22.0 Bayesian Statistics

- Answer Questions
- Review Midterm
- Bayesian Inference

22.1 Bayesian Inference

Recall Bayes' Theorem:

$$P(A_1|B) = \frac{P(B|A_1) * P(A_1)}{\sum_{i=1}^{k} P(B|A_i) * P(A_i)}$$

where the $A_1, \ldots, A_k$ are mutually exlcusive and

$$P(A_1 \text{ or } A_2 \text{ or } \cdots \text{ or } A_k) = 1.$$

This is a formalism for how we learn. $P(A_1)$ the prior probability of A_1, given information we have before event B. Then we combine our prior probability with the new information on event B through $P(B|A_1)$ to get our new opinion about the probabilty of A_1, written as $P(A_1|B)$.

$P(A_1)$ is our prior opinion, and $P(A_1|B)$ is our posterior opinion.

To review the use of the formula, remember the breathalyzer example. Suppose 20% of the people on the road after 2 a.m. are legally drunk. Suppose a police officer stops someone at random and admnisters at breathalyzer test. The breathalyzer test has probability .95 of identifying a person who is legally drunk, and probability .1 of misidentifying a person who is not legally drunk.

When the officer makes the stop, his prior probability of drunkeness is .2, the proportion of drivers who are drunk. How should that opinion change if the breathalyzer is positive?

Here A_1 is the event that a person is drunk, and B is the event that the test is positive. The A_2 is the event that the person is not drunk—this is mutually exclusive of A_1, and the probability of A_1 or A_2 is 1.

We apply Bayes' Theorem. Clearly

$$
\begin{aligned}
P(A_1|B) &= P(\ \textbf{Drunk} \mid \textbf{positive test}\ ) \\
&= \frac{P(B|A_1) * P(A_1)}{\sum_{i=1}^{k} P(B|A_i) * P(A_i)} \\
&= \frac{P(\ \textbf{pos. test} \mid \textbf{Drunk}\ ) * P(\ \textbf{Drunk}\ )}{P(\ \textbf{pos} \mid \textbf{D}\ ) * P(\ \textbf{D}\ ) + P(\ \textbf{pos} \mid \textbf{not D}\ ) * P(\ \textbf{not D}\ )} \\
&= \frac{.95 * .2}{.95 * .2 + .1 * .8} \\
&= .7037
\end{aligned}
$$

So after failing the test, the police office should believe you have about a 70% chance of being drunk.

The frequentist paradigm:

- sees science as objective
- defines probability as a long-run frequency in independent, identical trials
- looks at parameters (i.e., the true mean of the population, the true probability of heads) as fixed quantities

This paradigm leads one to specify the null and alternative hypotheses, collect the data, calculate the significance probability under the assumption that the null is true, and draw conclusions based on these significance probabilities using the size of the observed effects to guide decisions.

The Bayesian paradigm:

- sees science as subjective
- defines probability as a subjective belief (which must be consistent with all of one's other beliefs)
- looks at parameters (i.e., the true mean of the population, the true probability of heads) as random quantities because we can never know them with certainty.

This paradigm leads one to specify plausible models, to assign a prior probability to each model, to collect data, to calculate the probability of the data under each model, to use Bayes' theorem to calculate the posterior probability of each model, and to make inferences based on these posterior probabilities. The posterior probabilities enable one to make predictions about future observations and one uses one's loss function to make decisions that minimize the probable loss.

22.2 RU486 Example

The "morning after" contraceptive RU486 was tested in a clinical trial in Scotland. This discussion slightly simplifies the design.

Assume 800 women report to a clinic; they have each had sex within the last 72 hours. Half are randomly assigned to take RU486; half are randomly given the conventional theory (high doses of estrogen and synthetic progesterone).

Among the RU486 group, none became pregnant. Among the conventional therapy group, there were 4 pregnancies. Does this information show that RU486 is more effective than conventional treatment?

We shall compare the frequentist and Bayesian approaches.

