MDTS4411
Module 1
Bayesian Data Analysis

Dr. Durba Bhattacharya, St. Xavier's College (Autonomous), Kolkata.

Prerequisites to this Course

- Basic knowledge of Descriptive Statistics and Probability.
- Statistical Inference.
- Familiarity with calculus.

Overview of the Topics to be covered in this lecture

1. Independence of events and Conditional probability

2. Introduction – Bayes' Theorem and Applications

3. Ideas of Prior, Posterior and Predictive Probabilities

Independent Events:

Two events A and B are independent if the occurrence or non-occurrence of one of the events has no influence on the occurrence or non-occurrence of the other event.

Probabilistically, Two events A and B are independent if $Pr(A \cap B) = Pr(A) \times Pr(B)$

Example: P(breast cancer in India)= 0.14

P(BRCA1 gene in India) = 0.255

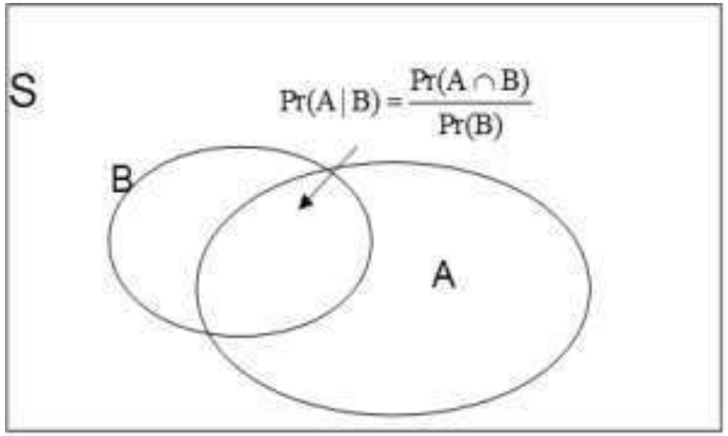
P(breast cancer ∩ BRCA1 gene in India) = 0.166

Clearly, P(breast cancer in India) X P(BRCA1 gene in India) = 0.14 X 0.255 = 0.036 ≠ P(breast cancer ∩ BRCA1 gene in India)

Conditional Probability

Conditional probabilities allow us to understand how the probability of an event A changes after it has been learned that some other event B has occurred.

The key concept for thinking about conditional probabilities is that the occurrence of B reshapes the sample space for subsequent events.



By definition: If A and B are two events such that $P(B) > 0$, then

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(\text{breast cancer} \mid \text{BRCA1 gene in India}) = \frac{P(\text{breast cancer} \cap \text{BRCA1 gene in India})}{P(\text{BRCA1 gene in India})}$$

$$= \frac{0.166}{0.255} = 0.651$$

Properties of Conditional Probability

1. **The Conditional Probability for Independent Events:** If A and B are independent then P(A|B) = P(A)

2. **The Multiplication Rule for Conditional Probabilities:** In an experiment involving two non-independent events A and B:

$$\Pr(A \cap B) = \Pr(B)\,\Pr(A|B) \text{ or}$$

$$\Pr(A \cap B) = \Pr(A)\,\Pr(B|A)$$

3. The set of events $\{A_1, A_2, \ldots, A_n\}$ are partition of sample space S,

where $\bigcup_{i=1}^{n} A_i = S$, then $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$.

Example:

D: *Diseased State*

$\mathbf{D^c}$:Disease not Present

P(E_1): Prevalence rate of the disease in the population.

+ : the result of a single clinical test is positive (suggesting the presence of the disease)

- : the result of the clinical test is negative (suggesting the absence of the disease)

$$P(+ \mid D) = \mathbf{0.900}, \qquad P(- \mid D^c) = \mathbf{0.875}$$

Sensitivity: the ability of a test to correctly identify patients with a disease

Specificity: the ability of a test to correctly identify people without the disease

Quick MCQs

1. A diagnostic test has very high specificity but low sensitivity.
Which of the following statements is correct?

A. The test rarely misses the disease
B. A positive result is likely to be false
C. Healthy individuals are rarely labelled as diseased
D. The test is suitable for mass screening

2. A screening test for a serious disease is designed so that almost no diseased individuals are missed, even if some healthy individuals are wrongly flagged.
Which property of the test must be high?

