Introduction to Bayesian Methods

Arthur Small

Spring 2020

SYS 6014 Decision Analysis

Big picture: data-driven decision-making

You have a set of possible actions you can take.

You have a prediction model or classification model that enables you to *differentiate* between cases, based on data.

Your decision model then enables you to discriminate — to make the "right" intervention for each different case, rather than just choosing the same action for all cases.

Predictive models

Predictive models may take many different forms.

- They may use many, many different types of data
- They may be built using different analytic techniques: statistical regression, machine learning,...

We embrace this diversity of approaches.

But: we want our predictions to include *uncertainty* information.

So: want outputs in form of *probability distributions* over decision-state variables

Probabilistic predictions

Bayesian Methods

Bayesian methods: Introduction via simple example

(Related readings: Hoff Ch. 1)

Suppose you want to estimate the fraction of a population that is infected with some disease.

$\theta \in [0, 1]$: true value

Test a randome sample of 20 from the population.

$Y \in \{0, 1, \ldots, 20\}$: # of positive results.

Question: What does realized value of Y tell us about the true value of θ?

$Y|\theta \sim \text{binomial}(20, \theta)$: For $y = 0, 1, \ldots, 20$,

$$l(y|\theta) = \Pr(Y = y|\theta) = \binom{20}{y}\theta^{y}(1-\theta)^{(20-y)}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

$l(y|\theta)$ called the *likelihood function*.

Idea: For any $0 < \theta < 1$, all values of Y are *possible*, but some are more likely than others.

The likelihood function tells us how likely is each possible observation, for a given θ.

If, say, $Y = 15$, that provides evidence that θ is not small.

Prior information

Suppose we have some background knowledge about the likely values of θ.

Represent this knowledge by means of a *prior distribution* $\pi(\theta)$ over $[0, 1]$.

Obviously, there are many (infinitely many) possible such distributions.

For convenience, we typically choose to model priors as chosen from a parametrized family of distributions.

The Beta distribution

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

For our case, let's suppose our prior beliefs correspond to:

$$\theta \sim \text{beta}(2, 20)$$

Bayes Theorem

Let $\pi(\theta|y)$ denote our *posterior distribution* over values of θ.

This means: our *updated* beliefs about the likelihood that θ takes various values, *after* we've received our test results.

Bayes Theorem says:

$$\pi(\theta|y) = \frac{l(y|\theta)\pi(\theta)}{Pr\{Y = y\}} = \frac{l(y|\theta)\pi(\theta)}{\int_{\Theta} l(y|\tilde{\theta})\pi(\tilde{\theta})d\tilde{\theta}}$$

Can be shown:

If $\theta \sim \text{beta}(2, 20)$ and $Y = 0$, then $\theta | y \sim \text{beta}(2, 40)$.

More generally:

If $\theta \sim \text{beta}(a, b)$ and $Y = y$, then $\theta | y \sim \text{beta}(a + y, b + 20 - y)$.