Appendix to *Once Upon a Time Series*

by Eric J. Daza, DrPH, MPS on 2024-09-05

The Problem

Let $Y_t = 1$ if I have a migraine attack on day t; let $Y_t = 0$ otherwise. Let $X_t = 1$ if I drink at least one cup of coffee that day; let $X_t = 0$ otherwise. The probability of a migraine on day t given how much coffee I drink that day X_t, and given the occurrence of an attack the day before Y_{t-1}, is $\Pr(Y_t = 1|X_t, Y_{t-1})$.

Suppose my own personally true causal mechanism is:

$$probit(\Pr(Y_t = 1|X_t, Y_{t-1})) = \beta_0 + \beta_X X_t + \beta_Y Y_{t-1} + \beta_{XY} X_t Y_{t-1} \quad (1)$$

Suppose we're interested in the APTE specified as the risk difference $\Pr(Y_t = 1|X_t = 1) - \Pr(Y_t = 1|X_t = 0)$ when X is randomized. Note that this quantity is marginalized over Y_{t-1}. How can I estimate this APTE when I never randomized X_t?

The Solution

Link

We often use a link function $\eta(\mu)$ to link the expectation of Y conditional on its predictors $\{X, W\}$, denoted $\mu = E(Y|X, W)$. When Y is binary, we typically use the link functions $\eta(\mu) = logit(\mu)$ or $\eta(\mu) = probit(\mu)$ to relate the conditional expectation of Y, μ, to its predictors. When Y is continuous, we typically use the identity link $\eta(\mu) = I(\mu) = \mu$.

Using this link function, we can specify the conditional APTE (i.e., conditioned on W) as $\eta(E(Y|X = 1, W)) - \eta(E(Y|X = 0, W))$ whenever X is not associated with Y—as when X is randomized. However, we are usually interested in the APTE defined using the marginalized quantity $E(Y|X = a)$, rather than the more conditional quantity $E(Y|X = a, W)$.

In particular, for a given link function η, we are usually interested in the APTE specified as $\delta_\eta = \eta(E(Y|X = 1)) - \eta(E(Y|X = 0))$ whenever X is not associated with Y. Because the value of X is set in this expression, each expectation is really just a function of Y. This may be clearer when using potential outcomes notation, which essentially indexes Y (as a superscript) based on a value of X: $\delta_\eta = \eta(E(Y^1)) - \eta(E(Y^0))$.

Law

Under the law of total expectation (LTE), we have $E(Y|X = a) = E_W\{E(Y|X = a, W)|X = a\}$.

When using the identity link, the LTE can be used to directly relate $\eta(E(Y|X = a, W))$ to $E(Y|X = a)$:

$$E_W\{\eta(E(Y|X = a, W))|X = a\} \quad = E_W\{E(Y|X = a, W)|X = a\}$$

$$= E(Y|X = a)$$

If X is randomized, we have $E_W\{E(Y|X = a, W)|X = a\} = E_W\{E(Y|X = a, W)\}$. G-computation simply sets the right-side expectation equal to the desired quantity: $E(Y|X = a)$ had X been randomized. In terms of potential outcomes, the g-formula states $E(Y^a) = E_W\{E(Y|X = a, W)\}$.

Probit

The probit link requires a bit more consideration. Let β represent the $p \times 1$ vector of p coefficient parameters, with first element β_0 for the intercept. Let V represent a conformable $1 \times p$ vector of coefficient values, with 1 as its first element and the other elements consisting of X and W variables. In equation (1), $X = X_t$, $W = Y_{t-1}$, $\beta = (\beta_0, \beta_X, \beta_Y, \beta_{XY})'$, and $V = (1, X_t, Y_{t-1}, X_t Y_{t-1})$. Let V^a represent V with X set to a. In equation (1), $V^a = (1, a, Y_{t-1}, aY_{t-1})$.

Let $Y \in \{0, 1\}$, and define $\mu(X) = E(Y|X) = \Pr(Y = 1|X)$ and $\mu(X, W) = E(Y|X, W) = \Pr(Y = 1|X, W)$. By the LTE, we have $\mu(X) = E_W\{\mu(X, W)|X\}$.

