Bayes theorem

Edoardo Milotti

January 18, 2025

1 Introduction

In probability theory, the simplest form of Bayes' theorem is

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \tag{1}$$

If we take a set of mutually exclusive (composite) events $\{B_n\}$ such that $\cup_n B_n = \Omega$, then

$$\begin{aligned} P(A) = P(A \cap \Omega) = P\left[A \cap \left(\bigcup_n B_n\right)\right] = P\left(\bigcup_n A \cap B_n\right) &= \sum_n P(AB_n) \\ &= \sum_n P(A|B_n)P(B_n). \end{aligned} \tag{2}$$

Then, we can write the expanded version of Bayes theorem

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_n P(A|B_n)P(B_n)}. \tag{3}$$

In Bayesian inference, equation (3) becomes

$$P(H_k|D) = \frac{P(D|H_k)P(H_k)}{\sum_n P(D|H_n)P(H_n)}, \tag{4}$$

which is the inferential form of Bayes' theorem, where H_k is the k-th hypothesis and D is the set of data. For a parametric set of hypotheses, this becomes

$$p(H(\boldsymbol{\theta})|D) = \frac{p(D|\boldsymbol{\theta})p[H(\boldsymbol{\theta})]}{\int_\Theta p(D|\boldsymbol{\theta'})p[H(\boldsymbol{\theta'})]d\boldsymbol{\theta'}} \tag{5}$$

These forms are definitely different from (3), because hypotheses and data are not quite the same thing as events in a probability space and because the prior distribution $P(H_n)$ is an unknown and must be guessed, possibly in an informed way. In other words, the quantities on the right are not clearly defined. While we can define a probability model for our data (the likelihood), the prior distribution is a matter of choice (more or less subjective), and this has an impact on the evidence as well. Can we justify the inferential statement in a somewhat rigorous way?

Consider a set of hypotheses $\{H(\boldsymbol{\theta})\}$, then we can introduce the pdf $p[H(\boldsymbol{\theta})]$ as a function of the parameters $\boldsymbol{\theta}$ that label elementary events in the probability space $\Omega_{\boldsymbol{\theta}}$. Similarly, $p(D) = p(\{x_1, x_2, \ldots, x_N\})$ where the x's are the individual data and the pdf's are those of the joint events $\{x_1, x_2, \ldots, x_N\}$ in the probability space $\Omega_{\boldsymbol{\theta}}$. Next, we define the product probability space $\Omega_H \times \Omega_{\boldsymbol{\theta}}$, which is the space of all pairs of events $(H(\boldsymbol{\theta}), \{x_1, x_2, \ldots, x_N\})$ with a proper assignment of a joint probability density $p[H(\boldsymbol{\theta}); \{x_1, x_2, \ldots, x_N\}]$. Apparently, this solves the problem of the poorly defined probability space, but we still have to consider the subjective quality of the prior distribution.

2 Cox's argument

It is clear that for Bayes' theorem to be logically sound and applicable, we need to justify the choice of the prior distribution. The straightforward answer is that this cannot be done in a compelling way, given that we can only estimate or guess the prior, but we can make all of this *logically plausible*. That this is actually so, was demonstrated in 1946 by Cox [1], and in this section I summarize his results (see also [2] for recent references on Cox's results).

The basic idea is that any approximate reasoning that assigns a probability to a logical statement (represented by a bold letter such as $\mathbf{a}$) should comply with the laws of logic, such as those listed below:

1. $\tilde{\tilde{\mathbf{a}}} = \mathbf{a}$
2. $\mathbf{a} \wedge \mathbf{b} = \mathbf{b} \wedge \mathbf{a} \quad \Rightarrow \quad \mathbf{ab} = \mathbf{ba}$
3. $\mathbf{a} \vee \mathbf{b} = \mathbf{b} \vee \mathbf{a}$
4. $\mathbf{a} \wedge \mathbf{a} = \mathbf{a} \quad \Rightarrow \quad \mathbf{aa} = \mathbf{a}$
5. $\mathbf{a} \vee \mathbf{a} = \mathbf{a}$
6. $\mathbf{a}(\mathbf{bc}) = (\mathbf{ab})\mathbf{c}$
7. $\mathbf{a} \vee (\mathbf{b} \vee \mathbf{c}) = (\mathbf{a} \vee \mathbf{b}) \vee \mathbf{c} = \mathbf{a} \vee \mathbf{b} \vee \mathbf{c}$
8. $\widetilde{\mathbf{ab}} = \tilde{\mathbf{a}} \vee \tilde{\mathbf{b}}$
9. $\widetilde{\mathbf{a} \vee \mathbf{b}} = \tilde{\mathbf{a}}\tilde{\mathbf{b}}$
10. $\mathbf{a}(\mathbf{a} \vee \mathbf{b}) = \mathbf{a}$
11. $\mathbf{a} \vee (\mathbf{ab}) = \mathbf{a}$

