Regression models

Arvid Sjölander

Department of Medical Epidemiology and Biostatistics
Karolinska Institutet

A short course on concepts and methods in Causal Inference

Assumptions so far

- We have considered ideal conditions
  - infinite samples
  - binary variables
- Under these ideal conditions, we can use non-parametric analyses
  - e.g. stratification

In reality

- In real studies, we often have
  - small/moderate sample sizes
  - continuous and categorical variables
- Under these conditions, non-parametric analyses are typically not feasible
  - few subjects in each stratum
  - unstable estimates (e.g. wide confidence intervals, large p-values)

Outcome models

- To obtain more stable estimates, regression models are commonly used
  - e.g. linear regression, logistic regression, Cox proportional hazards regression
- Standard regression model are models for the outcome
- By default, outcome models estimate conditional causal effects
  - a parametric analog to stratification
- Outcome regression models can also be used to estimate marginal causal effects
  - parametric standardization

Exposure models

- Less known, marginal causal effects can also be estimated with a regression model for the exposure
- Modeling the exposure is particularly attractive when
  - the mechanisms that bring about the exposure are well understood
  - when there are multiple exposures, and covariates that are affected by previous exposures and affect later exposures

Outline

Outline

Outline

Motivating example

- Suppose we carry out an observational study to estimate the causal effect of AZT on infection risk for AIDS patients
- 1000 subjects enrolled
- Baseline measures:
  - CD4 count (L; counts/μl)
  - AZT level (A; '0' for 'untreated', '1' for 'treated')
- At end of follow up we measure:
  - infection status (Y; '0' for infection, '1' for no infection)

Data

```
> aids=read.table("single.txt",header=TRUE)
> aids[1:10,]
     L A Y
1  405 0 0
2  412 0 1
3  301 1 1
4  253 1 1
5  307 0 0
6  392 0 0
7  361 0 1
8  363 1 1
9  267 1 1
10 355 0 0
```


Unadjusted analysis in R

```
> chisq.test(x=aids$A,y=aids$Y)

        Pearson's Chi-squared test with Yates' continuity cor

data:  aids$A and aids$Y
X-squared = 337.47, df = 1, p-value < 2.2e-16
```

- *Interpretation?*

The role of CD4 count

- Subjects with low CD4 count are more likely to get AZT, and more likely to get infections
- Arguable, CD4 count is an important confounder that we need to adjust for
- But in the data, very few subjects have the same CD4 count
  - stratification on CD4 count is not feasible
- Let's use a regression model

The logistic regression model

- Since the outcome is binary, it is natural to use the logistic regression model

$$\text{logit}\{\Pr(Y=1|A,L)\} = \alpha + \beta A + \gamma L$$

- *What are the interpretations of α, β, and γ in terms of probabilities?*

Solution

$$\text{logit}\{\Pr(Y=1|A,L)\} = \alpha + \beta A + \gamma L$$

$$\begin{aligned}\alpha &= \text{logit}\{\Pr(Y=1|A=0,L=0)\} \\ &= \log\left\{\frac{\Pr(Y=1|A=0,L=0)}{\Pr(Y=0|A=0,L=0)}\right\}\end{aligned}$$

$$\begin{aligned}\beta &= \text{logit}\{\Pr(Y=1|A=1,L)\} - \text{logit}\{\Pr(Y=1|A=0,L)\} \\ &= \log\left\{\frac{\Pr(Y=1|A=1,L)}{\Pr(Y=0|A=1,L)}\Big/\frac{\Pr(Y=1|A=0,L)}{\Pr(Y=0|A=0,L)}\right\}\end{aligned}$$

$$\begin{aligned}\gamma &= \text{logit}\{\Pr(Y=1|A,L+1)\} - \text{logit}\{\Pr(Y=1|A,L)\} \\ &= \log\left\{\frac{\Pr(Y=1|A,L+1)}{\Pr(Y=0|A,L+1)}\Big/\frac{\Pr(Y=1|A,L)}{\Pr(Y=0|A,L)}\right\}\end{aligned}$$

Underlying assumptions

$$\text{logit}\{\Pr(Y=1|A,L)\} = \alpha + \beta A + \gamma L$$

- *What assumptions do this model make?*

Solution

$$\text{logit}\{\Pr(Y=1|A,L)\} = \alpha + \beta A + \gamma L$$

- The increase in log odds of being infection free, comparing AZT with no AZT, at a given CD4 count L, is assumed to be constant ($= \beta$) across levels of L
- The increase in log odds for being infection free, comparing CD4 count $L+1$ with CD4 count L, at a given AZT level A, is assumed to be constant ($= \gamma$) across levels of A and L