Suppose we first model $\mu(X, W)$ using the probit link as $probit(\mu(X, W)) = \Phi^{-1}(\mu(X, W)) = V\beta$, where $\Phi(z)$ is the cumulative distribution function of the standard normal distribution. Hence, $\mu(X, W) = \Phi(V\beta)$ and:

$$\mu(X) \quad = E_W\{\Phi(V\beta)|X\}$$

$$= E_W(\Phi(\beta_0 + \beta_X X + \beta_W W + \beta_{XW} XW)|X) \text{ for equation (1)}$$

$$= \int \Phi(\beta_0 + \beta_X X + \beta_W w + \beta_{XW} Xw)dF_W(w|X)$$

$$= \int \Phi(c^X + d^X w)dF_W(w|X) \text{ where } c^X = \beta_0 + \beta_X X \text{ and } d^X = \beta_W + \beta_{XW} X$$

$$= \int \Phi(c^X + d^X w) dF_W(w|X)$$

$$= \sum_{w=0}^{1} \Phi(c^X + d^X w) \Pr(W = w|X) \text{ because } W = Y_{t-1} \text{ and } Y \in \{0, 1\} \text{ for all } t$$

$$= \Phi(c^X) \Pr(W = 0|X) + \Phi(c^X + d^X) \Pr(W = 1|X)$$

$$= \Phi(c^X)\{1 - \Pr(W = 1|X)\} + \Phi(c^X + d^X) \Pr(W = 1|X)$$

$$= \Phi(c^X) - \Phi(c^X) \Pr(W = 1|X) + \Phi(c^X + d^X) \Pr(W = 1|X)$$

$$= \Phi(c^X) + \{\Phi(c^X + d^X) - \Phi(c^X)\} \Pr(W = 1|X)$$

Redux

In terms of equation (1), we have:

$$\mu(X_t) = \Phi(c^X) + \{\Phi(c^X + d^X) - \Phi(c^X)\} p_X$$

where

- $\mu(X_t) = E(Y_t|X_t) = \Pr(Y_t = 1|X_t)$
- $c^X = \beta_0 + \beta_X X_t$
- $d^X = \beta_Y + \beta_{XY} X_t$
- $p_X = \Pr(Y_{t-1} = 1|X_t)$

If X is randomized, then $p_X = \pi$ and:

$$\mu(X_t) = \Phi(c^X) + \{\Phi(c^X + d^X) - \Phi(c^X)\}\pi$$

Consider the observed risk difference:

$$d = E(Y_t|X_t = 1) - E(Y_t|X_t = 0)$$

$$= \Pr(Y_t = 1|X_t = 1) - \Pr(Y_t = 1|X_t = 0)$$

$$= \mu(1) - \mu(0)$$

$$= [\Phi(c^1) + \{\Phi(c^1 + d^1) - \Phi(c^1)\}p_1] - [\Phi(c^0) + \{\Phi(c^0 + d^0) - \Phi(c^0)\}p_0]$$

$$= \Phi(c^1) + \{\Phi(c^1 + d^1) - \Phi(c^1)\}p_1 - \Phi(c^0) - \{\Phi(c^0 + d^0) - \Phi(c^0)\}p_0$$

$$= \Phi(c^1) + \Phi(c^1 + d^1)p_1 - \Phi(c^1)p_1 - \Phi(c^0) - \Phi(c^0 + d^0)p_0 + \Phi(c^0)p_0$$

$$= \Phi(c^1) - \Phi(c^1)p_1 - \Phi(c^0) + \Phi(c^0)p_0 + \Phi(c^1 + d^1)p_1 - \Phi(c^0 + d^0)p_0$$

$$= \Phi(c^1)(1 - p_1) - \Phi(c^0)(1 - p_0) + \Phi(c^1 + d^1)p_1 - \Phi(c^0 + d^0)p_0$$

$$= \Phi(\beta_0 + \beta_X)(1 - p_1) - \Phi(\beta_0)(1 - p_0) +$$

$$\Phi(\beta_0 + \beta_X + \beta_Y + \beta_{XY})p_1 - \Phi(\beta_0 + \beta_Y)p_0$$

