Chapter 16 - Bayesian Inference

- Recall that many probability distributions we use to model our data are governed by one or more parameters.
- Classical statistics treats the parameters as constants: fixed (but often unknown) numbers.
- Bayesian statistics treats the parameters as random variables and places a distribution on the parameters.
- Sometimes this reflects the fact that the parameter could actually vary across samples; in other cases it simply reflects our uncertainty about the parameter value.
- Historically, Bayesian inference dates back to the 1700s when the Rev. Thomas Bayes was performing inference about the binomial probability p.
- He treated p as having a ________ distribution.

- The French mathematician Pierre Laplace did something similar around the same time.

16.2 Bayesian Priors, Posteriors, and Estimators

- Recall the likelihood $L(\theta | y_1, \ldots, y_n)$ is mathematically the same as the joint pdf (or pmf) of the data:

- In Bayesian statistics, we specify a distribution $g(\theta)$ for the parameter θ.
- This is called a prior distribution and is specified before examining the sample data.
- For now, we will assume $g(\theta)$ is a continuous density whose parameter(s) are specified.
- By the definition of conditional and joint distributions, the joint pdf of the data $Y_1, \ldots, Y_n$ and the parameter Θ is:

- Then the marginal pdf (or pmf) of $Y_1, \dots, Y_n$ is

- The conditional density of θ given $y_1, \dots, y_n$ is called the posterior density of $\theta | y_1, \dots, y_n$ and is found by

- Note the resemblance to the Bayes' Rule formula from Sec. 2.10:

- The difference is that θ is continuous and thus the integral replaces the sum.
- The posterior represents the investigator's "updated belief" about the parameter after examining the sample.
- The information in the posterior is a combination of the ________ information and the ________ information.

Example 1: Let $Y_1, \ldots, Y_n$ be iid Bernoulli r.v.'s with $P(Y_i = 1) = p$. We want to estimate the unknown success probability p.

- A natural prior distribution for p would be a ________ distribution since ________.
- The investigator would specify the hyperparameters ________ based on his/her prior beliefs about p.

- The Bernoulli pmf is

So the likelihood is

- The prior $g(p)$ is a beta where p plays the role of the r.v.:

So the joint pdf of $Y_1, \dots, Y_n$ and p is:

- The marginal pdf of $Y_1, \ldots, Y_n$ is:

- So the posterior density of p is:

- We see that this posterior is a ________ pdf with parameters ________ and ________.
- The prior was also a ________.
- When the posterior has the same functional form as the prior (but with updated parameter values), we say the prior is a ________.

Side note: When the prior is a proper density (integrating to 1), the posterior must also be a proper density.

- The <u>kernel</u> of the posterior in this example is the part that

- Note the marginal pdf $m(y_1, \dots, y_n)$ does <u>not</u> depend on p: It is simply a <u>normalizing constant</u> that allows the posterior density to integrate to 1. So: If we can recognize the posterior's kernel

as being the kernel of a <u>known distribution</u>, then we don't actually need to calculate the marginal density.
- Any part of the posterior that does not depend on the target parameter can be treated as "a constant" and not kept track of, saving effort.

<u>However</u>: There is an advantage to finding the marginal density as it can be useful for checking model fit.

<u>Example 1 again:</u> Plant scientists are studying the probability that a certain type of plant will bear fruit. They treat the success probability as varying across plant specimens.

- The scientists guess that the average success probability is around 0.75. What would be a good prior distribution to use?
- Recall the expected value for a beta(α, β) distribution is
- Choices for $g(p)$ that give

- The beta(α, β) variance is

- Suppose they choose a beta (3, 1) prior.
- The scientists observe 20 plants and 18 bear fruit.
- From our previous derivation, the posterior for p is ________ with parameters

  and

  i.e.,

- The usual Bayesian point estimator of p is the posterior mean:

- Note the MLE of p here is the

- The prior mean of p here was

- Note the posterior Bayes estimate is between the prior mean and the sample estimate.

- In general for this model, note the posterior mean is

which is a weighted average of the prior mean and the MLE.
- The MLE is weighted more heavily when the sample is ________.
- This weighted-average property is common for many Bayesian estimators.
- Other point estimators are possible, such as:

but these are best found using a computer.