Remember

- **All models are wrong**
  - but if the model is approximately correct, then our conclusions are approximately valid
- Assumptions that we make should ideally be justified by both
  - subjects matter knowledge, and
  - data (e.g. diagnostic tests)

Fitting the model in R

```
> summary(glm(formula=Y~A+L,family=binomial,
    data=aids))

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.251557   0.707103  -3.184  0.00145 **
A            3.513298   0.240476  14.610  < 2e-16 ***
L            0.004962   0.001882   2.637  0.00836 **
```

- *Interpretation?*

Outline

A closer look at the model

- Adding an interaction term between *A* and *L* gives:

```
> summary(glm(formula=Y~A+L+A*L,family=binomial,
    data=aids))

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.987432   0.766467  -1.288 0.197645
A           -3.309358   1.734947  -1.907 0.056460 .
L            0.001564   0.002049   0.763 0.445355
A:L          0.021647   0.005667   3.820 0.000133 ***
```

- *Interpretation? Is the treatment beneficial or harmful?*

What to report?

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.987432   0.766467  -1.288 0.197645
A           -3.309358   1.734947  -1.907 0.056460 .
L            0.001564   0.002049   0.763 0.445355
A:L          0.021647   0.005667   3.820 0.000133 ***
```

- The main effect together with the interaction term?
  - unintuitive for non-statisticians
  - cumbersome if many covariates and interaction terms
- The effect at, say, the mean/median of L?
  - not very informative, unless most subjects are close to the mean/median

The marginal effect

$$\Pr(Y_1 = 1) \text{ vs } \Pr(Y_0 = 1)$$

- Arguably more intuitive than main effect + interaction term
- Can always be presented as one single number (e.g. one log odds ratio) regardless of the number of interactions
- More informative than the effect at the mean/median L, since it applies to the whole population

The standardization formula

- We have used the standardization formula

$$\Pr(Y_a = 1) = \sum_L \Pr(Y = 1|A = a, L)\Pr(L)$$

- to carry out standardization in a non-parametric fashion
- When L is high-dimensional (e.g. continuous), non-parametric standardization is not feasible
- But we can use a regression model for the outcome to estimate $\Pr(Y_a = 1)$
  - through an estimation technique called 'Maximum Likelihood' (ML)

Marginal effect through outcome model: ML

- **Step 1**: Fit a regression model for the outcome
- **Step 2**: Replace the factual exposure level with a, for each individual
- **Step 3**: Estimate $\Pr(Y = 1|A = a, L)$ for each individual (i.e. for each observed value of L)
- **Step 4**: Average these estimates over all individuals to obtain an estimate of $\Pr(Y_a = 1)$
- The estimate of $\Pr(Y_a = 1)$ is unbiased under conditional exchangeability, given L

In R

```
> #step 1
> fit=glm(formula=Y~A+L+A*L,family=binomial,
     data=aids)
> #step 2 for a=0
> aids0=aids
> aids0$A=0
> #step 3 for a=0
> pred0=predict(object=fit,newdata=aids0,
   type="respons")
> #step 4 for a=0
> p0=mean(pred0)
> p0
[1] 0.3916425
```


In R, cont'd