It is important to observe that not all these rules are independent. For instance, using rules 8 and 1, we find rule 9

$$\widetilde{\tilde{\mathbf{a}}\tilde{\mathbf{b}}} = \mathbf{a} \vee \mathbf{b} \quad \Rightarrow \quad \widetilde{\mathbf{a} \vee \mathbf{b}} = \tilde{\mathbf{a}}\tilde{\mathbf{b}} \tag{6}$$

Now let $f(\mathbf{a}|\mathbf{b})$ be a measure of the plausibility of $\mathbf{a}$ given $\mathbf{b}$, and consider the plausibility of the combined statement $f(\mathbf{cb}|\mathbf{a})$ which we expect to be a function of two variables

$$f(\mathbf{cb}|\mathbf{a}) = F[f(\mathbf{c}|\mathbf{ba}), f(\mathbf{b}|\mathbf{a})] \tag{7}$$

in analogy to the case of conditional probabilities $P(AB|C) = P(A|B,C)P(B|C)$.

Now, F is at least partly defined by its compliance with the logic of propositions described above. Consider the following

$$f(\mathbf{dcb}|\mathbf{a}) = f[\mathbf{d}(\mathbf{cb})|\mathbf{a}] = F[f(\mathbf{d}|\mathbf{cba}), f(\mathbf{cb}|\mathbf{a})] = F\{f(\mathbf{d}|\mathbf{cba}), F[f(\mathbf{c}|\mathbf{ba}), f(\mathbf{b}|\mathbf{a})]\} \tag{8}$$

$$= f[(\mathbf{dc})\mathbf{b}|\mathbf{a}] = F[f(\mathbf{dc}|\mathbf{ba}), f(\mathbf{b}|\mathbf{a})] = F\{F[f(\mathbf{d}|\mathbf{cba}), f(\mathbf{c}|\mathbf{ba})], f(\mathbf{b}|\mathbf{a})\} \tag{9}$$

i.e.,

$$F\{f(\mathbf{d}|\mathbf{cba}), F[f(\mathbf{c}|\mathbf{ba}), f(\mathbf{b}|\mathbf{a})]\} = F\{F[f(\mathbf{d}|\mathbf{cba}), f(\mathbf{c}|\mathbf{ba})], f(\mathbf{b}|\mathbf{a})\}. \tag{10}$$

We can simplify the notation by letting $x = f(\mathbf{d}|\mathbf{cba})$, $y = f(\mathbf{c}|\mathbf{ba})$, $z = f(\mathbf{b}|\mathbf{a})$ so that we finally obtain the functional equation

$$F[x, F(y,z)] = F[F(x,y), z]. \tag{11}$$

It can be shown (see the Appendix) that the general solution of equation (11) is

$$Cg[F(p,q)] = g(p)g(q) \tag{12}$$

where g is an arbitrary function and C is an arbitrary constant, therefore combining this result with equation (7)

$$Cg[f(\mathbf{cb}|\mathbf{a})] = Cg[F[f(\mathbf{c}|\mathbf{ba}), f(\mathbf{b}|\mathbf{a})]] = g(f(\mathbf{c}|\mathbf{ba}))g(f(\mathbf{b}|\mathbf{a})) \tag{13}$$

and simplifying the notation $g \circ f \to f$

$$Cf(\mathbf{cb}|\mathbf{a}) = f(\mathbf{c}|\mathbf{ba})f(\mathbf{b}|\mathbf{a}). \tag{14}$$

We can use the arbitrariness in the determination of C to make the solution look just as in the case of ordinary probabilities and therefore we set $C = 1$, i.e.,

$$f(\mathbf{cb}|\mathbf{a}) = f(\mathbf{c}|\mathbf{ba})f(\mathbf{b}|\mathbf{a}). \tag{15}$$

