

STAT 238 - Bayesian Statistics

Lecture Seven

Spring 2026, UC Berkeley

Aditya Guntuboyina

04 Feb 2026

1 Derivation of the Rules of Probability for Subjective Probability

In the last class, we sketched the derivation of the rules of probability from logical consistency **without** relying on any imagined frequency considerations for defining probability. The argument (taken from Jaynes [1, Chapter 2]) proceeded in the following way.

We denoted the “plausibility” of a proposition A given some information in the form of proposition B by $(A | B)$. We assumed that these plausibilities take values in the set of real numbers (no restriction to be in the interval $[0, 1]$) and that a higher value of plausibility represents greater belief.

To deduce the usual product rule of probability, we assumed the existence of a continuous coordinate-wise increasing function F of two real variables such that

$$(AB | C) = F((B | C), (A | BC))$$

for all A, B, C . This function can then be employed in two different sequences to calculate $(ABC | D)$ for four propositions A, B, C, D :

$$\begin{aligned}(ABC | D) &= F(F((C|D), (B|CD)), (A|BCD)) \\ (ABC | D) &= F((C|D), F((B|CD), (A|BCD))).\end{aligned}$$

It is therefore natural to require that the function F should be such that the right hand sides of the above two equations produce the same answer. From this, one gets the condition:

$$F(F(x, y), z) = F(x, F(y, z)) \text{ for all real numbers } x, y, z.$$

As proved in Jaynes [1, Section 2.1], the only functions F which satisfy the above equation are of the form

$$F(x, y) = w^{-1}(w(x)w(y)) \tag{1}$$

for a positive continuous increasing function w . Note that the function w is not unique. For example, if w satisfies (1), then $\tilde{w}(x) := (w(x))^\alpha$ (which means $\tilde{w}^{-1}(u) = w^{-1}(u^{1/\alpha})$) also satisfies (1) for every $\alpha > 0$. It can be proved that there is uniqueness up to this power transformation.

From here, we derived the following:

1. $w(AB | C) = w(A | C)w(B | AC) = w(B | C)w(A | BC)$.

2. $w(A | C)$ always lies between 0 and 1 with $w(\text{impossible}) = 0$ and $w(\text{certain}) = 1$.

In other words, the function w applied to the plausibilities leads to quantities which satisfy the first two rules of probability.

Next the goal is to derive the sum rule. Here we first assume that there exists a function $S : [0, 1] \rightarrow [0, 1]$ such that

$$w(A^c | C) = S(w(A | C))$$

for all A and C . Note that we are working with $w(A | C)$ instead of the raw plausibilities $(A | C)$. This allows us to use the product rule which has already been derived. We can assume that S is strictly decreasing, $S(0) = 1$, $S(1) = 0$ and that S is continuous.

To set up the characterizing equation for $S(\cdot)$, consider the setting of Figure 1. We are looking at A and B such that B^c is contained in A (or equivalently, A^c is contained in B).

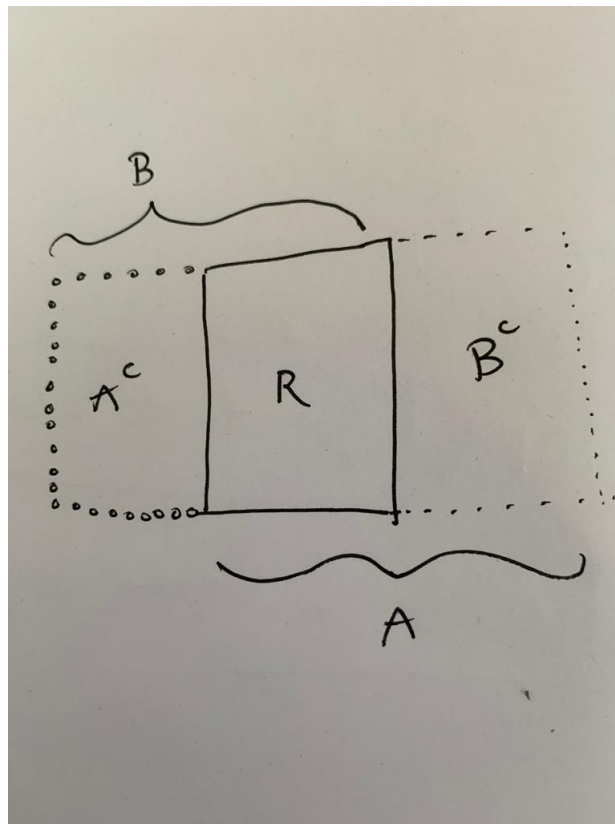


Figure 1: Setting for deriving the Sum Rule

In this setting, there are two different ways of calculating the plausibility of the proposition $R = AB$ in terms of $x := w(A)$ and $y := w(B)$ and the function S . Both these calculations use the product rule. The first method for calculating $w(R) = w(AB)$ is:

$$\begin{aligned} w(AB) &= w(A)w(B | A) \\ &= w(A)S(w(B^c | A)) \\ &= w(A)S\left(\frac{w(B^c)}{w(A)}\right) \\ &= w(A)S\left(\frac{S(w(B))}{w(A)}\right) = xS\left(\frac{S(y)}{x}\right). \end{aligned}$$

