Defining a model and deriving the likelihood for Q3

Define $X_{4i} = \begin{cases} 1 & \text{person } i \text{ has GPA} \geq 3.5 \\ 0 & \text{otherwise} \end{cases}$

$X_{1i} = \begin{cases} 1 & \text{person } i \text{ has used study drugs} \\ 0 & \text{otherwise} \end{cases}$

Model

$$X_{4i} \sim Ber(\theta)$$

$$X_{1i} \mid X_{4i} = 1 \sim Ber(\psi_1)$$

$$X_{1i} \mid X_{4i} = 0 \sim Ber(\psi_0)$$

Note! This is based on an assumption of drawing a person at random from the population. As was pointed out in class, θ, ψ_1, and ψ_0 are descriptive of the population and we can't infer causality.

Further, define

$Y_{4i} = \begin{cases} 1 & \text{person } i \text{ responds 1 to Q4} \\ 0 & \text{otherwise} \end{cases}$

$Y_{1i} = \begin{cases} 1 & \text{person } i \text{ responds 2 or 3 to Q1} \\ 0 & \text{otherwise} \end{cases}$

D_{11i}, D_{12i} die rolls on Q1 for person i

D_{41i}, D_{42i} " Q4 "

The likelihood is

$$\mathcal{L}(\theta, \psi_1, \psi_0) = \prod_{i=1}^{n} p(y_{4i}, y_{1i} \mid \theta, \psi_1, \psi_0)$$

(assuming people answer the survey independently)

$$\boxed{= \prod_{i=1}^{n} p(y_{4i} \mid \theta, \psi_1, \psi_0)\, p(y_{1i} \mid y_{4i}, \theta, \psi_1, \psi_0)}$$

Now let me simplify these.
I will temporarily drop the index i.

$Y_{4i} = 0$ or 1, so I need to find the probability of each in terms of the parameters. (This is similar to the calculation you did before.)

answer according to D_{42}

$$P(Y_4 = 1 \mid \theta, \psi_1, \psi_0) = P(Y_4 = 1 \mid D_{41} = 1 \text{ or } 2, \theta, \psi_1, \psi_0) \times P(D_{41} = 1 \text{ or } 2 \mid \theta, \psi_1, \psi_0)$$

$$+ P(Y_4 = 1 \mid D_{41} = 3, 4, 5 \text{ or } 6, \theta, \psi_1, \psi_0) \times P(D_{41} = 3, 4, 5 \text{ or } 6 \mid \theta, \psi_1, \psi_0)$$

I'm conditioning on all variables for completeness. You can see that most of them will drop out.

answer according to X_4

$$= P(D_{42} = 1)\, P(D_{41} = 1 \text{ or } 2)$$

$$+ P(X_4 = 1 \mid \theta)\, P(D_{41} = 3, 4, 5 \text{ or } 6)$$

$$= \frac{1}{6} \cdot \frac{1}{3} + \theta \cdot \frac{2}{3} = \frac{1}{18} + \frac{2}{3}\theta$$

So $\boxed{p(y_{4i} \mid \theta, \psi_1, \psi_0) = \left[\frac{1}{18} + \frac{2}{3}\theta\right]^{y_i} \left[1 - \frac{1}{18} - \frac{2}{3}\theta\right]^{1-y_i}}$

Moving now to $p(y_{1i} \mid y_{4i}, \theta, \psi_1, \psi_0)$ ...

As with Y_{4i}, Y_{1i} is either 1 or 0, so our problem is to find $P(Y_{1i} = 1 \mid Y_{4i}, \theta, \psi_1, \psi_0)$. Drop i again.

