■練習問題 3.2.1

練習問題 1.5.2（図 1.10）とそこにあるパラメターについて、

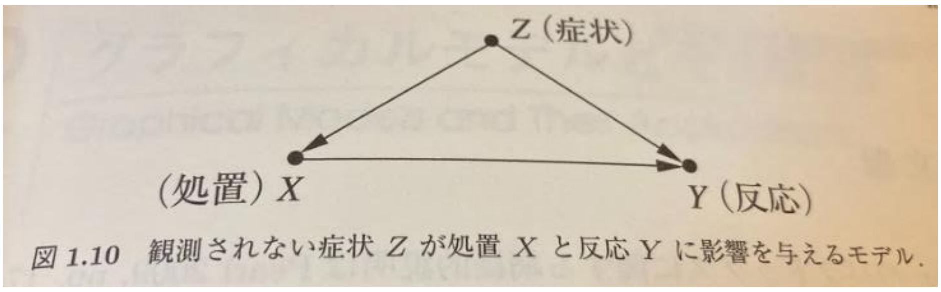

図 1.10 観測されない症状 Z が処置 X と反応 Y に影響を与えるモデル.

割合 r は死に至る症状 Z を持つ（Z は症状の有無。z_1:症状あり）

$$P(z_1) = r$$

X は延命効果のある薬の服薬有無（x_1:服薬）、Y は生死（y_1:死）を表す。各条件付き確率は、

$$P(Y = y_1 | X = x_0, Z = z_0) = p_1$$
$$P(Y = y_1 | X = x_1, Z = z_0) = p_2$$
$$P(Y = y_1 | X = x_0, Z = z_1) = p_3$$
$$P(Y = y_1 | X = x_1, Z = z_1) = p_4$$

で表されるとする。さらに症状のある患者は薬を避ける傾向があり、薬を飲む確率を下記とする

$$P(x_1 | z_0) = q_1$$
$$P(x_1 | z_1) = q_2$$

(a)モデルに do(x)の介入を施すことにより、すべての x,y の値について$P(y|do(x))$を計算せよ

介入前が図 3.3（Pと表す）、介入後が図 3.4（P_mと表す）である。教科書の議論の通り、

$$P_m(Y = y | Z = z, X = x) = P(Y = y | Z = z, X = x)$$
$$P_m(Z = z) = P(Z = z)$$

介入後のP_mにおいては X と Z が独立

であるから、

$$P\big(Y = y_1 \big| do(X = x_1)\big) = P_m(Y = y_1 | X = x_1) = \sum_z P_m(Y = y_1 | X = x_1, Z = z) P_m(Z = z | X = x_1)$$

$$= \sum_z P_m(Y = y_1 | X = x_1, Z = z) P_m(Z = z)$$

$$= \sum_z P(Y = y_1 | X = x_1, Z = z) P(Z = z)$$

$$P\big(Y = y_1 \big| do(X = x_0)\big) = P_m(Y = y_1 | X = x_0) = \cdots$$
$$P\big(Y = y_0 \big| do(X = x_1)\big) = P_m(Y = y_0 | X = x_1) = \cdots$$
$$P\big(Y = y_0 \big| do(X = x_0)\big) = P_m(Y = y_0 | X = x_0) = \cdots$$

(下 3 つも同様、具体的な計算は(b)へ)

(b)調整化公式 3.5 を使うことにより、$P(y|do(x))$をすべての x,y の値について計算せよ
調整化公式（または(a)の結果から）

$$P(Y=y|do(X=x)) = \sum_z P(Y=y|X=x, Z=z)P(Z=z)$$

より、

$$\begin{aligned}P(Y=y_1|do(X=x_1)) &= \sum_z P(Y=y_1|X=x_1, Z=z)P(Z=z)\\ &= P(Y=y_1|X=x_1, Z=z_0)P(Z=z_0) + P(Y=y_1|X=x_1, Z=z_1)P(Z=z_1)\\ &= p_2(1-r)+p_4 r\end{aligned}$$

$$\begin{aligned}P(Y=y_1|do(X=x_0)) &= \sum_z P(Y=y_1|X=x_0, Z=z)P(Z=z)\\ &= P(Y=y_1|X=x_0, Z=z_0)P(Z=z_0) + P(Y=y_1|X=x_0, Z=z_1)P(Z=z_1)\\ &= p_1(1-r)+p_3 r\end{aligned}$$

$$\begin{aligned}P(Y=y_0|do(X=x_1)) &= \sum_z P(Y=y_0|X=x_1, Z=z)P(Z=z)\\ &= P(Y=y_0|X=x_1, Z=z_0)P(Z=z_0) + P(Y=y_0|X=x_1, Z=z_1)P(Z=z_1)\\ &= (1-p_2)(1-r)+(1-p_4) r\end{aligned}$$

$$\begin{aligned}P(Y=y_0|do(X=x_0)) &= \sum_z P(Y=y_0|X=x_0, Z=z)P(Z=z)\\ &= P(Y=y_0|X=x_0, Z=z_0)P(Z=z_0) + P(Y=y_0|X=x_0, Z=z_1)P(Z=z_1)\\ &= (1-p_1)(1-r)+(1-p_3) r\end{aligned}$$

