cox-theorem-notes

October 6, 2022

Author: G. Jarrad

This document summarises some of the relations due to Cox's Theorem [1], particularly as they relate to the discussion by Van Horn [2]. Note that here we are aiming more for explicability than formal exactness.

1 The Calculus of Plausible Inference

1.1 Propositional calculus

We start with the notion of Boolean propositions, denoted by A, B, C, et cetera, that must take on values of either `True` or `False`. However, our knowledge about the truth values of these propositions is always contextual, and depends on some background knowledge, denoted by X. Hence, $A \mid X$ denotes the proposition that A is conditionally `True` given X. Conversely, $\neg A \mid X$ denotes the propostion that A is conditionally `False` given X. Similarly, $A \wedge B \mid X$ denotes *both* A and B being conditionally `True`, and $A \vee B \mid X$ denotes *either* A or B (or both) being conditionally `True`.

1.2 Credibility of propositions

In the case where we do not know with certainty that some proposition $A \mid X$ is definitely `True` or definitely `False`, then we instead have some degree of *belief* in its truth value. Formally, we encapsulate the degree of belief via some *credibility* (or plausibility) function $c(\cdot)$, such that $c(A \mid X)$ denotes our subjective measure of the cedibility of the proposition $A \mid X$.

In a unimodal calculus of credibility, we suppose that credibility is denoted by a single, real number, such that a higher number indicates more certainty that the proposition is `True`, and a lower number indicates more certainty that the proposition is `False`. However, we further require that the calculus of credibility remains consistent with the calculus of propositional logic. In particular, if we know that $A \mid X$ is `True`, then the assigned credibility must attain its maximimum value $T \doteq c(\texttt{True})$. Conversely, if we know that $A \mid X$ is `False`, then the assigned credibility must attain its minimimum value $F \doteq c(\texttt{False})$. Hence, in general, we have $F \leq c(A \mid X) \leq T$.

Note, however, that we have neither defined nor proscribed any given values of T and F. In particular, we have not ruled out $F = -\infty$ or $T = +\infty$. However, we can assume that $F < T$, otherwise we will always be stuck with a singular (and thus unexpressive) value of credibility for every proposition. Also note that we are assuming that credibility values are dense in the interval $[F, T]$. This is essentially the *universality* axiom.

1.3 Credibility of negation

Suppose that our background knowledge changes from X to X', such that our (conditional) degree of belief in some proposition A increases, i.e. $c(A \mid X') \geq c(A \mid X)$. It follows, for our unimodal calculus of credibility, that our degree of belief in the converse proposition $\neg A$ must decrease, i.e. $c(\neg A \mid X') \leq c(\neg A \mid X)$. In general, we suppose that our degrees of belief in the contrary propositions $A \mid X$ and $\neg A \mid X$ are related via some *complementation* function $s(\cdot)$, such that $c(\neg A \mid X) = s(c(A \mid X))$ and $c(A \mid X) = s(c(\neg A \mid X))$.

It immediately follows that $s(\cdot)$ is self-invertible in the sense that $s(s(x)) = x$ for any valid credibility $x = c(A \mid X)$. Clearly, by construction, $s(x)$ is a monitonically decreasing function of x. Furthermore, for consistency with the calculus of propositional logic, we observe that knowing that $A \mid X$ is `True` is equivalent to knowing that $\neg A \mid X$ is `False`, such that $s(T) = F$ and $s(F) = T$.

1.4 Credibility of conjunction

We now turn to the conjunctive proposition $A \wedge B \mid X$, and its credibility $c(A \wedge B \mid X)$. Observe that if $B \mid X =$ `True`, then it follows logically that $A \wedge B \mid X = A \mid X$. Conversely, if $B \mid X =$ `False` then $A \wedge B \mid X =$ `False`. Hence, we might suppose that $c(A \wedge B \mid X)$ depends upon $c(A \mid X)$ and $c(B \mid X)$.