There is still too much freedom in expression (15), for instance

$$[f(\mathbf{cb}|\mathbf{a})]^m = [f(\mathbf{c}|\mathbf{ba})]^m [f(\mathbf{b}|\mathbf{a})]^m. \tag{16}$$

works just as well for any power m. However, we can use additional constraints from our set of rules, in particular we note that $\tilde{\mathbf{a}}$ is completely determined by its complement, therefore

$$f(\tilde{\mathbf{b}}|\mathbf{a}) = S[f(\mathbf{b}|\mathbf{a})]. \tag{17}$$

Obviously, this implies

$$f(\mathbf{b}|\mathbf{a}) = f\left(\tilde{\tilde{\mathbf{b}}}|\mathbf{a}\right) = S\left\{S[f(\mathbf{b}|\mathbf{a})]\right\}, \tag{18}$$

or

$$x = S\left[S(x)\right], \tag{19}$$

after setting $x = f(\mathbf{b}|\mathbf{a})$. Equation (19) is insufficient to determine S, but we have more logical rules that we can use, in particular, we note that

$$S[f(\mathbf{c} \vee \mathbf{b}|\mathbf{a})] = f\left(\widetilde{\mathbf{c} \vee \mathbf{b}}|\mathbf{a}\right) = f\left(\tilde{\mathbf{c}}\tilde{\mathbf{b}}|\mathbf{a}\right) = f(\tilde{\mathbf{c}}|\tilde{\mathbf{b}}\mathbf{a})f(\tilde{\mathbf{b}}|\mathbf{a}) = S[f(\mathbf{c}|\tilde{\mathbf{b}}\mathbf{a})]S[f(\mathbf{b}|\mathbf{a})]. \tag{20}$$

The resulting expression can be further streamlined after we remark that

$$f(\mathbf{c}|\tilde{\mathbf{b}}\mathbf{a}) = \frac{f(\mathbf{c}\tilde{\mathbf{b}}|\mathbf{a})}{f(\tilde{\mathbf{b}}|\mathbf{a})} = \frac{f(\tilde{\mathbf{b}}\mathbf{c}|\mathbf{a})}{f(\tilde{\mathbf{b}}|\mathbf{a})} = \frac{f(\tilde{\mathbf{b}}|\mathbf{ca})f(\mathbf{c}|\mathbf{a})}{f(\tilde{\mathbf{b}}|\mathbf{a})} = \frac{S[f(\mathbf{b}|\mathbf{ca})]f(\mathbf{c}|\mathbf{a})}{S[f(\mathbf{b}|\mathbf{a})]}, \tag{21}$$

i.e.,

$$\frac{S[f(\mathbf{c} \vee \mathbf{b}|\mathbf{a})]}{S[f(\mathbf{b}|\mathbf{a})]} = S\left\{\frac{S[f(\mathbf{b}|\mathbf{ca})]f(\mathbf{c}|\mathbf{a})}{S[f(\mathbf{b}|\mathbf{a})]}\right\} \tag{22}$$

or also, recalling that $f(\mathbf{bc}|\mathbf{a}) = f(\mathbf{b}|\mathbf{ca})f(\mathbf{c}|\mathbf{a})$,

$$S\left[\frac{S[f(\mathbf{c} \vee \mathbf{b}|\mathbf{a})]}{S[f(\mathbf{b}|\mathbf{a})]}\right] = \frac{S[f(\mathbf{b}|\mathbf{ca})]f(\mathbf{c}|\mathbf{a})}{S[f(\mathbf{b}|\mathbf{a})]} = S\left[\frac{f(\mathbf{bc}|\mathbf{a})}{f(\mathbf{c}|\mathbf{a})}\right]\frac{f(\mathbf{c}|\mathbf{a})}{S[f(\mathbf{b}|\mathbf{a})]}. \tag{23}$$

Finally, we let $\mathbf{b} = \mathbf{cd}$ and the previous equation becomes

$$S\left[\frac{S[f(\mathbf{c}|\mathbf{a})]}{S[f(\mathbf{cd}|\mathbf{a})]}\right] = S\left[\frac{f(\mathbf{cd}|\mathbf{a})}{f(\mathbf{c}|\mathbf{a})}\right]\frac{f(\mathbf{c}|\mathbf{a})}{S[f(\mathbf{cd}|\mathbf{a})]}. \tag{24}$$