A. Specificity
B. Sensitivity
C. Positive predictive value
D. Accuracy

D = person actually has the disease
$\mathbf{D^c}$ = person does not have the disease
+ = test is positive
- = test is negative

P(D|+) = Positive Predictive Value **(PPV)** = probability the patient really has the disease given a positive test.

P(D|-) = 1 – Negative Predictive Value **(NPV)** = probability the patient has the disease given a negative test.

Consider the expression: **P(D|+)** – **P(D|-)** = **PPV – (1 – NPV)**

P($\mathbf{D^c}$ | -)

P(D|+) - P(D|-) : Measures the diagnostic power of the test.

Positive Predictive Value (PPV) = P(D|+) = $\frac{P(D)P(+|D)}{P(D)P(+|D)+P(D^c)P(+|D^c)}$

Negative Predictive Value (NPV) = P(D^c |-) = $\frac{P(D^c)P(-|D^c)}{P(D)P(-|D)+P(D^c)P(-|D^c)}$

Note that
Even with: Sensitivity = 99%, Specificity = 99%
If prevalence = 0.1%:
Most positives will be false
PPV will be low
This cannot be seen from sensitivity alone.

Sensitivity and specificity are intrinsic properties of a diagnostic test, whereas PPV and NPV depend on disease prevalence and describe the probability of true disease status given a test result.

$P(D|+) - P(D|-)$: Measures the diagnostic power of the test.

What does this difference really tell us?

1. Before any test, our best guess about whether the patient has the disease is simply the prevalence **$P(D)$**

2. After a positive test, our updated belief is **$P(D|+)$ = PPV**

$P(D^c \mid -)$

3. After a negative test, our updated belief is **$P(D|-) = 1 -$ NPV**

4. The quantity **PPV** – (**1 – NPV**) tells us how far apart the two possible posteriors are. → It measures how much the test result swings our belief in either direction.

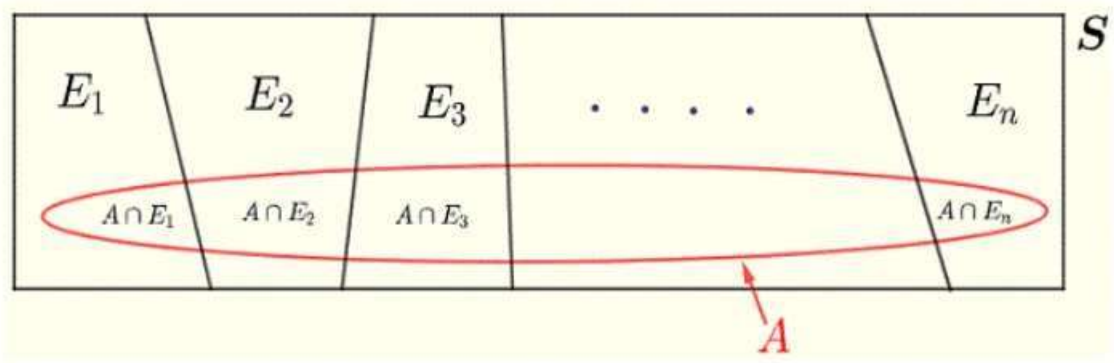
Bayes' Theorem

For any finite partition $\{E_j, j \in J\}$, of Ω, such that $P(E_j) > 0 \ \forall j$

The event A is such that P(A) > 0. $A = \bigcup_{j \in J}(A \cap E_j)$

$$P(E_i|A) = \frac{P(E_i \cap A)}{P(A)}$$

$$= \frac{P(A|E_i)P(E_i)}{\sum_{j \in J} P(A|E_j)P(E_j)}$$



Bayes' theorem may be written in the form $P(E_i|A) \propto P(A|E_i)P(E_i)$

Proportionality constant $[P(A)]^{-1} = \sum_{j \in J} P(A|E_j)P(E_j)$

APPLICATION

Initial beliefs

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{\boxed{\sum_{j\in J} P(A|E_j)P(E_j)}}$$

Modified Beliefs given the Data

Proportionality constant $P(A) = \sum_{j\in J} P(A|E_j)P(E_j)$

Suppose E_j; j∈ J: set of possible diseases from which a patient may be suffering

A: relevant piece of evidence, or data: Here, the outcome of a clinical test.

