

Session 10

How much it favors very conservative states of decision making when you refuse to quantify uncertainty? you need to quantify uncertainty. What you need to do is taking the uncertainties and quantifying them with the best you can, turning them into the manageable risks and figuring out strategies for managing those risks.

Sometimes you have some kind of model that you use, such as classification model, statistical model, and so on that delivers the output and you need to do some post-processing to convert those outputs to probabilities over the features.

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 random sample of 20 from the population.

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

Question: What does realize the value of Y tells us about the true value of θ ?

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

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

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 $\theta|y \sim \text{beta}(a + y, b + 20 - y)$.