If the two therapies (R and C, for RU486 and conventional) are equally effective, then the probability that an observed pregnacy came from the R group is the proportion of women in the R group, or .5; let

$$p = P[\ \textbf{an observed pregnancy came from group R}\ ].$$

A frequentist wants to make a hypothesis test. Specifically,

$$\mathbf{H}_0 : p \geq .5 \quad \textbf{vs.} \quad \mathbf{H}_A : p < .5$$

If the evidence supports the alternative, then RU486 is more effective than conventional treatment.

The data are 4 observations from a binomial, where p is the probability that a pregnancy is from group R.

How do we calculate the significance probability?

The significance probability is the chance of observing a result as or more supportive of the alternative than the one in the sample, when the null hypothesis is true.

Our sample had no children from the RU486 group, which is as supportive as we could have. So

$$P - value = P[\ \textbf{0 successes in 4 tries}\ |\mathbf{H}_0] = (1 - .5)^4 = .0625.$$

Most frequentists would fail to reject, since .0625 > .05.

Suppose we had observed 1 pregnancy in the R group. What would the P-value be then?

In the Bayesian analysis, we begin by listing the models we consider plausible. For example, suppose we thought we had no information a priori about the probability that a child came from the R group. In that case all values of p between 0 and 1 would be equally likely.

Without calculus we cannot do that case, so let us approximate it by assuming that each of the following values for p

$$.1, .2, .3, .4, .5, .6, .7, .8, .9$$

is equally likely. So we consider 9 models, one for each value of the parameter p.

If we picked one of the models, say with $p = .1$, then that means the probability of a sampole pregnancy coming from the R group is .1, and .9 that it comes from the C group. But we are not sure about the model.

Model	Prior	P(data—model)	Product	Posterior
p	P[model]	P[$k = 0$ — p]		P[model — data]
.1	1/9	.656	.0729	.427
.2	1/9	.410	.0455	.267
.3	1/9	.240	.0266	.156
.4	1/9	.130	.0144	.084
.5	1/9	.063	.0070	.041
.6	1/9	.026	.0029	.017
.7	1/9	.008	.0009	.005
.8	1/9	.002	.0002	.001
.9	1/9	.000	.0000	.000
	1		.1704	1

So the most probable of the nine models has $p = .1$. And the probability that $p < .5$ is $.427 + .267 + .156 + .084 = .934$.

Note that in performing the Bayes calculation,

- We were able to find the probability that $p \leq .5$, which we could not do in the frequentist framework.
- In calculating this, we used only the data that were observed. Data that were more extreme than what we observed plays no role in the calculation or the logic.
- Also note that the prior probability of $p = .5$ dropped from 1/9=.111 to .041. This illustrates how our prior belief changes after seeing the data.

Suppose a new person analyzes the same data. But their prior does not put equal weight on the 9 models; they put weight .52 on $p = .5$ and equal weight on the others.

Model	Prior	P(data—model)	Product	Posterior
p	P[model]	P[$k = 0$ — p]		P[model — data]
.1	.06	.656	.0349	.326
.2	.06	.410	.0246	.204
.3	.06	.240	.0144	.119
.4	.06	.130	.0078	.064
.5	.52	.063	.0325	.269
.6	.06	.026	.0015	.013
.7	.06	.008	.0005	.004
.8	.06	.002	.0001	.001
.9	.06	.000	.0000	.000
	1		.1208	1

Compared to the first analyst, this one now believes that the probability that $p = .5$ is .269, instead of .041. So the strong prior used by the second analyst has gotten a rather different result.

But the probability that $p = .5$ had dropped from .52 to .269, showing the evidence is running against the prior belief.

But in practice, what one really needs to know are predictive probabilities. For example, what is the probability that the next pregnancy comes from the RU486 group?

To calculate the predictive probability for the next pregnancy, one finds the weighted average of the different p values, using the posterior probabilities as the weights.

For the second Bayesian, she finds

$$\textbf{predictive probability} = .1 * .326 + .2 * .204 + \ldots .9 * .000 = .281.$$

This is a very useful quantity, and one that cannot be calculated within the frequentist paradigm.