```
> #step 1
> fit=glm(formula=Y~A+L+A*L,family=binomial,
     data=aids)
> #step 2 for a=1
> aids1=aids
> aids1$A=1
> #step 3 for a=1
> pred1=predict(object=fit,newdata=aids1,
   type="respons")
> #step 4 for a=1
> p1=mean(pred1)
> p1
[1] 0.9625101
```


The marginal causal log odds ratio

$$\text{logit}\{\Pr(Y=1|A,L)\} = \alpha + \beta A + \gamma L + \psi AL$$

$$\hat{\Pr}(Y_1=1) = 0.9625101 \qquad \hat{\Pr}(Y_0=1) = 0.3916425$$

- We can use the estimates of $\Pr(Y_1=1)$ and $\Pr(Y_0=1)$ to construct an estimate of the marginal causal log odds ratio

$$\log\left\{\frac{\hat{\Pr}(Y_1=1)}{1-\hat{\Pr}(Y_1=1)} \Big/ \frac{\hat{\Pr}(Y_0=1)}{1-\hat{\Pr}(Y_0=1)}\right\} = 3.68$$

- *Interpretation?*

Standard errors

- Standard errors can be obtained with some additional programming
  - sandwich formula
  - bootstrap
- Bootstrap:

$$s.e = 0.23$$

- 95% CI:

$$\text{estimate} \pm 1.96 \times s.e. = 3.68 \pm 1.96 \times 0.23 = (3.23, 4.13)$$

Other measures of marginal effects

$$\text{logit}\{\Pr(Y = 1|A, L)\} = \alpha + \beta A + \gamma L + \psi AL$$

$$\hat{\Pr}(Y_1 = 1) = 0.9625101 \qquad \hat{\Pr}(Y_0 = 1) = 0.3916425$$

- Once we have estimated $\Pr(Y_1 = 1)$ and $\Pr(Y_0 = 1)$ separately, we can estimate any measure of effect, e.g.

$$\text{causal risk difference} = \hat{\Pr}(Y_1 = 1) - \hat{\Pr}(Y_0 = 1) = 0.57$$

$$\text{causal risk ratio} = \hat{\Pr}(Y_1 = 1)/\hat{\Pr}(Y_0 = 1) = 2.46$$

even though the estimates were derived from a logistic regression model

Outcome models vs exposure models

- So far, we have considered models for the outcome, e.g

$$\text{logit}\{\Pr(Y = 1|A, L)\} = \alpha + \beta A + \gamma L$$

- Sometimes, it may be more natural to use a model for the exposure, e.g.

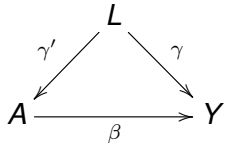
$$\text{logit}\{\Pr(A = 1|L)\} = \alpha' + \gamma' L$$

  - for instance, we may know more about the guidelines for AZT administration, than we know about the biological mechanisms underlying infection
- Marginal causal effects can be estimated with a regression model for the exposure
  - through an estimation technique called 'Inverse Probability Weighting' (IPW)

Outcome models vs exposure models in a DAG

$$\text{logit}\{\Pr(Y = 1|A, L)\} = \alpha + \beta A + \gamma L$$

$$\text{logit}\{\Pr(A = 1|L)\} = \alpha' + \gamma' L$$



Marginal effect through exposure model: IPW

- **Step 1**: Fit a regression model for the exposure
- **Step 2**: For each subject, use the fitted exposure model to estimate a subject-specific weight

$$W = 1/\hat{\Pr}(A|L)$$

  - for instance, suppose that $\hat{\Pr}(A = 1|L = 305) = 0.2$
  - subjects with $(A = 1, L = 305)$ then get the weight $1/0.2 = 5$, and subjects with $(A = 0, L = 305)$ get the weight $1/(1 - 0.2) = 1.25$
- **Step 3**: Use $\Pr(Y = 1|A = a)$ in the weighted sample as an estimate of $\Pr(Y_a = 1)$, for $a = 1$ and $a = 0$
- The estimate of $\Pr(Y_a = 1)$ is unbiased under conditional exchangeability, given L

In R

