23.0 More Bayesian Statistics

- Answer Questions
- Bayes Review
- GUSTO
- Review Midterm

23.1 Bayesian Inference: Review

In the Bayesian analysis, one lists the possible models and assigns them prior probabilities.

Then one collects the data.

Then one calculates the posterior probability of each model using Bayes' rule.

We considered nine possible situations (models) for the RU486 data:

$$p = .1, .2, .3, .4, .5, .6, .7, .8, \text{ or } .9$$

Each possible value of p is the chance that a pregnancy in the data comes from the RU486 group rather than the conventional therapy group.

For purposes of illustration we assumed that $p = .5$ had prior probability .52, and all the other models were equally likely (prior probability .06). A different analyst might choose different prior probabilities.

Then one collects the data. There were 4 pregnancies, all from the conventional therapy group.

To find the posterior probability that $p = .1$, given the four observed pregnancies, one uses Bayes' rule:

$$
\begin{aligned}
P(p = .1 | 4 \textbf{ in C }) &= \frac{P(\textbf{ 4 in C } | p = .1) * P(p = .1)}{\sum_{i=1}^{9} P(\textbf{ 4 from C } | p = 1/10) * P(p = i/10)} \\
&= \frac{\textbf{(bin: 4 in C for } p\textbf{=.1)} * P[p = .1]}{\sum_{i=1}^{9} \textbf{(bin: 4 in C for } p = \frac{i}{10}\textbf{)} * P[p = \frac{i}{10}]} \\
&= \frac{.656 * .06}{.656 * .06 + .410 * .06 + \cdots + .000 * .06} \\
&= \frac{.0394}{.1211} \\
&= .326
\end{aligned}
$$

The first binomial probability is $\binom{4}{0} * (.1)^0 * (.9)^4 = .6561.$

Model	Prior	P(data—model)	Product	Posterior
p	P[model]	P[$k = 0$ — p]		P[model — data]
.1	.06	.656	.0394	.326
.2	.06	.410	.0246	.204
.3	.06	.240	.0144	.119
.4	.06	.130	.0078	.064
.5	.52	.063	.0328	.269
.6	.06	.026	.0015	.013
.7	.06	.008	.0005	.004
.8	.06	.002	.0001	.001
.9	.06	.000	.0000	.000
	1		.1211	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 this 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. It is the best guess about the probability that the next child will come from the RU486 group.

23.2 GUSTO Study

The GUSTO study illustrates how Bayesian analysis applies to important decisions. It used calculus-based methods to do the same kinds of calculation we did in the RU486 example.

What were some of the key elements in the problem?

- different costs for streptokinase and t-PA
- nearly equivalent effects
- information from previous studies

A major issue in the analysis is that Genentech decided that the clinically meaningful difference was saving one percent more lives than the other. It is unclear how this was arrived at, but it raises the important question of the loss function.

A Bayesian is supposed to make the decision that minimizes the expected loss. In this case one would need to consider the cost of the drug in terms of the value of the residual life of the patient and the probability of successful treatment.

This leads to complex ethical issues. Consider the different points of view (or loss functions) held by the head of Genentech, the head of an insurance agency, a patient with heart disease, and a patient with cancer whose treatment cost equals that of the next 200 t-PA patients.

Bayesian methods are far more general than the toy problems we have raised here. Instead of using just binomial models, the same ideas apply to every other kind of statistical model you can imagine (normal data, regression data, and so forth).

Also, the explicit use of the loss function is something that a decision-maker needs to consider. In the Bayes framework, this is automatic (although we've not really done it). In the frequentist world, it only shows up when one chooses the threshold at which the significance probability is small enough to reject the null.

Finally, the Bayesian can answer questions like: "What is the probability that t-PA is better than steptokinase?" or "What is the probability that t-PA is more than 1% better?" A frequentist analysis cannot.