ADVANCED BAYESIAN MODELING

A Simple Bayesian Example

Your doctor orders a diagnostic test for a rare disease.

Suppose the test result is positive.

What are the chances that you actually have the disease?

More information:

- The prevalence of the disease is about 0.01%.
- 99% of people with the disease test positive (*true positive rate* or *sensitivity* of 0.99).
- 95% of people without the disease test negative (*true negative rate* or *specificity* of 0.95).

Let's formalize:

$$\theta = \begin{cases} 1 & \text{if you have the disease} \\ 0 & \text{if not} \end{cases}$$

$$y = \begin{cases} 1 & \text{if you test positive} \\ 0 & \text{if not} \end{cases}$$

Note: y observed, θ not observed

For a randomly-selected person,

$$\Pr(\theta = 1) \quad = \quad 0.0001$$

This is a **marginal** (or **unconditional**) probability.

For a randomly-selected person,

$$\Pr(\theta = 1) \quad = \quad 0.0001$$

This is a **marginal** (or **unconditional**) probability.

The diagnostic test characteristics:

$$\begin{aligned} \Pr(y = 1 \mid \theta = 1) &\quad = \quad 0.99 \\ \Pr(y = 0 \mid \theta = 0) &\quad = \quad 0.95 \end{aligned}$$

These are **conditional** probabilities.

For a randomly-selected person,

$$\Pr(\theta = 1) \quad = \quad 0.0001$$

This is a **marginal** (or **unconditional**) probability.

The diagnostic test characteristics:

$$\begin{aligned} \Pr(y = 1 \mid \theta = 1) &= 0.99 \\ \Pr(y = 0 \mid \theta = 0) &= 0.95 \end{aligned}$$

These are **conditional** probabilities.

The probability you seek:

$$\Pr(\theta = 1 \mid y = 1)$$

By **Bayes' rule**:

$$\begin{aligned}\Pr(\theta = 1 \mid y = 1) &= \frac{\Pr(\theta = 1, y = 1)}{\Pr(y = 1)} \\ &= \frac{\Pr(\theta = 1)\Pr(y = 1 \mid \theta = 1)}{\Pr(y = 1)}\end{aligned}$$

By **Bayes' rule**:

$$\begin{aligned}
\Pr(\theta = 1 \mid y = 1) &= \frac{\Pr(\theta = 1, y = 1)}{\Pr(y = 1)} \\
&= \frac{\Pr(\theta = 1)\Pr(y = 1 \mid \theta = 1)}{\Pr(y = 1)}
\end{aligned}$$

For the denominator:

$$\begin{aligned}
\Pr(y = 1) &= \Pr(y = 1, \theta = 1) + \Pr(y = 1, \theta = 0) \\
&= \Pr(\theta = 1)\Pr(y = 1 \mid \theta = 1) \\
&\quad + \Pr(\theta = 0)\Pr(y = 1 \mid \theta = 0)
\end{aligned}$$

Applying this,

$$
\begin{aligned}
& \Pr(\theta = 1 \mid y = 1) \\
& \qquad = \frac{0.0001 \times 0.99}{0.0001 \times 0.99 \; + \; (1 - 0.0001) \times (1 - 0.95)} \\
& \qquad \approx 0.002
\end{aligned}
$$

Conclusion: There is little chance you have the disease, even if the (fairly accurate) test is positive.

But are you really “randomly-selected”?

Did your doctor have suspicions? Do you have predisposing characteristics (e.g., family history)?

Suppose instead that

$$\Pr(\theta = 1) \;\; = \;\; 0.1$$

Then (try yourself)

$$\Pr(\theta = 1 \mid y = 1) \;\; \approx \;\; 0.69$$

Your chances depend substantially on what you initially assume!

Looking Forward

$$y = \textbf{data} \qquad\qquad \theta = \textbf{parameter}$$

Goal: Gain knowledge of unobserved θ from observed y

- **sampling** (or **data**) **distribution**: conditional distribution of y given θ – how the data would be distributed under different scenarios for θ
- **prior distribution**: marginal distribution of θ – our state of knowledge prior to observing y
- **posterior distribution**: conditional distribution of θ given y – our state of knowledge after observing y