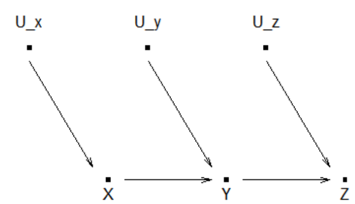
Stat 265 - HW 3 Solutions/Comments (Spring 2018)

1. **Graphs.**
   (a) Parents are nodes with a direct link to the given node. W and Y are parents of Z.
   (b) Ancestors include nodes whose descendents include a parent of the given node; W, X, Y are ancestors of Z.
   (c) Children are nodes that the given node links directly to. Y and Z are children of W.
   (d) Descendents include nodes that are children of the node's children and so on; Y, Z, T are descendents of W.
   (e) Note paths here do not need to be directed (i.e., we can allow for colliders and forks). Paths are: XWZ, XWYZ, XYZ, XYTZ, XYWZ, XWYTZ.
   (f) Directed paths between X and T are: XYT, XYZT, XWYT, XWZT, XWYZT.
   (g) According to the product decomposition the joint distribution is given by multiplying the conditional distribution of each node given its parents. $p(x, w, y, z, t) = p(x)p(w|x)p(y|x, w)p(z|w, y)p(t|y, z)$.

2. **Structural causal models.**
   (a) Graph:



   (b) $E(X) = E(U_x) = 0;\ \ E(Y) = E(X)/3 + E(U_y) = 0;\ \ E(Z) = E(Y)/16 + E(U_z) = 0;$
   (c) $E(Z|Y = 3) = 3/16 + E(U_z) = 3/16$
   (d) $E(Z|X = 3) = E(Y|X = 3)/16 + E(U_z) = 1/16 + E(U_y/16) + E(U_z) = 1/16$
   (e) $E(Z|X = 1, Y = 3) = 3/16$ because Z depends only on Y and U_z. More formally this is true because we can tell from the graph or the joint distribution ($p(x, y, z) = p(x)p(y|x)p(z|x, y)$) that X and Z are independent given Y.

3. **d-separation.**
   (a) W and T are not marginally independent. We know that they are (likely) dependent because of the directed path $W \to Y \to T$. Recall that because the actual conditional distributions (from 1(g)) are not known we can't say for sure that they not independent. Our approach is to look for independencies (or conditional independencies) that hold for any choice of the distributions. The only way to mathematically prove this would be to derive $p(w, t)$ from 1(g) and show that it is not equal to $p(w)p(t)$. Here you just get some messy formulas!
   (b) Conditioning on Z blocks several of the paths from W to T but it does not block the path mentioned above. Thus W and T are not conditionally independent given Z.
   (c) Conditioning on Y and Z blocks all of the paths through which W impacts T so W and T are conditionally independent given Y and Z. Proof:

$$p(w, t|y, z) = \frac{p(w, t, y, z)}{p(y, z)} = \frac{\sum_x p(x)p(w|x)p(y|x, w)p(z|w, y)p(t|y, z)}{p(y, z)} = \frac{\sum_x p(x, w, y)p(z|w, y)p(t|y, z)}{p(y, z)}$$
$$= \frac{p(w, y)p(z|w, y)p(t|y, z)}{p(y, z)} = \frac{p(w, y)p(z|w, y)}{p(y, z)}p(t|y, z) = p(w|y, z)p(t|y, z)$$

   (d) Notice that there are a series of directed paths from X to T and there are all blocked by choosing the conditioning set to be $(W, Y), (Y, Z), (W, Y, Z)$. Any of those sets will d-separate X and T. This means that X and T are conditionally independent given any of those sets.

4. **Interventions.** Note that the paragraph translates to:
$Pr(Z=1)=r, Pr(X=1|Z=0)=q_1, Pr(X=1|Z=1)=q_2,$
$Pr(Y=1|Z=0,X=0)=p_1, Pr(Y=1|Z=0,X=1)=p_2,$
$Pr(Y=1|Z=1,X=0)=p_3, Pr(Y=1|Z=1,X=1)=p_4.$

(a) The joint distribution is specified in the table below.