Just as in the previous discussion, we introduce auxiliary variables $x = f(\mathbf{c}|\mathbf{a})$, $y = S[f(\mathbf{cd}|\mathbf{a})]$ and the previous equation becomes

$$yS\left[\frac{S(x)}{y}\right] = xS\left[\frac{S(y)}{x}\right] \tag{25}$$

It can be shown (see the Appendix) that the general solution of equation (25) is

$$S(t) = (1 - t^m)^{1/m} \tag{26}$$

where m is an arbitrary constant. Then, *using the previous results and we require that the f's are normalized*, i.e.,

$$f(\mathbf{b}|\mathbf{a})^m + f(\tilde{\mathbf{b}}|\mathbf{a})^m = f(\mathbf{b}|\mathbf{a})^m + S[f(\tilde{\mathbf{b}}|\mathbf{a})]^m = 1 \tag{27}$$

where m is again an arbitrary parameter. **Constraining m to be the same in equations** (26) **and** (27), we find that

$$f(\mathbf{b}|\mathbf{a})^m + \left[(1 - f(\mathbf{b}|\mathbf{a})^m)^{1/m}\right]^m = f(\mathbf{b}|\mathbf{a})^m + [1 - f(\mathbf{b}|\mathbf{a})^m] = 1 \tag{28}$$

as required. This means that we can redefine f absorbing the m parameter and let $f \to p$. Then, the previous result writes

$$p(\mathbf{b}|\mathbf{a}) + p(\tilde{\mathbf{b}}|\mathbf{a}) = 1 \tag{29}$$

implying

$$p(\mathbf{a}|\mathbf{a}) + p(\tilde{\mathbf{a}}|\mathbf{a}) = 1, \tag{30}$$

and therefore, assuming $p(x) \geq 0 \quad \forall x$, this means that $0 \leq p(\mathbf{b}|\mathbf{a}) \leq 1$, $p(\mathbf{a}|\mathbf{a}) = 1$, and $p(\tilde{\mathbf{a}}|\mathbf{a}) = 0$.

Overall, we have the following collection of assumptions and rules

- $p(\text{certainty}) = 1$
- $p(\text{impossibility}) = 0$
- $p(\mathbf{b}|\mathbf{a}) + p(\tilde{\mathbf{b}}|\mathbf{a}) = 1$
- $p(\mathbf{cb}|\mathbf{a}) = p(\mathbf{c}|\mathbf{ba})p(\mathbf{b}|\mathbf{a})$

from which all the other usual rules of probability follow.

Conclusion: "credibilities" behave just like probabilities and we can use Bayes' theorem without violating the rules of logic.

A Solution of the Cox's functional equations

The functional equations (11) and ... are solved with the assumption that F has continuous second derivatives.

A.1 Solution of $F[x, F(y,z)] = F[F(x,y), z]$

Throughout this proof, I use the notation F_i to denote the derivative with respect to the i-th argument to help solve equation (11)

$$F[x, F(y,z)] = F[F(x,y), z]. \tag{31}$$

As a first step, we define the auxiliary variables $u = F(y,z)$ and $v = F(x,y)$. Then, equation (31) becomes

$$F[x,u] = F[v,z]. \tag{32}$$

Taking the first derivative of both members with respect to x, y, and z, yields

$$F_1[x,u] = F_1[v,z]\frac{\partial v}{\partial x} = F_1[v,z]F_1[x,y] \tag{33}$$

$$F_2[x,u]\frac{\partial u}{\partial y} = F_2[x,u]F_1[y,z] = F_1[v,z]\frac{\partial v}{\partial y} = F_1[v,z]F_2[x,y] \tag{34}$$

$$F_2[x,u]\frac{\partial u}{\partial z} = F_2[x,u]F_2[y,z] = F_2[v,z] \tag{35}$$

Next, moving to the 9 second derivatives, we find

$$(xx) \quad F_{11}[x,u] = F_{11}[v,z](F_1[x,y])^2 + F_1[v,z]F_{11}[x,y] \tag{36}$$

$$(xy) \quad F_{12}[x,u]\frac{\partial u}{\partial y} = F_{12}[x,u]F_1[y,z] = F_{11}[v,z]F_1[x,y]F_2[x,y] + F_1[v,z]F_{12}[x,y] \tag{37}$$

$$(xz) \quad F_{12}[x,u]\frac{\partial u}{\partial z} = F_{12}[x,u]F_2[y,z] = F_{12}[v,z]F_1[x,y] \tag{38}$$

$$(yx) \quad F_{12}[x,u]F_1[y,z] = F_{11}[v,z]F_1[x,y]F_2[x,y] + F_1[v,z]F_{12}[x,y] \tag{39}$$

$$(yy) \quad F_{22}[x,u](F_1[y,z])^2 + F_2[x,u]F_{11}[y,z] = F_{11}[v,z](F_2[x,y])^2 + F_1[v,z]F_{22}[x,y] \tag{40}$$

$$(yz) \quad F_{22}[x,u]F_1[y,z]F_2[y,z] + F_2[x,u]F_{12}[y,z] = F_{12}[v,z]F_2[x,y] \tag{41}$$

$$(zx) \quad F_{12}[x,u]F_2[y,z] = F_{12}[v,z]F_1[x,y] \tag{42}$$

$$(zy) \quad F_{22}[x,u]F_1[y,z]F_2[y,z] + F_2[x,u]F_{11}[y,z] = F_{12}[v,z]F_2[x,y] \tag{43}$$

$$(zz) \quad F_{22}[x,u](F_2[y,z])^2 + F_2[x,u]F_{22}[y,z] = F_{22}[v,z] \tag{44}$$