answer according to D_{12}

$$P(Y_1 = 1 \mid Y_4, \theta, \psi_1, \psi_0) = P(Y_1 = 1 \mid Y_4, \theta, \psi_1, \psi_0, D_{11} = 1 \text{ or } 2) \times P(D_{11} = 1 \text{ or } 2 \mid Y_4, \theta, \psi_1, \psi_0)$$

$$+ P(Y_1 = 1 \mid Y_4, \theta, \psi_1, \psi_0, D_{11} = 3, 4, 5, \text{ or } 6) \times P(D_{11} = 3, 4, 5 \text{ or } 6 \mid Y_4, \theta, \psi_1, \psi_0)$$

answer according to X_1

$$= P(D_{12} = 3, 4, 5, 6)\, P(D_{11} = 1 \text{ or } 2) + P(X_1 = 1 \mid Y_4, \theta, \psi, \psi_0)\, P(D_{11} = 3, 4, 5 \text{ or } 6)$$

$$= \frac{2}{3} \cdot \frac{1}{3} + \frac{2}{3} \underbrace{P(X_1 = 1 \mid Y_4, \theta, \psi_1, \psi_0)}_{\text{need to simplify this}}$$

$$P(X_1 = 1 \mid Y_4, \theta, \psi_1, \psi_0)$$

* now I'm going to condition on the truth for Q4

$$= \overbrace{P(X_1 = 1 \mid X_4 = 1, Y_4, \theta, \psi_1, \psi_0)}^{=\psi_1} P(X_4 = 1 \mid Y_4, \theta, \psi_1, \psi_0)$$

$$+ \underbrace{P(X_1 = 1 \mid X_4 = 0, Y_4, \theta, \psi_1, \psi_0)}_{=\psi_0} P(X_4 = 0 \mid Y_4, \theta, \psi_1, \psi_0)$$

$$P(X_4 = 1 \mid Y_4 = 1, \theta, \psi_1, \psi_0) = \frac{P(X_4 = 1, Y_4 = 1 \mid \theta)}{P(Y_4 = 1 \mid \theta)}$$

To obtain these, I used the table on the following page.

$$= \frac{\frac{13}{18}\theta}{\frac{1}{18} + \frac{2}{3}\theta} \equiv g_1(\theta)$$

$$P(X_4 = 1 \mid Y_4 = 0, \theta, \psi_1, \psi_0) = \frac{P(X_4 = 1, Y_4 = 0 \mid \theta)}{P(Y_4 = 0 \mid \theta)}$$

$$= \frac{\frac{5}{18}\theta}{1 - \frac{1}{18} - \frac{2}{3}\theta} \equiv g_0(\theta)$$

X_4	D_{41}	D_{42}	Y_4 (determined by the others)	probability
0	1-2	1	1	$(1-\theta)\frac{1}{3}\frac{1}{6}$
0	1-2	2-6	0	$(1-\theta)\frac{1}{3}\frac{5}{6}$
0	3-6	1	0	$(1-\theta)\frac{2}{3}\frac{1}{6}$
0	3-6	2-6	0	$(1-\theta)\frac{2}{3}\frac{5}{6}$
1	1-2	1	1	$\theta\frac{1}{3}\frac{1}{6}$
1	1-2	2-6	0	$\theta\frac{1}{3}\frac{5}{6}$
1	3-6	1	1	$\theta\frac{2}{3}\frac{1}{6}$
1	3-6	2-6	1	$\theta\frac{2}{3}\frac{5}{6}$

Shortening the notation from page 4, I can write

$$X_4 \mid Y_4, \theta, \psi_1, \psi_0 \sim \text{Ber}\left(g_{Y_4}(\theta) \right)$$

i.e., use g_1 or g_0 depending on Y_4

Putting it all together,

$$P(Y_1 = 1 \mid Y_4, \theta, \psi_1, \psi_0) = \frac{2}{9} + \frac{2}{3} P(X_1 = 1 \mid Y_4, \theta, \psi_1, \psi_0)$$

(bottom of page 3)

$$= \frac{2}{9} + \frac{2}{3}\left[\psi_1 \, g_{Y_4}(\theta) + \psi_0 \left(1 - g_{Y_4}(\theta) \right) \right]$$

$$\equiv f(\theta, \psi_1, \psi_0, Y_4)$$

and the contribution to the likelihood is

$$\boxed{p(y_{1i} \mid y_{4i}, \theta, \psi_1, \psi_0) = f(\theta, \psi_1, \psi_0, y_{4i})^{y_{1i}} \left[1 - f(\theta, \psi_1, \psi_0, y_{4i}) \right]^{1-y_{1i}}}$$