However, if $B \mid X =$ `True`, then B is consistent with, and adds nothing to, the background knowledge X. In other words, the new knowledge $X' \doteq B \wedge X$ is still just X. Hence, if $B \mid X =$ `True` then it follows that $A \mid X = A \mid B \wedge X$. Similarly, if $B \mid X =$ `True`, then it follows that $B \mid A \wedge X =$ `True` for any proposition A that is consistent with X. Consequently, we might suppose that $c(A \wedge B \mid X)$ also depends upon both $c(A \mid B \wedge X)$ and $c(B \mid A \wedge X)$.

In general, we therefore suppose that the credibility of the proposition $A \wedge B \mid X$ depends on the atomic credibilities of $x = c(A \mid X)$, $y = c(B \mid X)$ and $w = c(A \mid B \wedge X)$, $z = c(B \mid A \wedge X)$. We thus posit a *conjunctive* function $f(\cdot)$ such that $c(A \wedge B \mid X) = f(x, y, w, z)$. Since each atomic credibility, taken in turn, might or might not influence $c(A \wedge B \mid X)$, we see that there are $2^4 = 16$ distinct functions, each with a distinct argument signature. These arbitrary functions, enumerated as f_i for $i = 1, 2 \ldots, 16$, are listed in the first column of the table below.

Note that if $c(A \wedge B \mid X)$ is defined by any particular function f_i, then this relation must hold for all propositions A and B, and any (consistent) background knowledge X. In particular, if we suppose that $B \mid X =$ `True`, then (from above) we have $y = c(B \mid X) = T$ and $z = c(B \mid A \wedge X) = T$, and also $w = c(A \mid B \wedge X) = c(A \mid X) = x$, and finally $c(A \wedge B \mid X) = c(A \mid X) = x$.

These values give us the constraints listed in the second column of the table. We may therefore immediately rule out functions f_1, f_3, f_5 and f_{10} as being inconsistent, since their respective constraints imply that a constant left-hand side must equal a variable right-hand side.

By a similar argument, we deduce that if, alternatively, we suppose that $A \mid X =$ `True`, then we obtain $x = c(A \mid X) = T$, $w = c(A \mid B \wedge X) = T$, $z = c(B \mid A \wedge X) = c(B \mid X) = y$ and $c(A \wedge B \mid X) = c(B \mid X) = y$. These values give us the constraints listed in the third column of the table. Consequently, we may further rule out functions f_2, f_4 and f_7 as being inconsistent.

Continuing these arguments, suppose now that $A \mid X = \neg B \mid X$. Then we obtain that $c(A \wedge B \mid X) = F$, along with $x = c(A \mid X) = c(\neg B \mid X) = s(c(B \mid X)) = s(y)$, and $w = c(A \mid B \wedge X) = F$ and $z = c(B \mid A \wedge X) = F$. These values form the constraints listed in the fourth column of the table. We now rule out function f_6 on the basis that the constraint implies that both $f_6(x, y) = F$

and $y = s(x)$; informally, the intersection of these two curves cannot hold for the dense subset $x \in [F, T]$ (as is required by the universality axiom). Note that Van Horn [2] uses a more detailed, multi-faceted argument to formally make this conclusion.

Finally, suppose now that $A \mid X = B \mid X$. Consequently, we obtain $x = c(A \mid X) = c(B \mid X) = y$, along with $w = c(A \mid B \wedge X) = T$, $z = c(B \mid A \wedge X) = T$, and $c(A \wedge B \mid X) = c(A \mid X) = x$. These values form the constraints listed in the fifth column of the table. Hence, we now also eliminate function f_{11}.