Overall, the set of equation is

$$
\begin{aligned}
&(x) \quad F_1[x,u] = F_1[v,z]F_1[x,y] && (45)\\
&(y) \quad F_2[x,u]F_1[y,z] = F_1[v,z]F_2[x,y] && (46)\\
&(z) \quad F_2[x,u]F_2[y,z] = F_2[v,z] && (47)\\
&(xx) \quad F_{11}[x,u] = F_{11}[v,z](F_1[x,y])^2 + F_1[v,z]F_{11}[x,y] && (48)\\
&(xy) \quad F_{12}[x,u]F_1[y,z] = F_{11}[v,z]F_1[x,y]F_2[x,y] + F_1[v,z]F_{12}[x,y] && (49)\\
&(xz) \quad F_{12}[x,u]F_2[y,z] = F_{12}[v,z]F_1[x,y] && (50)\\
&(yx) \quad F_{12}[x,u]F_1[y,z] = F_{11}[v,z]F_1[x,y]F_2[x,y] + F_1[v,z]F_{12}[x,y] && (51)\\
&(yy) \quad F_{22}[x,u](F_1[y,z])^2 + F_2[x,u]F_{11}[y,z] = F_{11}[v,z](F_2[x,y])^2 + F_1[v,z]F_{22}[x,y] && (52)\\
&(yz) \quad F_{22}[x,u]F_1[y,z]F_2[y,z] + F_2[x,u]F_{12}[y,z] = F_{12}[v,z]F_2[x,y] && (53)\\
&(zx) \quad F_{12}[x,u]F_2[y,z] = F_{12}[v,z]F_1[x,y] && (54)\\
&(zy) \quad F_{22}[x,u]F_1[y,z]F_2[y,z] + F_2[x,u]F_{11}[y,z] = F_{12}[v,z]F_2[x,y] && (55)\\
&(zz) \quad F_{22}[x,u](F_2[y,z])^2 + F_2[x,u]F_{22}[y,z] = F_{22}[v,z] && (56)
\end{aligned}
$$

where the derivatives of u and v are highlighted in red. Because of the exchange symmetry in mixed derivatives, there are three expressions that can be eliminated, and the new system of independent equations is

$$
\begin{aligned}
&(x) \quad F_1[x,u] = F_1[v,z]F_1[x,y] && (57)\\
&(y) \quad F_2[x,u]F_1[y,z] = F_1[v,z]F_2[x,y] && (58)\\
&(z) \quad F_2[x,u]F_2[y,z] = F_2[v,z] && (59)\\
&(xx) \quad F_{11}[x,u] = F_{11}[v,z](F_1[x,y])^2 + F_1[v,z]F_{11}[x,y] && (60)\\
&(xy) \quad F_{12}[x,u]F_1[y,z] = F_{11}[v,z]F_1[x,y]F_2[x,y] + F_1[v,z]F_{12}[x,y] && (61)\\
&(xz) \quad F_{12}[x,u]F_2[y,z] = F_{12}[v,z]F_1[x,y] && (62)\\
&(yy) \quad F_{22}[x,u](F_1[y,z])^2 + F_2[x,u]F_{11}[y,z] = F_{11}[v,z](F_2[x,y])^2 + F_1[v,z]F_{22}[x,y] && (63)\\
&(yz) \quad F_{22}[x,u]F_1[y,z]F_2[y,z] + F_2[x,u]F_{12}[y,z] = F_{12}[v,z]F_2[x,y] && (64)\\
&(zz) \quad F_{22}[x,u](F_2[y,z])^2 + F_2[x,u]F_{22}[y,z] = F_{22}[v,z] && (65)
\end{aligned}
$$

Next, we go through the following steps:

1. **Eliminate** $F_{12}[x,u]$ **using equations** (xy) **and** (xz): here we multiply equation (xy) times $F_2[y,z]$ and equation (xz) times $F_1[y,z]$ and subtract the second from the first

$$F_{11}[v,z]F_1[x,y]F_2[x,y]F_2[y,z] + F_1[v,z]F_{12}[x,y]F_2[y,z] = F_{12}[v,z]F_1[x,y]F_1[y,z] \tag{66}$$

2. **Eliminate $F_{12}[v,z]$ from (66) using equation (yz):**

$$F_{11}[v,z]F_1[x,y]F_2[x,y]F_2[y,z] + F_1[v,z]F_{12}[x,y]F_2[y,z] = \\ = F_1[x,y]F_1[y,z]\frac{F_{22}[x,u]F_1[y,z]F_2[y,z] + F_2[x,u]F_{12}[y,z]}{F_2[x,y]} \tag{67}$$

3. **Rearrange equation (67):**

$$F_{11}[v,z]F_1[x,y](F_2[x,y])^2F_2[y,z] + F_1[v,z]F_{12}[x,y]F_2[x,y]F_2[y,z] = \\ = F_{22}[x,u]F_1[x,y](F_1[y,z])^2F_2[y,z] + F_2[x,u]F_{12}[y,z]F_1[x,y]F_1[y,z] \tag{68}$$

4. **Rearrange equation (68):**

$$F_1[x,y]F_2[y,z]\left(F_{11}[v,z](F_2[x,y])^2 - F_{22}[x,u](F_1[y,z])^2\right) = \\ = F_2[x,u]F_{12}[y,z]F_1[x,y]F_1[y,z] - F_1[v,z]F_{12}[x,y]F_2[x,y]F_2[y,z] \tag{69}$$

5. **Rearrange equation (yy):**

$$F_{22}[x,u](F_1[y,z])^2 - F_{11}[v,z](F_2[x,y])^2 = F_1[v,z]F_{22}[x,y] - F_2[x,u]F_{11}[y,z] \tag{70}$$

6. **Combine equations (69) and (70):**

$$F_2[x,u]F_{11}[y,z]F_1[x,y]F_2[y,z] - F_1[v,z]F_{22}[x,y]F_1[x,y]F_2[y,z] = \\ = F_2[x,u]F_{12}[y,z]F_1[x,y]F_1[y,z] - F_1[v,z]F_{12}[x,y]F_2[x,y]F_2[y,z] \tag{71}$$

7. **Rearrange equation (71):**

$$\frac{F_2[x,u]}{F_1[v,z]} = \frac{F_{22}[x,y]F_1[x,y]F_2[y,z] - F_{12}[x,y]F_2[x,y]F_2[y,z]}{F_{11}[y,z]F_1[x,y]F_2[y,z] - F_{12}[y,z]F_1[x,y]F_1[y,z]} \tag{72}$$

8. **Use equation (y) to get rid of the $F_2[x,u]/F_1[v,z]$ ratio:**

$$\frac{F_2[x,y]}{F_1[y,z]} = \frac{F_{22}[x,y]F_1[x,y]F_2[y,z] - F_{12}[x,y]F_2[x,y]F_2[y,z]}{F_{11}[y,z]F_1[x,y]F_2[y,z] - F_{12}[y,z]F_1[x,y]F_1[y,z]} \tag{73}$$

9. **Remove the color (no longer useful) and replace the compact notation with the explicit derivatives:**

$$\frac{\frac{\partial v}{\partial y}}{\frac{\partial u}{\partial y}} = \frac{\frac{\partial^2 v}{\partial y^2}\frac{\partial v}{\partial x}\frac{\partial u}{\partial z} - \frac{\partial^2 v}{\partial x \partial y}\frac{\partial v}{\partial y}\frac{\partial u}{\partial z}}{\frac{\partial^2 u}{\partial y^2}\frac{\partial v}{\partial x}\frac{\partial u}{\partial z} - \frac{\partial^2 u}{\partial y \partial z}\frac{\partial v}{\partial x}\frac{\partial u}{\partial y}} = \frac{\frac{\partial u}{\partial z}\left(\frac{\partial^2 v}{\partial y^2}\frac{\partial v}{\partial x} - \frac{\partial^2 v}{\partial x \partial y}\frac{\partial v}{\partial y}\right)}{\frac{\partial v}{\partial x}\left(\frac{\partial^2 u}{\partial y^2}\frac{\partial u}{\partial z} - \frac{\partial^2 u}{\partial y \partial z}\frac{\partial u}{\partial y}\right)} \tag{74}$$