```
> #step 1
> fit=glm(formula=A~L,family=binomial,data=aids)
> #step 2
> pred=predict(object=fit,type="respons")
> w=1/(aids$A*pred+(1-aids$A)*(1-pred))
> #step 3 for a=0
> p0=weighted.mean(x=aids$Y[aids$A==0],
   w=w[aids$A==0])
> p0
[1] 0.3920298
> #step 3 for a=1
> p1=weighted.mean(x=aids$Y[aids$A==1],
   w=w[aids$A==1])
> p1
[1] 0.9636674
```


The marginal causal log odds ratio

$$\text{logit}\{\Pr(A=1|L)\} = \alpha' + \gamma' L$$

$$\hat{\Pr}(Y_1 = 1) = 0.9636674 \qquad \hat{\Pr}(Y_0 = 1) = 0.3920298$$

- We can use the estimates of $\Pr(Y_1 = 1)$ and $\Pr(Y_0 = 1)$ to construct an estimate of the marginal causal log odds ratio

$$\log\left\{\frac{\hat{\Pr}(Y_1=1)}{1-\hat{\Pr}(Y_1=1)} \Big/ \frac{\hat{\Pr}(Y_0=1)}{1-\hat{\Pr}(Y_0=1)}\right\} = 3.72$$

- *Interpretation?*

Standard errors

- Standard errors can be obtained with some additional programming
  - sandwich formula
  - bootstrap
- Bootstrap:

$$s.e = 0.23$$

- 95% CI:

$$\text{estimate} \pm 1.96 \times s.e. = 3.72 \pm 1.96 \times 0.23 = (3.27, 4.17)$$

Why IPW works

- $\Pr(Y = 1|A = a)$ is equal to $\Pr(Y_a = 1)$ if exposed and unexposed are exchangeable
- This is in fact true in the weighted sample
  - assuming that conditional exchangeability, given L, holds in the unweighted sample
- The weighting eliminates all confounding by L

A cautionary note on IPW

- IPW can give very unstable estimates (e.g. large standard errors, wide confidence intervals) for non-binary exposures
- IPW often produces considerably less stable estimates than ML

Outline

Outline

Motivating example

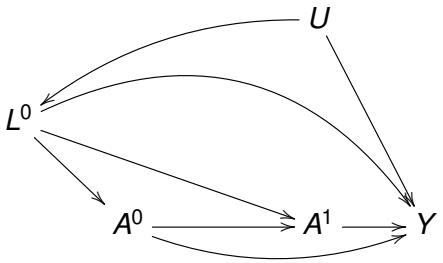
- Suppose we carry out an observational study to estimate the causal effect of AZT on infection risk for AIDS patients
- 300 subjects enrolled
- At $t = 0$ (baseline) and $t = 1$ we measure:
  - CD4 count (L^t; counts/μl)
  - AZT level (A^t; '0' for 'untreated at t', '1' for 'treated at t')
- At end of follow up we measure :
  - infection status (Y; '0' for infection, '1' for no infection)

Data

```
> aids=read.table("multiple.txt",header=TRUE)
> aids[1:10,]
    L0 A0  L1 A1 Y
1  366  0 320  0 0
2  371  0 364  0 0
3  353  0 320  0 0
4  357  0 315  0 1
5  316  1 275  1 1
6  389  0 362  1 0
7  332  0 220  0 1
8  264  1 446  1 1
9  419  0 348  0 0
10 382  0 344  0 0
```


Ignoring L^1

- Let's first ignore L^1, and assume the DAG below
- In this DAG, joint and direct effects can be estimated by standard adjustment for L^0



Standard adjustment in R

$$\text{logit}\{\Pr(Y=1|A^0,A^1,L^0)\} = \alpha + \beta_0 A^0 + \beta_1 A^1 + \gamma_0 L^0$$

```
> summary(glm(formula=Y~A0+A1+L0,family=binomial,
     data=aids))

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.702174   0.563003  -3.023 0.002500 **
A0          -0.884715   0.427294  -2.071 0.038405 *
A1           2.402227   0.407327   5.898 3.69e-09 ***
L0           0.006026   0.001563   3.854 0.000116 ***
```

- *Compute the conditional joint effect of A^0 and A^1, given L^0, as a log odds ratio*
- *Compute the conditional direct effect of A^0, at a fixed level of A^1, given L^0, as a log odds ratio*

Solution

$$\text{logit}\{\Pr(Y=1|A^0,A^1,L^0)\} = \alpha + \beta_0 A^0 + \beta_1 A^1 + \gamma_0 L^0$$

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.702174   0.563003  -3.023 0.002500 **
A0          -0.884715   0.427294  -2.071 0.038405 *
A1           2.402227   0.407327   5.898 3.69e-09 ***
L0           0.006026   0.001563   3.854 0.000116 ***
```

- The conditional joint effect of A^0 and A^1, given L^0, as a log odds ratio:

$$\begin{aligned}
&\log\left\{\frac{\Pr(Y_{11}=1|L^0)}{\Pr(Y_{11}=0|L^0)} \Big/ \frac{\Pr(Y_{00}=1|L^0)}{\Pr(Y_{00}=0|L^0)}\right\} \\
&= \text{logit}\{\Pr(Y_{11}=1|L^0)\} - \text{logit}\{\Pr(Y_{00}=1|L^0)\} \\
&= \text{logit}\{\Pr(Y=1|A^0=1,A^1=1,L^0)\} - \text{logit}\{\Pr(Y=1|A^0=0,A^1=0,L^0)\} \\
&= (\alpha+\beta_0+\beta_1+\gamma_0 L) - (\alpha+\gamma_0 L) = \beta_0+\beta_1 = -0.88+2.40 = 1.52
\end{aligned}$$