$c(A \wedge B \mid X)$	$B \mid X = \texttt{True}$	$A \mid X = \texttt{True}$	$A \mid X = \neg B \mid X$	$A \mid X = B \mid X$
$f_1()$	$\cancel{f_1() = x}$	$\cancel{f_1() = y}$	$f_1() = F$	$\cancel{f_1() = x}$
$f_2(x)$	$f_2(x) = x$	$\cancel{f_2(T) = y}$	$\cancel{f_2(x) = F}$	$f_2(x) = x$
$f_3(y)$	$\cancel{f_3(T) = x}$	$f_3(y) = y$	$\cancel{f_3(y) = F}$	$f_3(y) = y$
$f_4(w)$	$f_4(w) = w$	$\cancel{f_4(T) = y}$	$f_4(F) = F$	$\cancel{f_4(T) = x}$
$f_5(z)$	$\cancel{f_5(T) = x}$	$f_5(z) = z$	$f_5(F) = F$	$\cancel{f_5(T) = x}$
$f_6(x, y)$	$f_6(x, T) = x$	$f_6(T, y) = y$	$\cancel{f_6(x, s(x)) = F}$	$f_6(x, x) = x$
$f_7(x, w)$	$f_7(x, x) = x$	$\cancel{f_7(T, T) = y}$	$f_7(x, F) = F$	$f_7(x, T) = x$
$f_8(x, z)$	$f_8(x, T) = x$	$f_8(T, z) = z$	$f_8(x, F) = F$	$f_8(x, T) = x$
$f_9(y, w)$	$f_9(T, w) = w$	$f_9(y, T) = y$	$f_9(y, F) = F$	$f_9(y, T) = y$
$f_{10}(y, z)$	$\cancel{f_{10}(T, T) = x}$	$f_{10}(y, y) = y$	$f_{10}(y, F) = F$	$f_{10}(y, T) = y$
$f_{11}(w, z)$	$f_{11}(w, T) = w$	$f_{11}(T, z) = z$	$f_{11}(F, F) = F$	$\cancel{f_{11}(T, T) = x}$
$f_{12}(x, y, w)$	$f_{12}(x, T, x) = x$	$f_{12}(T, y, T) = y$	$f_{12}(x, s(x), F) = F$	$f_{12}(x, x, T) = x$
$f_{13}(x, y, z)$	$f_{13}(x, T, T) = x$	$f_{13}(T, y, y) = y$	$f_{13}(x, s(x), F) = F$	$f_{13}(x, x, T) = x$
$f_{14}(x, w, z)$	$f_{14}(x, x, T) = x$	$f_{14}(T, T, z) = z$	$f_{14}(x, F, F) = F$	$f_{14}(x, T, T) = x$
$f_{15}(y, w, z)$	$f_{15}(T, w, T) = w$	$f_{15}(y, T, y) = y$	$f_{15}(y, F, F) = F$	$f_{15}(y, T, T) = y$
$f_{16}(x, y, w, z)$	$f_{16}(x, T, x, T) = x$	$f_{16}(T, y, T, y) = y$	$f_{16}(x, s(x), F, F) = F$	$f_{16}(x, x, T, T) = x$

So far, we have ignored the symmetry that $A \wedge B \mid X = B \wedge A \mid X$. This symmetry allows us to swap the labels A and B to obtain the same results, implying that we may exchange the pairs $x = c(A \mid X) \leftrightarrow y = c(B \mid X)$ and $w = c(A \mid B \wedge X) \leftrightarrow z = c(B \mid A \wedge X)$. This exchangability then induces some pairings between formulae:

$$\begin{aligned}
c(A \wedge B \mid X) &= f_8(x, z) = f_8(y, w) = f_9(y, w) = f_9(x, z)\,, \\
c(A \wedge B \mid X) &= f_{12}(x, y, w) = f_{12}(y, x, z) = f_{13}(x, y, z) = f_{13}(y, x, w)\,, \\
c(A \wedge B \mid X) &= f_{14}(x, w, z) = f_{14}(y, z, w) = f_{15}(y, w, z) = f_{15}(x, z, w)\,, \\
c(A \wedge B \mid X) &= f_{16}(x, y, w, z) = f_{16}(y, x, z, w)\,.
\end{aligned}$$

At this juncture, Van Horn [2] disputes the claims of Tribus [3] to have eliminated the final three models.

