Review

# Mini Review:
## Bayes' Theorem and Independence by 3Blue1Brown

*Reviewed by Swati Sircar*

Bayes' theorem (15:45min):
https://www.youtube.com/watch?v=HZGCoVF3YvM

The quick proof of Bayes' theorem (3:47min):
https://www.youtube.com/watch?v=U_85TaXbeIo

3Blue1Brown is a YouTube channel that discusses various topics and problems from undergraduate level mathematics. The key features are very good visuals – animated pictures and diagrams that illustrate various concepts – and balanced, logical arguments.

Bayes' Theorem is one of the most counter intuitive results in probability. Many struggle to make sense of the complicated formula of this theorem,

$$P(H|E) = \frac{(P(H) \times P(E|H))}{P(H) \times P(E|H) + P(H') \times P(E|H')}$$

where $H$ and $E$ are two events of interest, P($H$) is the probability of $H$ happening, P$(E|H)$ is the conditional probability of $E$ happening given $H$ has happened, $H'$ is the complement of $H$ (which is referred to as "¬H" in the video), etc. This theorem is widely applied in various fields and is useful for anyone especially during a pandemic!

Let us say $H$ is the event that you have Covid and $E$ is the event that you have been tested positive. Now, P$(H)$ or probability of you having Covid can be computed based on empirical data and P$(H')$ = 1 – P$(H)$.

P$(E|H)$ i.e. the probability of testing positive if you actually have Covid should be very high and P$(E|H')$ i.e. probability of false positive should be low. (Otherwise the test would not be recommended!) Now, if you do get tested positive, then you would be interested in the probability that you actually have Covid or P$(H|E)$. It may seem that this probability should also be quite high. But if P$(H)$ i.e. the proportion of people actually infected is low, say 5% then P$(H|E)$ can actually be closer to half.

Many of the formulas and results in probability can be understood through Venn diagrams. That's exactly what the above-mentioned videos do. They make the connection between probability and proportion and then take the help of geometry to represent the proportions as areas and lengths. Thus, they provide a strong intuitive understanding of the heart of the matter from which the seemingly complicated formula can be easily derived.

More importantly, the first video starts by mentioning uses of this result, followed by outlining the levels of understanding as:

i. What is it saying?

ii. Why is it true?

iii. When is it useful?

Then it exemplifies with a case and numbers (not percentages) along with aptly drawn visuals. This is followed by the transition from numbers or counts to area and how that helps generalise the result – all within 5 minutes! It then systematically derives the formula with the example, counts, and visuals.

The second video, which is like a footnote for the first video, deals with P$(H$ and $E)$ or the probability of both events happening. As a result, it also touches upon the notion of independence. This is also illustrated with a suitable example.

## How does this help a teacher?

It helps mainly in three ways:

First, by providing a visual than can be used to model any situation involving conditional probability and can help one understand what such probabilities really mean (Figure 1).

1. It uses the unit square as a sample space which can be partitioned in vertical strips given any partition $E_1, E_2 \ldots E_k$ such that the widths of the strips are proportional to their probabilities. Note that both the probabilities of the partition and the widths of the strips add up to 1.
2. Now each strip can be split into two strips (top and bottom) according to the conditional probabilities P$(A|E_1)$, P$(A|E_2) \ldots$ P$(A|E_k)$. That is, the vertical dimensions of the bottom strips should be proportional to these conditional probabilities.
3. So, the bottom strips represent the events $A$ and $E_1$, $A$ and $E_2 \ldots A$ and $E_k$ whose areas are P$(E_1) \times$ P$(A|E_1)$, P$(E_2) \times$ P$(A|E_2) \ldots$ P$(E_k) \times$ P$(A|E_k)$, respectively. Also, event A is represented by the collection of the bottom strips.
4. Now, Bayes' theorem can be deduced by considering the sum of the areas of the bottom strips i.e. P$(A)$ = P$(E_1) \times$ P$(A|E_1)$ + P$(E_2) \times$ P$(A|E_2) + \ldots +$ P$(E_k) \times$ P$(A|E_k)$ and the area of any of the bottom strips i.e. P$(E_i) \times$ P$(A|E_i)$ for any $i = 1, 2 \ldots k$.
5. So, if the heights (vertical dimension) of the bottom strips vary, then the partition has some effect on the event $A$. And therefore, $A$ is not independent of the partition. But if the heights are same for all the bottom strips, then the partition has no effect on $A$. In that case, $A$ is independent of the partition and the common height can be factored out as P$(A)$. So, in that case, P$(A$ and $E_i)$ = P$(E_i) \times$ P$(A)$.

