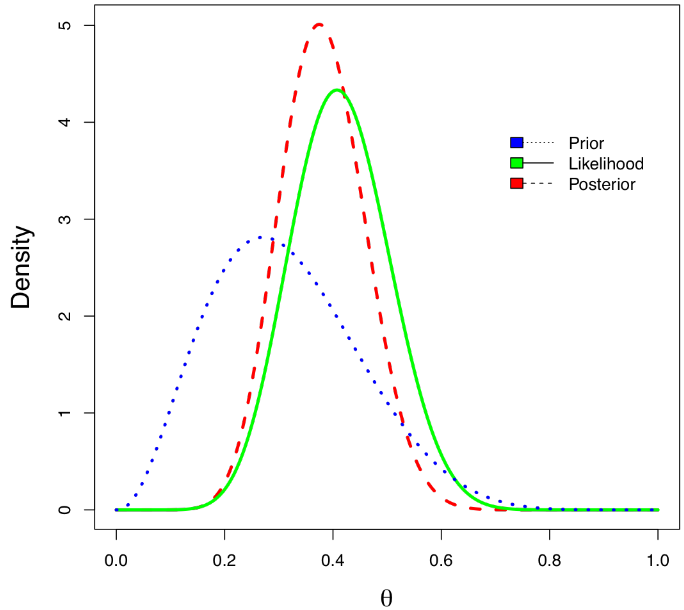
Introduction to Bayesian Inference

Confidence Intervals vs. Credible Intervals

- Model:

  Assume that the prior on θ is a Beta(3.3,7.2):

$$
\begin{aligned}
X \mid \theta &\sim \mathrm{Binomial}(27, \theta) \\
\theta &\sim \mathrm{Beta}(3.3, 7.2) \\
\theta \mid x &\sim \mathrm{Beta}(x + 3.3, 27 - x + 7.2) \\
\theta \mid 11 &\sim \mathrm{Beta}(14.3, 23.2)
\end{aligned}
$$



- Thus, the posterior distribution is

  → $\theta \mid 11 \sim \mathrm{Beta}(11 + 3.3,\ 27 - 11 + 7.2) = \mathrm{Beta}(14.3,\ 23.2)$.

Introduction to Bayesian Inference

Confidence Intervals vs. Credible Intervals

- R Code:

```
th = seq(0,1,length=500)
a = 3.3
b = 7.2
n = 27
x = 11
prior = dbeta(th,a,b)
like = dbeta(th,x+1,n-x+1)
post = dbeta(th,x+a,n-x+b)
quartz()
plot(th,post,type="l",ylab="Density",lty=2,lwd=3,xlab =
expression(theta),col="red",cex.lab=1.5)
lines(th,like,lty=1,lwd=3,col="green")
lines(th,prior,lty=3,lwd=3,col="blue")
legend(0.7,4,c("Prior","Likelihood","Posterior"),lty=c(3,1,2),lwd=c(1,1,1),fill =
c("blue","green","red"),cex=1,bty='n')
```

Introduction to Bayesian Inference

Confidence Intervals vs. Credible Intervals

- Based on the previous results, find a **90% credible interval**
- We need to solve for ***L** (lower)* and ***U** (upper)* such that:

$$P(\theta < L|x) = 0.05 \text{ and } P(\theta > U|x) = 0.05$$

→ Need to solve numerically:

$$\int_0^L \mathrm{Beta}(14.3, 23.3)d\theta = 0.05 \text{ and } \int_U^1 \mathrm{Beta}(14.3, 23.3)d\theta = 0.05$$

```
a = 3.3;b = 7.2
n = 27;x = 11
a.star = x+a
b.star = n-x+b
L = qbeta(0.05,a.star,b.star)
U = qbeta(1-0.05,a.star,b.star)
```

R Code gives us the values:
L= 0.256 and *U= 0.514*

Introduction to Bayesian Inference

Hypothesis Testing

- The frequentist approach to hypothesis testing would compare a null hypothesis H_0 with an alternative H_1 through a test statistic T which typically obtains a larger value when H_1 is true than when H_0 is true. The null hypothesis is rejected with a level α if the observed value of the test statistic, t_{obs}, is larger than the critical value t_c where $Pr(T > t_c | H_0) = \alpha$. The p-value, $p = Pr(T \geq t_{obs} | H_0)$.

- In frequentist statistics, we do not assign probabilities to hypotheses. In particular, the p-value cannot be interpreted as $p(H_0)$.

Introduction to Bayesian Inference

Hypothesis Testing

- In the Bayesian approach, we may assign the prior probabilities p(H) and p(H), and, using Bayes' theorem, compute the posterior probabilities.

$$p(H_i|x) = \frac{p(H_i)p(x|H_i)}{p(H_0)p(x|H_0) + p(H_1)p(x|H_1)}, \quad i = 0, 1$$

- In the frequentist approach it is not absolutely necessary to specify an alternative hypothesis. Further, if an alternative is specified, the p-value is independent of it. In the Bayesian approach, the both hypotheses must be fully specified.

Introduction to Bayesian Inference

Hypothesis Testing

- In the frequentist approach it is not absolutely necessary to specify an alternative hypothesis. Further, if an alternative is specified, the p-value is independent of it. In the Bayesian approach, the both hypotheses must be fully specified.

- One usually computes the posterior odds

$$\frac{p(H_1|x)}{p(H_0|x)} = \frac{p(x|H_1)}{p(x|H_0)} \times \frac{p(H_1)}{p(H_0)}$$

which depends on the data y only through the **Bayes factor**

$$B = \frac{p(x|H_1)}{p(x|H_0)}$$

B is the likelihood ratio of H_0 against H_1 i.e. the odds in favor of H_0 against H_1 that are given by the data. See: H. Jeffreys' scale interpretation

Introduction to Bayesian Inference

Hypothesis Testing

(Example)

Visualization(http://rpsychologist.com/d3/bayes/)

Introduction to Bayesian Inference

Linear Regression Example

(Example)

Visualization(https://towardsdatascience.com/introduction-to-bayesian-linear-regression-e66e60791ea7)