Van Horn then goes on to prove the standard result of Cox [1], namely that there exists a non-negative function $g(\cdot)$ with $g(F) = 0$, such that $g(f_8(x, z)) = g(x)g(z)$. Roughly speaking, we may therefore define the probability function $p(\cdot) \doteq g(c(\cdot))$, such that

$$p(A \wedge B \mid X) = p(A \mid X)\, p(B \mid A \wedge X) = p(B \mid X)\, p(A \mid B \wedge X)\,,$$

where we have chosen g to satisfy $g(T) = g(c(\texttt{True})) = p(\texttt{True}) = 1$. Clearly, we then have $p(\texttt{False}) = g(c(\texttt{False})) = g(F) = 0$.

At this point, however, we note that functions f_{12} to f_{16} cannot have this same factorisation. If they did, then we observe from the table above for $B \mid X = \texttt{True}$, for instance, that

$$g(x) \quad = \quad g(f_{16}(x,T,x,T)) \quad = \quad [g(T)\, g(x)]^2 \ ,$$

which cannot hold true for all $x \in [F,T]$ unless g is everywhere zero, unity or infinity.

In fact, a simple dimensional analysis (e.g. supposing that $c(A \mid X)$ has dimensional units $[A]$) shows that

$$\begin{aligned}
g(f_{12}(x,y,w)) &= g(y)\,\sqrt{g(x)\,g(w)}\,, \\
g(f_{13}(x,y,z)) &= g(x)\,\sqrt{g(y)\,g(z)}\,, \\
g(f_{14}(x,w,z)) &= g(z)\,\sqrt{g(x)\,g(w)}\,, \\
g(f_{15}(y,w,z)) &= g(w)\,\sqrt{g(y)\,g(z)}\,, \\
g(f_{16}(x,y,w,z)) &= \sqrt{g(x)\,g(w)}\,\sqrt{g(y)\,g(z)}\,,
\end{aligned}$$

are more suitable factorisations that satisfy the constraints of the above table. Furthermore, if we accept that $p(A \wedge B \mid X) = g(f_8(x,z)) = g(f_9(y,w))$, then we observe that

$$\begin{aligned}
g(f_{12}(x,y,w)) &= \sqrt{g(x)\,g(z)}\sqrt{g(y)\,g(w)}\,\sqrt{\frac{g(y)}{g(z)}} \;=\; p(A \wedge B \mid X)\,\sqrt{\frac{p(B \mid X)}{p(B \mid A \wedge X)}}\,, \\
g(f_{13}(x,y,z)) &= \sqrt{g(x)\,g(z)}\,\sqrt{g(y)\,g(w)}\,\sqrt{\frac{g(x)}{g(w)}} \;=\; p(A \wedge B \mid X)\,\sqrt{\frac{p(A \mid X)}{p(A \mid B \wedge X)}}\,, \\
g(f_{14}(x,w,z)) &= \sqrt{g(x)\,g(z)}\,\sqrt{g(y)\,g(w)}\,\sqrt{\frac{g(z)}{g(y)}} \;=\; p(A \wedge B \mid X)\,\sqrt{\frac{p(B \mid A \wedge X)}{p(B \mid X)}}\,, \\
g(f_{15}(y,w,z)) &= \sqrt{g(x)\,g(z)}\,\sqrt{g(y)\,g(w)}\,\sqrt{\frac{g(w)}{g(x)}} \;=\; p(A \wedge B \mid X)\,\sqrt{\frac{p(A \mid B \wedge X)}{p(A \mid X)}}\,, \\
g(f_{16}(x,y,w,z)) &= \sqrt{g(x)\,g(z)}\,\sqrt{g(y)\,g(w)} \;=\; p(A \wedge B \mid X)\,.
\end{aligned}$$

Thus, as Van Horn [2] notes, f_8 is clearly the simplest function (along with its equivalent f_9) out of the set $\{f_8, f_9, f_{12}, \ldots, f_{16}\}$ of plausible options.

2 References

[1] R.T. Cox (1946): "*Probability, frequency and reasonable expectation*"

[2] K.S. Van Horn (2003): "*Constructing a logic of plausible inference: a guide to Cox's Theorem*"

[3] M. Tribus (1969): "*Rational descriptions, decisions and designs*"