If X is randomized, then $p_X = \pi$ for $X \in \{0, 1\}$, and we have the APTE:

$$\delta_{probit} = \Phi(c^1)(1 - \pi) - \Phi(c^0)(1 - \pi) + \Phi(c^1 + d^1)\pi - \Phi(c^0 + d^0)\pi$$

$$= (\Phi(c^1) - \Phi(c^0))(1 - \pi) + (\Phi(c^1 + d^1) - \Phi(c^0 + d^0))\pi$$

$$= (\Phi(\beta_0 + \beta_X) - \Phi(\beta_0))(1 - \pi) +$$

$$(\Phi(\beta_0 + \beta_X + \beta_Y + \beta_{XY}) - \Phi(\beta_0 + \beta_Y))\pi$$

Note that $(1 - p_X) - (1 - \pi) = 1 - p_X - 1 + \pi = \pi - p_X$. Hence, the difference between this APTE and the non-randomized observed risk difference is:

$$d - \delta_{probit} = \{\Phi(c^1)(1 - p_1) - \Phi(c^0)(1 - p_0) + \Phi(c^1 + d^1)p_1 - \Phi(c^0 + d^0)p_0\} -$$

$$\{\Phi(c^1)(1 - \pi) - \Phi(c^0)(1 - \pi) + \Phi(c^1 + d^1)\pi - \Phi(c^0 + d^0)\pi\}$$

$$= \Phi(c^1)(\pi - p_1) - \Phi(c^0)(\pi - p_0) + \Phi(c^1 + d^1)(p_1 - \pi) - \Phi(c^0 + d^0)(p_0 - \pi)$$

$$= \Phi(\beta_0 + \beta_X)(\pi - p_1) - \Phi(\beta_0)(\pi - p_0) +$$

$$\Phi(\beta_0 + \beta_X + \beta_Y + \beta_{XY})(p_1 - \pi) - \Phi(\beta_0 + \beta_Y)(p_0 - \pi)$$

Bayes

Bayes' theorem states:

$$p_a \equiv \Pr(Y_{t-1} = 1 | X_t = a)$$

$$= \frac{\Pr(X_t = a | Y_{t-1} = 1)\Pr(Y_{t-1} = 1)}{\Pr(X_t = a)}$$

$$= \frac{\Pr(X_t=a|Y_{t-1}=1)\pi}{\Pr(X_t=a)}$$

Let $\pi_X = \Pr(X_t = 1) = \Pr(X = 1)$ represent the overall probability of drinking more than one cup of coffee on any given day. Hence:

$$p_1 - p_0 = \Pr(Y_{t-1} = 1|X_t = 1) - \Pr(Y_{t-1} = 1|X_t = 0)$$

$$= \frac{\Pr(X_t=1|Y_{t-1}=1)\pi}{\Pr(X_t=1)} - \frac{\Pr(X_t=0|Y_{t-1}=1)\pi}{\Pr(X_t=0)}$$

$$= \frac{\Pr(X_t=1|Y_{t-1}=1)\pi}{\pi_X} - \frac{\{1-\Pr(X_t=1|Y_{t-1}=1)\}\pi}{1-\pi_X}$$

$$= \frac{\Pr(X_t=1|Y_{t-1}=1)\pi}{\pi_X} - \frac{\{1-\Pr(X_t=1|Y_{t-1}=1)\}\pi}{1-\pi_X}$$

$$= \frac{\Pr(X_t=1|Y_{t-1}=1)\pi(1-\pi_X)}{\pi_X(1-\pi_X)} - \frac{\pi_X\{1-\Pr(X_t=1|Y_{t-1}=1)\}\pi}{\pi_X(1-\pi_X)}$$

$$= \frac{\Pr(X_t=1|Y_{t-1}=1)\pi(1-\pi_X)-\pi_X\{1-\Pr(X_t=1|Y_{t-1}=1)\}\pi}{\pi_X(1-\pi_X)}$$