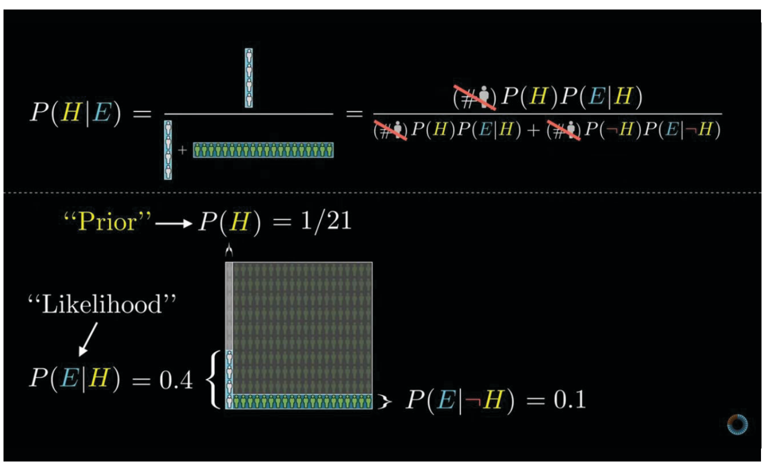


Figure 1

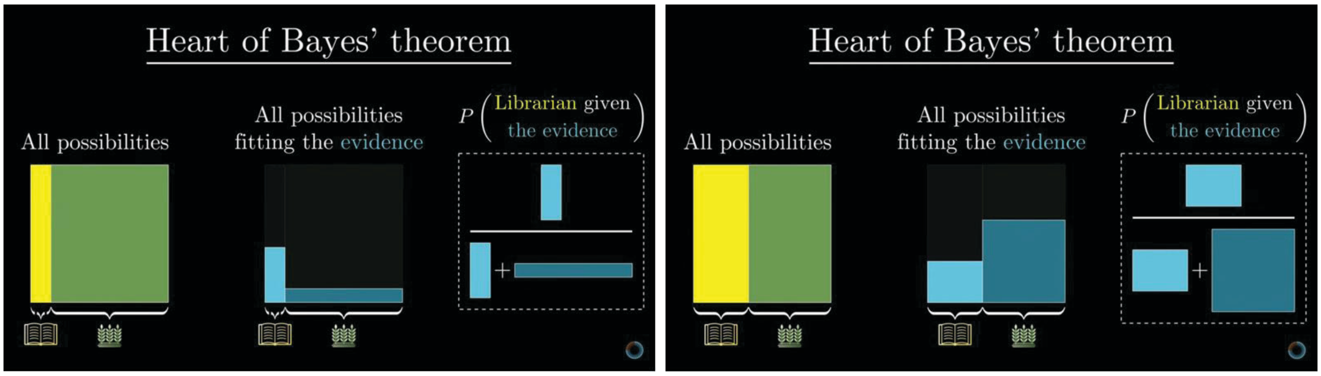


Figure 2

Second, it also draws attention to the fact that people understand proportions better in terms of odds (example: 7 out of 25) than as percentages (e.g. 28%), which is a very useful pedagogic tip.

Finally, it wraps up by linking this theorem to beliefs *(H)* and evidence *(E)* and how the latter should influence the former.

The second video discusses why there is a wide misconception that P*(H* and *E)* = P*(H)* × P*(E)* (regardless of independence).

These videos will undoubtedly help the teacher explain concepts in probability with visual representation and logical reasoning, both of which exemplify mathematical processes.

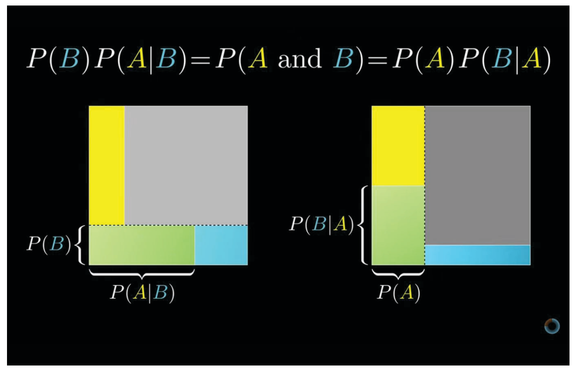


Figure 3

**SWATI SIRCAR** is Assistant Professor at the School of Continuing Education and University Resource Centre of Azim Premji University. Math is the 2nd love of her life (1st being drawing). She is a B.Stat-M.Stat from Indian Statistical Institute and an MS in math from University of Washington, Seattle. She has been doing mathematics with children and teachers for more than a decade and is deeply interested in anything hands on - origami in particular.