The second method for calculating $w(R) = w(AB)$ simply switches the roles of A and B in

the first method:

$$\begin{aligned}
w(AB) &= w(B)w(A | B) \\
&= w(B)S(w(A^c | B)) \\
&= w(B)S\left(\frac{w(A^c)}{w(B)}\right) \\
&= w(B)S\left(\frac{S(w(A))}{w(B)}\right) = yS\left(\frac{S(x)}{y}\right).
\end{aligned}$$

It is therefore natural to assume that S satisfies:

$$xS\left(\frac{S(y)}{x}\right) = yS\left(\frac{S(x)}{y}\right).$$

Recall that here $x = w(A)$ and $y = w(B)$. The setting is such that x and y cannot be completely arbitrary. Indeed because $B^c \subseteq A$, we must have

$$w(B^c) \leq w(A) \quad \text{or equivalently} \quad S(y) \leq x.$$

Our condition on S is therefore

$$xS\left(\frac{S(y)}{x}\right) = yS\left(\frac{S(x)}{y}\right) \quad \text{for all } 0 \leq x, y \leq 1 \text{ with } 0 \leq S(y) \leq x.$$

It is now proved in Jaynes [1, Section 2.2] that the above condition implies that

$$S(x) = (1 - x^\alpha)^{1/\alpha} \quad \text{for } x \in [0, 1]$$

for some $\alpha > 0$. This α is uniquely determined by S .

We have thus proved that

$$w(A^c|C) = (1 - w^\alpha(A|C))^{1/\alpha}$$

which is equivalent to

$$w^\alpha(A^c|C) = 1 - w^\alpha(A|C) \quad \text{or} \quad w^\alpha(A^c|C) + w^\alpha(A|C) = 1.$$

It can now be noted that the first two rules that are satisfied by $w(A|B)$ are also satisfied by $w^\alpha(A|B)$. Thus $w^\alpha(A|B)$ satisfies all the three rules

1. $0 \leq w^\alpha(A|B) \leq 1$, $w^\alpha(\text{impossible}) = 0$, and $w^\alpha(\text{certain}) = 1$,
2. $w^\alpha(AB|C) = w^\alpha(A|C)w^\alpha(B|AC) = w^\alpha(B|C)w^\alpha(A|BC)$, and
3. $w^\alpha(A^c|C) + w^\alpha(A|C) = 1$.

Denoting w^α by \mathbb{P} , we have

1. $0 \leq \mathbb{P}(A|B) \leq 1$, $\mathbb{P}(\text{impossible}) = 0$, and $\mathbb{P}(\text{certain}) = 1$,
2. **Product Rule:** $\mathbb{P}(AB|C) = \mathbb{P}(A|C)\mathbb{P}(B|AC) = \mathbb{P}(B|C)\mathbb{P}(A|BC)$, and
3. **Sum Rule:** $\mathbb{P}(A^c|C) + \mathbb{P}(A|C) = 1$.

We have thus derived the rules of probability without invoking any relationship between probability and long-run frequency.

To summarize: If we argue in terms of plausibilities but we take care to ensure some natural constraints for logical consistency, then we cannot manipulate the plausibilities arbitrarily but have to reason according to the usual rules of probability after an appropriate transformation (this transformation is given by the function w^α).

The sum rule of probability is usually stated as

$$\mathbb{P}(A \cup B|C) = \mathbb{P}(A|C) + \mathbb{P}(B|C) - \mathbb{P}(AB|C) \quad (2)$$

where $A \cup B$ denotes the proposition “at least one of A and B is true” (Jaynes uses the notation $A + B$ for $A \cup B$). (2) can be derived from the stated rules as

$$\begin{aligned} \mathbb{P}(A \cup B|C) &= 1 - \mathbb{P}((A \cup B)^c|C) \\ &= 1 - \mathbb{P}(A^c \cap B^c|C) \\ &= 1 - \mathbb{P}(B^c|A^c C)\mathbb{P}(A^c|C) \\ &= 1 - (1 - \mathbb{P}(B|A^c C))\mathbb{P}(A^c|C) \\ &= 1 - \mathbb{P}(A^c|C) + \mathbb{P}(B|A^c C)\mathbb{P}(A^c|C) \\ &= \mathbb{P}(A|C) + \mathbb{P}(A^c B|C) \\ &= \mathbb{P}(A|C) + \mathbb{P}(B|C)\mathbb{P}(A^c|BC) \\ &= \mathbb{P}(A|C) + \mathbb{P}(B|C)(1 - \mathbb{P}(A|BC)) \\ &= \mathbb{P}(A|C) + \mathbb{P}(B|C) - \mathbb{P}(B|C)\mathbb{P}(A|BC) = \mathbb{P}(A|C) + \mathbb{P}(B|C) - \mathbb{P}(AB|C). \end{aligned}$$

References

- [1] Jaynes, E. T, (2003) *Probability theory: the logic of science*. Cambridge University Press, 2003.