Equation (74) can be rearranged to obtain

$$\left(\frac{\partial u}{\partial y}\right)^{-1}\frac{\partial^2 u}{\partial y^2}-\left(\frac{\partial u}{\partial z}\right)^{-1}\frac{\partial^2 u}{\partial y\partial z} = \frac{\partial}{\partial y}\ln\frac{\partial u}{\partial y}-\frac{\partial}{\partial y}\ln\frac{\partial u}{\partial z} = \frac{\partial}{\partial y}\ln\frac{\frac{\partial u}{\partial y}}{\frac{\partial u}{\partial z}} =$$
$$= \left(\frac{\partial v}{\partial y}\right)^{-1}\frac{\partial^2 v}{\partial y^2}-\left(\frac{\partial v}{\partial x}\right)^{-1}\frac{\partial^2 v}{\partial x\partial y} = \frac{\partial}{\partial y}\ln\frac{\partial v}{\partial y}-\frac{\partial}{\partial y}\ln\frac{\partial v}{\partial x} = \frac{\partial}{\partial y}\ln\frac{\frac{\partial v}{\partial y}}{\frac{\partial v}{\partial x}} \tag{75}$$

where — going back to the F notation — the equality between the two blue terms can be rewritten as

$$\frac{\partial}{\partial y}\ln\frac{F_1(y,z)}{F_2(y,z)} = -\frac{\partial}{\partial y}\ln\frac{F_1(x,y)}{F_2(x,y)} \tag{76}$$

10. **Integrate equation** (76): Equality (76) means that both sides can only depend on y, therefore we define the ratio $\Phi(y) = F_1(x,y)/F_2(x,y)$ and we obtain

$$\frac{\partial}{\partial y}\ln\frac{F_1(x,y)}{F_2(x,y)} = \frac{d}{dy}\ln\Phi(y) \tag{77}$$
$$-\frac{\partial}{\partial y}\ln\frac{F_1(y,z)}{F_2(y,z)} = \frac{d}{dy}\ln\Phi(y) \tag{78}$$

Then, with a circular permutation of x, y, and z in equation (78) we obtain also

$$\frac{\partial}{\partial x}\ln\frac{F_1(x,y)}{F_2(x,y)} = -\frac{d}{dx}\ln\Phi(x) \tag{79}$$

Therefore, from (77) and (79), we obtain

$$d\left[\ln\frac{F_1(x,y)}{F_2(x,y)}\right] = \frac{\partial}{\partial x}\ln\frac{F_1(x,y)}{F_2(x,y)}dx+\frac{\partial}{\partial y}\ln\frac{F_1(x,y)}{F_2(x,y)}dy =$$
$$= -\frac{d}{dx}\ln\Phi(x)dx+\frac{d}{dy}\ln\Phi(y)dy = d\ln\frac{\Phi(y)}{\Phi(x)} \tag{80}$$

This equation can be integrated to give

$$\frac{F_1(x,y)}{F_2(x,y)} = C_1\frac{\Phi(y)}{\Phi(x)} \tag{81}$$

where C_1 is an integration constant.

11. **Proceed with the final manipulations**: Dividing equation (34) by equation (35), we find

$$\frac{F_1[y,z]}{F_2[y,z]} = \frac{F_1[v,z]}{F_2[v,z]}\frac{\partial v}{\partial y}, \tag{82}$$

therefore, using equation (81), we obtain

$$\frac{\Phi(z)}{\Phi(y)} = \frac{\partial v}{\partial y}\frac{\Phi(z)}{\Phi(v)}, \tag{83}$$

i.e.,

$$\frac{\partial}{\partial y}F(x,y) = \frac{\Phi\left[F(x,y)\right]}{\Phi(y)} \tag{84}$$

Similarly, from equations (33) and (34) we obtain

$$\frac{F_1[x,u]}{F_2[x,u]} = \frac{F_1[x,y]}{F_2[x,y]}\frac{\partial u}{\partial y}, \tag{85}$$

and therefore

$$\frac{\partial}{\partial y}F(y,z) = \frac{\Phi\left[F(y,z)\right]}{\Phi(y)} \tag{86}$$

which becomes

$$\frac{\partial}{\partial x}F(x,y) = \frac{\Phi\left[F(x,y)\right]}{\Phi(x)} \tag{87}$$

after a circular permutation.