(c)ACE を計算せよ

$$ACE = P(y_1|do(x_1)) - P(y_1|do(x_0))$$

(b)の結果より、下記の通り。

$$ACE = P(y_1|do(x_1)) - P(y_1|do(x_0)) = (p_2(1-r)+p_4 r) - (p_1(1-r)+p_3 r)$$

また、リスク差（RD:risk difference）を計算せよ

$$RD = P(y_1|x_1) - P(y_1|x_0)$$

$$\mathrm{RD} = \mathrm{P}(\mathrm{y}_1|\mathrm{x}_1) - \mathrm{P}(\mathrm{y}_1|\mathrm{x}_0) = \sum_z \left(P(y_1, z|x_1) - P(y_1, z|x_0)\right) = \sum_z \left(\frac{P(y_1, z, x_1)}{P(x_1)} - \frac{P(y_1, z, x_0)}{P(x_0)}\right)$$

$$= \sum_z \left(\frac{P(y_1|z, x_1)P(z, x_1)}{P(x_1)} - \frac{P(y_1|z, x_0)P(z, x_0)}{P(x_0)}\right)$$

$$= \sum_z \left(P(y_1|z, x_1)\frac{P(x_1|z)P(z)}{\sum_{z'} P(x_1|z')P(z')} - P(y_1|z, x_0)\frac{P(x_0|z)P(z)}{\sum_{z'} P(x_0|z')P(z')}\right)$$

$$= \frac{\mathrm{p}_2(1-r)q_1 + p_4 r q_2}{q_1(1-r) + q_2 r} - \frac{\mathrm{p}_1(1-r)(1-q_1) + p_3 r(1-q_2)}{(1-q_1)(1-r) + (1-q_2)r}$$

1 式目から 4 式目くらいまで(3.3)式の変形でも OK

ACE と RD の違いは何か。パラメターをどのような値にすればこの差を最小化することができるか

$$\mathrm{ACE} = (\mathrm{p}_2(1-r) + p_4 r) - (\mathrm{p}_1(1-r) + p_3 r)$$

$$\mathrm{RD} = \frac{\mathrm{p}_2(1-r)q_1 + p_4 r q_2}{q_1(1-r) + q_2 r} - \frac{\mathrm{p}_1(1-r)(1-q_1) + p_3 r(1-q_2)}{(1-q_1)(1-r) + (1-q_2)r}$$

赤字の通り、RD は条件付き確率（症状のある患者は薬を避ける傾向にあり、薬を飲む確率がそれぞれq_1とq_2）で重み付けがされている。

X と Z が独立なとき、ACE と RD が一致することがわかる（パラメターでいうところの$q_1 = q_2$。症状のある患者は薬を避ける傾向が特にない状態）

(d)練習問題 1.5.2(c)のように、Simpson のパラドックスが起きるようなパラメターを見つけよ。新薬の因果効果は層別しない統合データから得られることをはっきりと示せ

リスク差

$$\mathrm{RD} = \mathrm{P}(\mathrm{y}_1|\mathrm{x}_1) - \mathrm{P}(\mathrm{y}_1|\mathrm{x}_0) = \frac{\mathrm{p}_2(1-r)q_1 + p_4 r q_2}{q_1(1-r) + q_2 r} - \frac{\mathrm{p}_1(1-r)(1-q_1) + p_3 r(1-q_2)}{(1-q_1)(1-r) + (1-q_2)r}$$

が負（全体としては服薬効果あり）かつ、と

$\mathrm{P}(\mathrm{y}_1|\mathrm{x}_1, z_1) - \mathrm{P}(\mathrm{y}_1|\mathrm{x}_0, z_1) = \mathrm{p}_4 - p_3$

$\mathrm{P}(\mathrm{y}_1|\mathrm{x}_1, z_0) - \mathrm{P}(\mathrm{y}_1|\mathrm{x}_0, z_0) = \mathrm{p}_2 - p_1$