Solution, cont'd

$$\text{logit}\{\Pr(Y=1|A^0,A^1,L^0)\} = \alpha + \beta_0 A^0 + \beta_1 A^1 + \gamma_0 L^0$$

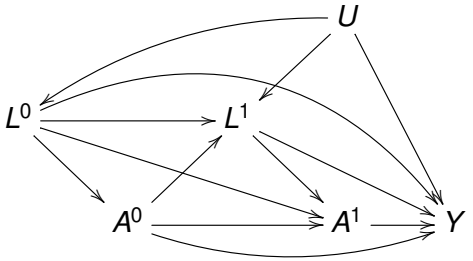
```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.702174   0.563003  -3.023 0.002500 **
A0          -0.884715   0.427294  -2.071 0.038405 *
A1           2.402227   0.407327   5.898 3.69e-09 ***
L0           0.006026   0.001563   3.854 0.000116 ***
```

- The conditional direct effect of A^0, at a fixed level of A^1, given L^0, as a log odds ratio:

$$\begin{aligned}
&\log\left\{\frac{\Pr(Y_{1a^1}=1|L^0)}{\Pr(Y_{1a^1}=0|L^0)} \Big/ \frac{\Pr(Y_{0a^1}=1|L^0)}{\Pr(Y_{0a^1}=0|L^0)}\right\} \\
&= \text{logit}\{\Pr(Y_{1a^1}=1|L^0)\} - \text{logit}\{\Pr(Y_{0a^1}=1|L^0)\} \\
&= \text{logit}\{\Pr(Y=1|A^0=1,A^1=a^1,L^0)\} - \text{logit}\{\Pr(Y=1|A^0=0,A^1=a^1,L^0)\} \\
&= (\alpha+\beta_0+\beta_1 a^1+\gamma_0 L) - (\alpha+\beta_1 a^1+\gamma_0 L) = \beta_0 = -0.88
\end{aligned}$$

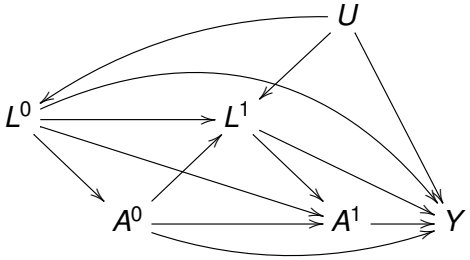
The consequence of ignoring L^1

- Ignoring L^1 makes the analysis simpler, but is likely to cause bias
- To see this, consider the more realistic DAG below, which includes L^1
- *If the DAG below is correct, what are the problems of ignoring L^1?*



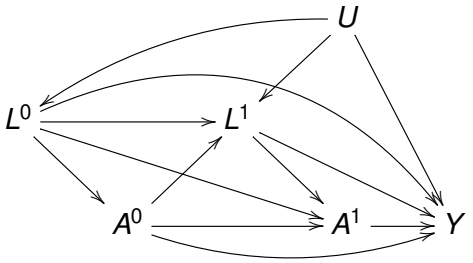
Solution

$$\text{logit}\{\Pr(Y=1|A^0,A^1,L^0)\} = \alpha + \beta_0 A^0 + \beta_1 A^1 + \gamma_0 L^0$$



- The coefficient for A^1 becomes biased due to the open non-causal path $A^1 \leftarrow L^1 \leftarrow U \rightarrow Y$
- The coefficient for A^0 becomes biased due to the open non-causal path $A^0 \rightarrow A^1 \leftarrow L^1 \leftarrow U \rightarrow Y$

Paying attention to L^1



- We proceed by paying attention to L^1, and we assume the DAG above
- In this DAG, we have sequential exchangeability
  - sequential adjustment gives the conditional causal effect of A^t, given the observed past
  - standard adjustment does not give joint and direct effects
  - sequential standardization gives marginal joint and direct effects

