Chapter 7

Bayes in Practice; Random Variables

Lecture 7 | Devore 9e: 2.4–3.1

Recall from last time:

$$\underbrace{\mathbb{P}(A_i \mid B)}_{\text{posterior}} = \frac{\overbrace{\mathbb{P}(B \mid A_i)}^{\text{likelihood}}\ \overbrace{\mathbb{P}(A_i)}^{\text{prior}}}{\sum_j \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}.$$

7.1 Disease testing, finished

Example 7.1 (Rare disease)**.** A disease affects 1% of the population:

$$A = \{\text{person has the disease}\}, \qquad \mathbb{P}(A) = 0.01.$$

The test satisfies $\mathbb{P}(B \mid A) = 0.99$ (sensitivity) and $\mathbb{P}(B^c \mid A^c) = 0.95$ (specificity), where $B = \{\text{test is positive}\}$; hence $\mathbb{P}(B \mid A^c) = 0.05$. A patient tests positive. What is $\mathbb{P}(A \mid B)$?

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A)\mathbb{P}(B \mid A)}{\mathbb{P}(A)\mathbb{P}(B \mid A) + \mathbb{P}(A^c)\mathbb{P}(B \mid A^c)} = \frac{(0.01)(0.99)}{(0.01)(0.99) + (0.99)(0.05)} = \frac{0.0099}{0.0594} = \boxed{\frac{1}{6} \approx 16.7\%}$$

Remark 7.2 (The base-rate trap)**.** The test is right 99% of the time on sick people, yet a positive result means only a 1-in-6 chance of being sick. Here is the back-of-envelope version, and it is worth teaching this way:

$$\text{positive result} \implies \begin{array}{lll} \text{either} & \text{the person is sick} & (1\% \text{ of people}) \\ \text{or} & \text{the test is wrong} & (5\% \text{ of the other } 99\%) \end{array}$$

and $\frac{1}{1+5} = \frac{1}{6}$. Because the disease is rare, the (small) false-positive *rate* acts on a (huge) healthy population, and the false positives outnumber the true positives five to one.

Now check the negative result. With $\mathbb{P}(B^c \mid A) = 0.01$,

$$\mathbb{P}(A^c \mid B^c) = \frac{(0.99)(0.95)}{(0.99)(0.95) + (0.01)(0.01)} = \frac{0.9405}{0.9406} \approx 99.99\%.$$

So for a *rare* disease:

positive $\Rightarrow$ has the disease $\times$ (only 16.7%)
negative $\Rightarrow$ does not $\checkmark$ (99.99%)

Example 7.3 (Common disease — the pattern reverses)**.** Same test, but now $\mathbb{P}(A) = 0.99$. Then

$$\mathbb{P}(A \mid B) = \frac{(0.99)(0.99)}{(0.99)(0.99) + (0.01)(0.05)} = \frac{0.9801}{0.9806} \approx 99.9\%,$$

$$\mathbb{P}(A^c \mid B^c) = \frac{(0.01)(0.95)}{(0.01)(0.95) + (0.99)(0.01)} = \frac{0.0095}{0.0194} \approx 49\%.$$

Now a positive result is conclusive and a *negative* result is a coin flip.

Remark 7.4. The test never changed. What changed is $\mathbb{P}(A)$. This is the sentence to remember: $\mathbb{P}(A \mid B)$ *is not a property of the test; it is a property of the test* ***and*** *the population.*

Remark 7.5 (Careful: the shortcut does not survive the switch)**.** The "$\frac{1}{1+5}$" reasoning above works for the *rare* disease because $\mathbb{P}(A) = 1\%$ is small enough that the healthy group is essentially the whole population, so its false-positive mass is just 5%. Running the same shortcut on the common disease gives

$$\frac{99}{99+5} = \frac{99}{104} \approx 95\%,$$

which is **wrong** — it charges the full 5% false-positive rate to a group that is only 1% of the population. The correct weighting is $\mathbb{P}(A^c)\mathbb{P}(B \mid A^c) = (0.01)(0.05) = 0.0005$, giving 99.9% as computed above. When the base rate is not small, do the arithmetic in full.