Having specified P(A | E_i) and P(Ei), the mechanism of the theorem provides a solution to the problem of how to learn from data.

Prior, Posterior And Predictive Probabilities

reflects how beliefs
about the data, D, vary over the different
underlying hypotheses,
thus defining the "relative likelihoods" of
the latter.

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{\sum_{j\in J} P(A|E_j)P(E_j)}$$

If $\{E_j, j \in J\}$ are exclusive and exhaustive events (hypotheses), then for any event (data) A,

(i) $P(E_j)$, $j \in J$, are called the prior probabilities of the E_j, $j \in J$

(ii) $P(A \mid E_j)$, $j \in J$, are called the likelihoods of the E_j, $j \in J$, given A;

(iii) $P(E_j|A)$, $j \in J$, are called the posterior probabilities of the E_j, $j \in J$

(iv) P(A) is called the predictive probability of A implied by the likelihoods and the prior probabilities.

Bayesian Spam Filtering

Uses Bayes' Theorem to determine the probability that any given email is spam

Given an email (yet to be classified as spam/non-spam)

and a list of words that appear frequently in spam emails.

Bayesian spam filtering calculates the individual probabilities of the email containing each suspicious word.

$$P(\text{Spam}|\text{Word}) = \frac{P(\text{word}|\text{spam}).P(\text{spam})}{P(\text{word})}$$

Bayesian Search as an Application of Bayes' Theorem

$$P(L_i|A) = \frac{P(L_i)\times P(A|L_i)}{\sum_j P(L_j)\times P(A|L_j)}$$

Where L_i : Location/ State i

$P(L_i)$: Prior probability that the target is at L_i

A: No detection after searching some area

$P(A \mid L_i)$: likelihood of not detecting the target, given it is in L_i

Detection probability $d_i = 1 - P(A \mid L_i)$

Bayesian search uses prior knowledge + new evidence to continuously update the probability of where the target is.

Bayesian Search as an Application of Bayes' Theorem

$$P(L_i|A) = \frac{P(L_i)\times P(A|L_i)}{\sum_j P(L_j)\times P(A|L_j)}$$

Depends on geography, weather, last known position etc.

Where L_i : Location/ State i

$P(L_i)$: Prior probability that the target is there

A: No detection after searching some area

$P(A \mid L_i)$: likelihood of not detecting the target, given it is in L_i

Likelihood depends on detection capability (sonar, radar, drones).

Detection probability $d_i = 1 - P(A \mid L_i)$

Bayesian Search as an Application of Bayes' Theorem

$$P(L_i|A) = \frac{P(L_i)\times P(A|L_i)}{\sum_j P(L_j)\times P(A|L_j)}$$; Where L_i : Location/ State i

$P(L_i)$: Prior probability that the target is there

A: No detection after searching some area

$P(A \mid L_i)$: likelihood of not detecting the target, given it is in L_i

Detection probability $d_i = 1 - P(A \mid L_i)$

Ex: Drones fly over an area and take images. If the area is covered by trees or at nighttime, the drone misses things. d might be 0.80 in open fields, but may be only 0.40 in forests.

Likelihood depends on detection capability ⇒ Probability of detecting (or missing) the target in a region depends on how powerful your search technology is.

These detection probabilities directly determine the likelihood term in Bayes' theorem and therefore affect how your posteriors change after each search.

Thinking Assignment!

Bayesian Search as an Application of Bayes' Theorem

$$P(L_i|A) = \frac{P(L_i)\times P(A|L_i)}{\sum_j P(L_j)\times P(A|L_j)},$$ Where $P(L_i)$: Prior probability that the target is at L_i

A: No detection after searching some area

$P(A \mid L_i)$: likelihood of not detecting the target, given it is in L_i

Bayesian search cycle -

1. Start with prior probabilities over locations.

2. Search the most promising area.

3. If the target is not found, update the probabilities.

4. Choose the next area with highest posterior.

5. Repeat.

Example 1: Both Region **A** and Region **B** are searched using a weak equipment.

Given: P(A)=0.6, P(B)=0.4

Detection probabilities - In A: d_A = 0.20, In B: d_B = 0.20. Suppose we find nothing in either region.