Sequential adjustment at $t = 0$ in R

$$\text{logit}\{\Pr(Y = 1|L^0, A^0)\} = \alpha + \underbrace{\beta_0 A^0}_{\text{causal effect}} + \underbrace{\gamma_0 L^0}_{\text{observed past}}$$

```
> summary(glm(formula=Y~A0+L0,family=binomial,
      data=aids))

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.497247   0.552822  -2.708 0.006761 **
A0           1.082838   0.241974   4.475 7.64e-06 ***
L0           0.005802   0.001538   3.773 0.000161 ***
```

- *What is the causal interpretation of β_0?*

Solution

$$\text{logit}\{\Pr(Y = 1|L^0, A^0)\} = \alpha + \underbrace{\beta_0 A^0}_{\text{causal effect}} + \underbrace{\gamma_0 L^0}_{\text{observed past}}$$

- β_0 is the conditional causal effect of A^0, given L^0, as a log odds ratio:

$$\begin{aligned}
&\log\left\{\frac{\Pr(Y_{a^0=1} = 1|L^0)}{\Pr(Y_{a^0=1} = 0|L^0)} \Big/ \frac{\Pr(Y_{a^0=0} = 1|L^0)}{\Pr(Y_{a^0=0} = 0|L^0)}\right\} \\
&= \text{logit}\{\Pr(Y_{a^0=1} = 1|L^0)\} - \text{logit}\{\Pr(Y_{a^0=0} = 1|L^0)\} \\
&= \text{logit}\{\Pr(Y = 1|L^0, A^0 = 1)\} - \text{logit}\{\Pr(Y = 1|L^0, A^0 = 0)\} \\
&(\alpha + \beta_0 + \gamma_0 L^0) - (\alpha + \gamma_0 L^0) = \beta_0
\end{aligned}$$

Sequential adjustment at $t = 1$ in R

$$\text{logit}\{\Pr(Y = 1|L^0, A^0, L^1, A^1)\} = \alpha + \underbrace{\beta_1 A^1}_{\text{causal effect}} + \underbrace{\beta_0 A^0 + \gamma_0 L^0 + \gamma_1 L^1}_{\text{observed past}}$$

```
> summary(glm(formula=Y~A1+L0+A0+L1,
      family=binomial,data=aids))

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.836716   0.760601  -5.044 4.55e-07 ***
A0          -0.993092   0.434425  -2.286   0.0223 *
A1           2.432397   0.411373   5.913 3.36e-09 ***
L0           0.006540   0.001589   4.115 3.87e-05 ***
L1           0.005895   0.001372   4.298 1.72e-05 ***
```

- *What is the causal interpretation of β_1?*

Solution

$$\text{logit}\{\Pr(Y = 1|L^0, A^0, L^1, A^1)\} = \alpha + \underbrace{\beta_1 A^1}_{\text{causal effect}} + \underbrace{\beta_0 A^0 + \gamma_0 L^0 + \gamma_1 L^1}_{\text{observed past}}$$

- β_1 is the conditional causal effect of A^1, given (L^0, A^0, L^1), as a log odds ratio:

$$\begin{aligned}
&\log\left\{\frac{\Pr(Y_{a^1=1} = 1|L^0, A^0, L^1)}{\Pr(Y_{a^0=1} = 0|L^0, A^0, L^1)} \Big/ \frac{\Pr(Y_{a^1=0} = 1|L^0, A^0, L^1)}{\Pr(Y_{a^1=0} = 0|L^0, A^0, L^1)}\right\} \\
&= \text{logit}\{\Pr(Y_{a^1=1} = 1|L^0, A^0, L^1)\} - \text{logit}\{\Pr(Y_{a^1=0} = 1|L^0, A^0, L^1)\} \\
&= \text{logit}\{\Pr(Y = 1|L^0, A^0 = 1, L^1, A^1)\} - \text{logit}\{\Pr(Y = 1|L^0, A^0 = 0, L^1, A^1)\} \\
&(\alpha + \beta_1 + \beta_0 A^0 + \gamma_0 L^0 + \gamma_1 L^1) - (\alpha + \beta_0 A^0 + \gamma_0 L^0 + \gamma_1 L^1) = \beta_1
\end{aligned}$$

Outline

Sequential standardization

- Sequential adjustment gives
  - the conditional causal effect of A^0, given L^0, and
  - the conditional causal effect of A^1, given (L^0, A^0, L^1)
- But we may want to estimate
  - the joint effect of (A^0, A^1), or
  - the direct effect of A^0, not mediated through A^1
- Sequential standardization gives joint and direct effects, marginally over (L^0, L^1)