7.1.1 Testing twice

Example 7.6. Rare disease again, $\mathbb{P}(A) = 0.01$. Test the patient twice; let B_1, B_2 be the two positive results, and assume the two tests are *conditionally independent* given the true status:

$$\mathbb{P}(B_1 \cap B_2 \mid A) = \mathbb{P}(B_1 \mid A)\mathbb{P}(B_2 \mid A), \qquad \mathbb{P}(B_1 \cap B_2 \mid A^c) = \mathbb{P}(B_1 \mid A^c)\mathbb{P}(B_2 \mid A^c).$$

Then

$$\mathbb{P}(A \mid B_1 \cap B_2) = \frac{\mathbb{P}(B_1 \mid A)\mathbb{P}(B_2 \mid A)\mathbb{P}(A)}{\mathbb{P}(B_1 \mid A)\mathbb{P}(B_2 \mid A)\mathbb{P}(A) + \mathbb{P}(B_1 \mid A^c)\mathbb{P}(B_2 \mid A^c)\mathbb{P}(A^c)} = \frac{(0.99)^2(0.01)}{(0.99)^2(0.01) + (0.05)^2(0.99)}.$$

Numerically $\dfrac{0.009801}{0.009801 + 0.002475} \approx \boxed{80\%}$.

One positive: 16.7%. Two positives: 80%. This is why screening programs re-test rather than trusting a single result.

7.2 Random variables

We rarely care about the outcome ω itself; we care about some *number* attached to it. That number is a random variable.

Example 7.7. $S = \{\text{heart}, \text{diamond}, \text{spade}, \text{club}\}$ and

$$X(\text{heart}) = X(\text{diamond}) = 0, \qquad X(\text{spade}) = X(\text{club}) = 1.$$

The new "sample space" — the set of values — is $\widehat{S} = \{0, 1\}$.

Example 7.8. A gambler plays n games: $S = \{W, L\}^n$, so $|S| = 2^n$. Let X be the number of wins. Then $X(WLL \cdots L) = 1$, $X(LWLL \cdots L) = 1$, and $\widehat{S} = \{0, 1, 2, \ldots, n\}$, of size $n + 1$.

Note the compression: 2^n outcomes collapse to $n + 1$ values. That is the point — a random variable throws away the information we do not need.

Example 7.9 (Continuous)**.** Throw a dart at the board $S = \{(x, y) : |x| \leq 1,\ |y| \leq 1\}$ and record the distance from the center:

$$X(x, y) = \sqrt{x^2 + y^2}, \qquad \widehat{S} = [0, \sqrt{2}\,].$$

Definition 7.10 (Random variable)**.** Let $(S, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X : S \to \mathbb{R}$ is a **random variable** if for every $x \in \mathbb{R}$,

$$\{\omega \in S : X(\omega) \leq x\} \in \mathcal{F}.$$

Remark 7.11. Why the condition? Because $\mathbb{P}$ is only defined on $\mathcal{F}$. If the set $\{X \leq x\}$ were not an event, the symbol $\mathbb{P}(X \leq x)$ would be meaningless — and that symbol is the entire subject of the next lecture. When $\mathcal{F} = \mathcal{P}(S)$ (any finite or countable S) the condition is automatic, so in this course it is never an obstacle; it is there so the definition also works on $\mathbb{R}$.

Remark 7.12 (Notation)**.** In lecture I often write ξ (xi) for a random variable, and so do several homework problems; the textbook and the exams write X. They mean the same thing. Uppercase is the random variable (a *function*); lowercase x is a particular value it might take (a *number*). Writing $\mathbb{P}(X \leq x)$ is meaningful; writing $\mathbb{P}(x \leq x)$ is not.