We combine equations (84) and (87) to obtain

$$dF = \frac{\partial}{\partial x}F(x,y)dx + \frac{\partial}{\partial y}F(x,y)dy = \frac{\Phi\left[F(x,y)\right]}{\Phi(x)}dx + \frac{\Phi\left[F(x,y)\right]}{\Phi(y)}dy, \tag{88}$$

i.e.,

$$\frac{dF}{\Phi\left[F(x,y)\right]} = \frac{dx}{\Phi(x)} + \frac{dy}{\Phi(y)}. \tag{89}$$

If we let

$$\ln f(t) = \int \frac{dt}{\Phi(t)} \quad \Rightarrow \quad \frac{df}{f} = \frac{dt}{\Phi(t)} \tag{90}$$

then

$$\ln f[F(x,y)] = C_2 + \ln f(x) + \ln f(y) \quad \Rightarrow \quad f[F(x,y)] = Cf(x)f(y) \tag{91}$$

A.2 Solution of $yS[S(x)/y] = xS[S(y)/x]$

Just as above, we define the following auxiliary variables $u = S(x)/y$ and $v = S(y)/x$, so that equation (25) becomes

$$yS[u] = xS[v] \tag{92}$$

and we go through the following steps:

1. **Take derivatives with respect to x, y, and both x and y, and simplify**:

$$S'[x]S'[u] = S[v] - vS'[v] \tag{93}$$
$$S[u] - uS'[u] = S'[y]S'[v] \tag{94}$$
$$\frac{u}{y}S'[x]S''[u] = \frac{v}{x}S'[y]S''[v] \tag{95}$$

2. **Combine equations (92) and (95)**: By multiplying these equations we get rid of the denominators and obtain

$$uS[u]S'[x]S''[u] = vS[v]S'[y]S''[v] \tag{96}$$

3. **Eliminate the middle factors $S'[x]$ and $S'[y]$ using (93) and (94)**:

$$S'[x] = \frac{S[v] - vS'[v]}{S'[u]} \tag{97}$$
$$S'[y] = \frac{S[u] - uS'[u]}{S'[v]} \tag{98}$$

Therefore,

$$uS[u]\left(\frac{S[v] - vS'[v]}{S'[u]}\right)S''[u] = vS[v]\left(\frac{S[u] - uS'[u]}{S'[v]}\right)S''[v] \tag{99}$$

from which we obtain

$$\frac{uS[u]S''[u]}{(S[u] - uS'[u])S'[u]} = \frac{vS[v]S''[v]}{(S[v] - vS'[v])S'[v]} \tag{100}$$

where the variables have been separated. The equality means that each term is equal to the same constant (k).

4. **Solve the final differential equation**:

$$uS[u]S''[u] = k(S[u] - uS'[u])S'[u]. \tag{101}$$

Dividing each term by $uS[u]S'[u]$, we find

$$\frac{S''[u]}{S'[u]} = \frac{k}{u} - k\frac{S'[u]}{S[u]}, \tag{102}$$

i.e.,

$$\frac{dS'}{S'} = k\frac{du}{u} - k\frac{dS}{S}, \tag{103}$$

and therefore we can integrate at once to find

$$\ln S'[u] = \ln u^k - \ln S[u]^k + \text{const} \tag{104}$$

and the new differential equation

$$S'[u] = A\left(\frac{u}{S[u]}\right)^k, \tag{105}$$

or, equivalently

$$S'[u](S[u])^k = Au^k \tag{106}$$

from which we finally obtain

$$(S[u])^m = Au^m + B \tag{107}$$

where $m = k + 1$ and A, B are integration constants.

5. **Determine the integration constants**: Substituting the result (107) in equation (92), we find

$$\begin{aligned} yS[u] = y(Au^m + B)^{1/m} = y\left[A\left(\frac{S[x]}{y}\right)^m + B\right]^{1/m} &= [A(Ax^m + B) + By^m]^{1/m} \\ = xS[v] = x(Av^m + B)^{1/m} &= [A(Ay^m + B) + Bx^m]^{1/m} \end{aligned} \tag{108}$$

Comparing terms, we see that $A^2 = B$. Moreover, the equivalence $x = S[S[x]]$ means that

$$x = (AS[x]^m + B)^{1/m} = [A(Ax^m + B) + B]^{1/m} \quad \Rightarrow \quad x^m = A^2x^m + AB + B \tag{109}$$

and this implies $|A| = 1$ and $A = -1$.

To conclude, we find $S[x]^m = -x^m + 1$ and therefore $x^m + S[x]^m = 1$.

References

[1] Richard T Cox. Probability, frequency and reasonable expectation. *American journal of physics*, 14(1):1–13, 1946.

[2] Kevin S Van Horn. Constructing a logic of plausible inference: a guide to Cox's theorem. *International Journal of Approximate Reasoning*, 34(1):3–24, 2003.