Z	X	Y	$Pr(Z=z,X=x,Y=y)$	Z	X	Y	$Pr(Z=z,X=x,Y=y)$
0	0	0	$(1-r)(1-q_1)(1-p_1)$	1	0	0	$r(1-q_2)(1-p_3)$
0	0	1	$(1-r)(1-q_1)p_1$	1	0	1	$r(1-q_2)p_3$
0	1	0	$(1-r)q_1(1-p_2)$	1	1	0	$rq_2(1-p_4)$
0	1	1	$(1-r)q_1p_2$	1	1	1	rq_2p_4

(b) This is looking at conditional probabilities (i.e., probabilities in the relevant subpopulations (takers and nontakers) but averages over syndrome status.

$$Pr(Y=1|X=1)-Pr(Y=1|X=0)=\frac{Pr(Y=1,X=1)}{Pr(X=1)}-\frac{Pr(Y=1,X=0)}{Pr(X=0)}$$

$$=\frac{\sum_z Pr(Z=z,X=1,Y=1)}{\sum_z\sum_y Pr(Z=z,X=1,Y=y)}-\frac{\sum_z Pr(Z=z,X=0,Y=1)}{\sum_z\sum_y Pr(Z=z,X=0,Y=y)}$$

$$=\frac{(1-r)q_1p_2+rq_2p_4}{(1-r)q_1+rq_2}-\frac{(1-r)(1-q_1)p_1+r(1-q_2)p_3}{(1-r)(1-q_1)+r(1-q_2)}$$

(c) This is easily computed since it is a function of the specified probabilities: $Pr(Y=1|X=1,Z=1)-Pr(Y=1|X=0,Z=1)=p_4-p_3$

(d) This is easily computed since it is a function of the specified probabilities: $Pr(Y=1|X=1,Z=0)-Pr(Y=1|X=0,Z=0)=p_2-p_1$

(e) The key to coming up with an example is to note that the expression in (b) is comparing a weighted average of p_2 and p_4 to a weighted average of p_1 and p_3. We know $p_4<p_3$ and $p_2<p_1$. For the weighted averages to reverse the association we must have a big difference between the pairs (i.e., p_3,p_4 larger than p_1,p_2 or vice versa) and then have the q's be unequal in an appropriate way. If p_3,p_4 are the large numbers, then we want q_2 large and q_1 small. For example, suppose that $p_1=0.2, p_2=0.1, p_3=0.9, p_4=0.8$ so that the drug lowers death rates in both syndrome patients (from .9 to .8) and nonsyndrome patients (from .2 to .1) ... but the death rate is much higher in those with the syndrome. Choosing $r=0.5$ (just to elimiinate from the formula in (b)),$q_1=.2, q_2=.8$ so the syndrome makes people more likely to take the drug, we find that the drug effect appears harmful (i.e., increases the death rate) from .34 in the no drug group to .66 in the drug group.

(f) Recall that for this graph $Pr(Y=y|do(X=x))$ is obtained using the adjustment formula and averaging over Z, $Pr(Y=y|do(X=x))=\sum_z Pr((Y=y|X=x,Z=z)Pr(Z=z)$. This yields: $Pr(Y=1|do(X=1))=(1-r)p_2+rp_4$, $Pr(Y=0|do(X=1))=(1-r)(1-p_2)+r(1-p_4)=1-Pr(Y=1|do(X=1))$, $Pr(Y=1|do(X=0))=(1-r)p_1+rp_3$, $Pr(Y=0|do(X=0))=(1-r)(1-p_1)+r(1-p_3)$.

(g) The average treatment effect is $Pr(Y=1|do(X=1))-Pr(Y=1|do(X=0))=(1-r)(p_2-p_1)+r(p_4-p_3)$. There was some confusion here. The $do(\cdot)$ approach is not equivalent to conditioning on X. As described above, conditioning on X restricts attention to the subpopulation with that value of X (without regard to other variables like Z that may differ in the $X=0$ and $X=1$ condition). The $do(\cdot)$ calculation is identifying what would happen in the **entire population** if we carried out an intervention that made everyone in the population take the drug. This makes us weight the treatment effects according to the distribution of the pre-treatment (and very important) covariate Z. Thus the $do(\cdot)$ approach is the one that we want for a correct causal interpretation.