Step 1: Likelihoods: P(No detect | A) = 1− d_A = 1 − 0.20 = 0.80

P(No detect | B) = 1− d_B = 1−0.20 = 0.80

Step 2: Total probability of no detection = P(No detect) = 0.8 x (0.6)+0.8 x (0.4) = 0.80

Step 3: Posterior for Region A: P(A | No detect) = $\frac{0.8 \times (0.6)}{0.80} = 0.6$

Interpretation: Even after searching Region A and failing to find the object:

- Almost no change (P(A)=0.6 → P(A | No detect) =0.6)
  because both tools are weak, so not detecting anything gives little information.

- The decrease is small because the tool is weak; even if the object *were* in A, there was an 80% chance of missing it.

  So A still remains the most likely region.

Example 2: Strong Tool in A ($\boldsymbol{d_A}$ = 0.90, $\boldsymbol{d_B}$ = 0.20)

Given: P(A)=0.6, P(B)=0.4 [Same as Example 1]
Detection probabilities - In A: $\boldsymbol{d_A}$ = 0.90 (Strong), In B: $\boldsymbol{d_B}$ = 0.20 (Weak). We search both the regions and find nothing!

Step 1: Likelihoods: P(No detect | A) = 1− $\boldsymbol{d_A}$ = 1 − 0.90 = 0.10
P(No detect | B) = 1− $\boldsymbol{d_B}$ = 1−0.20 = 0.80

Step 2: Total probability of no detection = P(No detect) = 0.10 x (0.6)+0.8 x (0.4) = 0.38

Step 3: Posterior for Region A: P(A | No detect) = $\frac{0.10 \times (0.6)}{0.38}$ = 0.158

Interpretation: Probability drops from P(A) = 0.60 → P(A | No detect) = 0.158 [a big change!]

Reason: A strong tool would almost certainly detect the object in **A** if it were there. Since it didn't, we revise belief against **A**.

Example 3: A lost drone is believed to be in one of three zones: **A**, **B**, or **C**.

Suppose we search Zone A, but do not find the drone.

Zone	Prior ($P(L_i)$)	Detection probability (d_i)
A	0.5	0.6
B	0.3	0.4
C	0.2	0.3

Step 1: Compute likelihoods of "no detection":

P(No Detect | L_A) = 1 – 0.6 = 0.4

P(No Detect | L_B) = 1– 0.4 = 0.6

P(No Detect | L_C) = 1 – 0.3 = 0.7

Step 2: Apply Bayes' theorem - P(No Detect) = 0.4 x (0.5) + 0.6 x (0.3) + 0.7 x (0.2) = 0.52

Posteriors: P(L_A | No detect) = 0.3846 ⇒ Probability for Zone A dropped from 0.50 to 0.38

P(L_B | No detect) = 0.3461 ⇒ Zone B increased from 0.30 to 0.35

P(L_C | No detect) = 0.2692 ⇒ Zone C increased from 0.20 to 0.27 → **Search next in Zone B**, **not A!**

Real-World Applications of Bayesian Search Algorithm

1. Search for MH370: Bayesian drift modelling of debris.
2. Search in robotics: localisation of robots in unknown environment.
3. Cybersecurity: locating intrusion source.
4. Industrial systems: fault localisation.
5. Bayesian Search ingredients ⇒ priors + detection likelihoods + iterative updating.
6. Always search next in area with highest posterior.
7. Useful for complex search-and-rescue and engineering problems.

References

1. Casella, G., & Berger, R. L. (2002). Statistical inference (2nd ed.). Duxbury Press.
2. DeGroot, M. H., & Schervish, M. J. (2012). Probability and statistics (4th ed.). Pearson.
3. Downey, A. B. (2013). Think Bayes: Bayesian statistics in Python. O’Reilly Media. https://greenteapress.com/thinkbayes/
4. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2nd ed.). Springer. https://www.statlearning.com/
5. Stone, L. D. (2012). Bayesian search theory (2nd ed.). Wiley.
6. Online References
7. Hajek, A. (2023). Bayesian epistemology. In The Stanford Encyclopedia of Philosophy (Fall 2023 ed.). https://plato.stanford.edu/entries/epistemology-bayesian/
8. Graham, P. (2002). A plan for spam. https://paulgraham.com/spam.html