$$= \frac{\Pr(X_t=1|Y_{t-1}=1)\pi-\Pr(X_t=1|Y_{t-1}=1)\pi\pi_X-\pi\pi_X+\Pr(X_t=1|Y_{t-1}=1)\pi\pi_X}{\pi_X(1-\pi_X)}$$

$$= \frac{\Pr(X_t=1|Y_{t-1}=1)\pi-\pi\pi_X}{\pi_X(1-\pi_X)}$$

$$= \frac{\pi\{\Pr(X_t=1|Y_{t-1}=1)-\pi_X\}}{\pi_X(1-\pi_X)}$$

If X_t and Y_{t-1} are unassociated, then $\Pr(X_t = 1|Y_{t-1} = 1) = \Pr(X_t = 1) = \pi_X$. Hence:

$$p_1 - p_0 = \frac{\pi\{\Pr(X_t=1|Y_{t-1}=1)-\pi_X\}}{\pi_X(1-\pi_X)}$$

$$= \frac{\pi(\pi_X-\pi_X)}{\pi_X(1-\pi_X)}$$

$$= \frac{0}{\pi_X(1-\pi_X)}$$

$$= 0$$

All Together Now

The relevant parameters for the example we used in the section “A Head of My Self” are:

- $\beta_X = 0$ because "drinking more coffee didn't increase my migraine chances" (i.e., direct effect)
- $\Pr(X_t = 1 | Y_{t-1} = 1) > \pi_X$ because "getting a migraine the day before caused me to sleep less—and this tended to make me drink more coffee the next day"
- $\beta_Y > 0$ because "a migraine attack yesterday also directly increased my chances of getting a migraine today"
- $\beta_{XY} > 0$ because "whenever I had a migraine attack the day before, this gave coffee the ability to cause a migraine attack the next day" (i.e., effect modification)

These imply:

$$
\begin{aligned}
d &= \Phi(\beta_0)(1 - p_1) - \Phi(\beta_0)(1 - p_0) + \\
&\qquad \Phi(\beta_0 + \beta_Y + \beta_{XY})p_1 - \Phi(\beta_0 + \beta_Y)p_0 \\
&= \Phi(\beta_0)\{(1 - p_1) - (1 - p_0)\} + \\
&\qquad \Phi(\beta_0 + \beta_Y + \beta_{XY})p_1 - \Phi(\beta_0 + \beta_Y)p_0 \\
&= \Phi(\beta_0)(p_0 - p_1) + \Phi(\beta_0 + \beta_Y + \beta_{XY})p_1 - \Phi(\beta_0 + \beta_Y)p_0 \\
\delta_{probit} &= (\Phi(\beta_0) - \Phi(\beta_0))(1 - \pi) + (\Phi(\beta_0 + \beta_Y + \beta_{XY}) - \Phi(\beta_0 + \beta_Y))\pi \\
&= (\Phi(\beta_0 + \beta_Y + \beta_{XY}) - \Phi(\beta_0 + \beta_Y))\pi \\
&= \Phi(\beta_0 + \beta_Y + \beta_{XY})\pi - \Phi(\beta_0 + \beta_Y)\pi \\
d - \delta_{probit} &= \Phi(\beta_0)(\pi - p_1) - \Phi(\beta_0)(\pi - p_0) + \\
&\qquad \Phi(\beta_0 + \beta_Y + \beta_{XY})(p_1 - \pi) - \Phi(\beta_0 + \beta_Y)(p_0 - \pi) \\
&= \Phi(\beta_0)(p_0 - p_1) + \Phi(\beta_0 + \beta_Y + \beta_{XY})(p_1 - \pi) - \Phi(\beta_0 + \beta_Y)(p_0 - \pi)
\end{aligned}
$$

In the example, we assumed $d >> \delta_{probit}$. This made me conclude that drinking more coffee increased my migraine chances—a spurious conclusion about a direct effect that was actually zero. (The overall average effect was slightly greater than zero due to effect modification by lagged migraine attacks.)