が両方とも正（症状 Z の有無で条件つけると、服薬が逆効果に）となるパラメターを探す。

p1	p2	p3	p4	q1	q2	r
0.1	0.3	0.7	0.9	0.8	0.2	0.3

このとき、

RD	−0.12088
$p_4 - p_3$	0.2
$p_2 - p_1$	0.2
ACE	0.2

となり Simpson のパラドックスとなっている。

ちなみに$q_1 = q_2$のときは下記の通り。(c)の通り、RD=ACE となる

p1	p2	p3	p4	q1	q2	r
0.1	0.3	0.7	0.9	0.8	0.8	0.3

RD	0.2
$p_4 - p_3$	0.2
$p_2 - p_1$	0.2
ACE	0.2

$$\mathrm{ACE} = (p_2(1-r) + p_4 r) - (p_1(1-r) + p_3 r) = (p_2 - p_1)(1-r) + (p_4 - p_3)r$$

より、

$P(y_1|x_1, z_1) - P(y_1|x_0, z_1) = p_4 - p_3$

$P(y_1|x_1, z_0) - P(y_1|x_0, z_0) = p_2 - p_1$

が両方とも正（症状 Z の有無で条件つけると、服薬が逆効果に）のとき、ACE も正となり、層別しない統合データから因果効果が得られている事がわかる。

■練習問題 4.3.1

図 4.3 のモデルにおいてU_1とU_2は共に正規分布に従う独立した確率変数であるとする。これはともに平均 0、分散 1 を持つ

(a)スキルレベルが Z=z であるような労働者が、もし x 年の大学教育を受けていたならば年収の期待値はいくらになるか。

ヒント：e を Z=z として定理 4.3.2 を使う。また任意の 2 つの正規分布する確率変数、例えば X と Z について$E[X|Z = z] = E[x] + R_{XZ}(z - E[Z])$である。3.8.2 項と 3.8.3 項での議論を参考に、すべての回帰係数を構造方程式の係数を用いて表現せよ。そして$E[Y_X|Z = z] = abx + bz/(1 + a^2)$を示せ。

(b)全問(a)の結果を利用し、特定のスキルレベルにおいて教育が年収に与える影響はスキルレベルに依存しないことを示せ。

※図 4.3 のモデルは大学教育（X)、スキル（Z)、年収（Y）の間の因果関係を示したモデル

$$X = U_1$$
$$Z = aX + U_2$$
$$Y = bZ$$

X は大学教育の有無、U_2は職務経験の有無、Z はスキルレベル、といったイメージ

(a) $E[Y_X|Z = z]$を求めていく。以下の定理 4.3.2 より、

定理 4.3.2

X から Y の総合効果の傾きをτとする。

$$\tau = E[Y|do(x + 1)] - E[Y|do(x)]$$

このときどの証拠$Z = e$についても、

$$E[Y_{X=x}|Z = e] = E[Y|Z = e] + \tau(x - E[X|Z = e])$$

が成り立つ。

$E[Y_X|Z = z] = E[Y|Z = z] + \tau(x - E[X|Z = z])$

ここでヒント（とU_1とU_2の独立性）より、

$$E[X|Z = z] = E[X] + R_{XZ}(z - E[Z]) = 0 + a\frac{\sigma_X^2}{\sigma_Z^2}(z - 0) = a\frac{1}{E[(aX + U_2)^2]}z = \frac{az}{(a^2 + 1)}$$

また線形システムであることから、$\tau = ab$

（$Y = bZ = b(aX + U_1) = abX + bU_1$）

以上より、

$$\mathrm{E}[Y_X|\mathrm{Z}=\mathrm{z}] = \mathrm{E}[bZ|\mathrm{Z}=\mathrm{z}] + ab\left(x - \frac{az}{(a^2+1)}\right) = \mathrm{bz} + abx - \frac{a^2bz}{(a^2+1)} = abx + \frac{bz}{(a^2+1)}$$

(b)

(a)の結果より、

$$\mathrm{E}[\mathrm{Y}_1 - \mathrm{Y}_0|Z=z] = \mathrm{E}[\mathrm{Y}_1|Z=z] - \mathrm{E}[\mathrm{Y}_0|Z=z] = ab + \frac{bz}{(a^2+1)} - \frac{bz}{(a^2+1)} = ab$$

（スキルレベル z に依存しないことがわかる

（削除）

$\mathrm{E}[\mathrm{Y}_1 - \mathrm{Y}_0|Z=z] = \mathrm{E}[\mathrm{Y}_1 - \mathrm{Y}_0]$を示せば良い。

$$\mathrm{E}[\mathrm{Y}_1 - \mathrm{Y}_0] = \sum_z \mathrm{E}[\mathrm{Y}_1 - \mathrm{Y}_0|Z=z]\,P(Z=z) = ab\sum_z P(Z=z) = ab$$

よって、$\mathrm{E}[\mathrm{Y}_1 - \mathrm{Y}_0|Z=z] = \mathrm{E}[\mathrm{Y}_1 - \mathrm{Y}_0]$