Outcome models vs exposure models

- As for ordinary standardization, it is often desirable to use regression models for sequential standardization
- And as for ordinary standardization, we can use either outcome models or exposure models
- But unlike ordinary standardization, there are some serious disadvantages of outcome models for sequential standardization

Outcome model

- The G-formula for two time points:

$$\Pr(Y_{a^0a^1} = 1) = \sum_{L^0,L^1} \Pr(Y = 1|L^0, A^0 = a^0, L^1, A^1 = a^1)\Pr(L^0)\Pr(L^1|L^0, A^0 = a^0)$$

- We need one model for the outcome, given the whole observed past:

$$\Pr(Y = 1|L^0, A^0, L^1, A^1),$$

and one model for the covariate, at each time point, given the observed past at that time point:

$$\Pr(L^0)$$

and

$$\Pr(L^1|L^0, A^0)$$

Example

$$\text{logit}\{\Pr(Y=1|L^0,A^0,L^1,A^1)\} = \alpha + \beta_0 A^0 + \beta_1 A^1 + \gamma_0 L^0 + \gamma_1 L^1$$
$$L^0 \sim N(\mu, \sigma^2)$$
$$L^1 \mid (L^0, A^0) \sim N\left(\mu + \psi L^0 + \phi A^0, \sigma^2\right)$$

- Fit the models, plug the estimates into

$$\Pr(Y_{a^0a^1}=1) = \sum_{L^0,L^1} \left[\overbrace{\frac{e^{\alpha+\beta_0 a^0+\beta_1 a^1+\gamma_0 L^0+\gamma_1 L^1}}{1+e^{\alpha+\beta_0 a^0+\beta_1 a^1+\gamma_0 L^0+\gamma_1 L^1}}}^{\Pr(Y=1|L^0,A^0=a^0,L^1,A^1=a^1)} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-\left(L^0-\mu\right)^2}{2\sigma^2}}}_{\Pr(L^0)} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-\left\{L^1-\left(\mu+\psi L^0+\phi a^0\right)\right\}^2}{2\sigma^2}}}_{\Pr(L^1|L^0,A^0=a^0)} \right]$$

Computational difficulties

$$\Pr(Y_{a^0a^1}=1) = \sum_{L^0,L^1} \left[\overbrace{\frac{e^{\alpha+\beta_0 a^0+\beta_1 a^1+\gamma_0 L^0+\gamma_1 L^1}}{1+e^{\alpha+\beta_0 a^0+\beta_1 a^1+\gamma_0 L^0+\gamma_1 L^1}}}^{\Pr(Y=1|L^0,A^0=a^0,L^1,A^1=a^1)} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-\left(L^0-\mu\right)^2}{2\sigma^2}}}_{\Pr(L^0)} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-\left\{L^1-\left(\mu+\psi L^0+\phi a^0\right)\right\}^2}{2\sigma^2}}}_{\Pr(L^1|L^0,A^0=a^0)} \right]$$

- To calculate the right-hand side we need to solve a two-dimensional integral
  - impossible analytically
  - awkward numerically

Interpretational difficulties

$$\Pr(Y_{a^0a^1}=1) = \sum_{L^0,L^1} \left[\overbrace{\frac{e^{\alpha+\beta_0 a^0+\beta_1 a^1+\gamma_0 L^0+\gamma_1 L^1}}{1+e^{\alpha+\beta_0 a^0+\beta_1 a^1+\gamma_0 L^0+\gamma_1 L^1}}}^{\Pr(Y=1|L^0,A^0=a^0,L^1,A^1=a^1)} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-\left(L^0-\mu\right)^2}{2\sigma^2}}}_{\Pr(L^0)} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-\left\{L^1-\left(\mu+\psi L^0+\phi a^0\right)\right\}^2}{2\sigma^2}}}_{\Pr(L^1|L^0,A^0=a^0)} \right]$$

- There is no simple interpretation of the right-hand side parameters in terms of causal effects
  - e.g. not clear how a particular value of (β_0, β_1) translates into an effect of (a^0, a^1)

Marginal Structural Models

- We can bypass both problems with the outcome model approach by using a Marginal Structural Model (MSM)
- A MSM is a model for the potential outcome $Y_{a^0a^1}$, e.g.

$$\text{logit}\{\Pr(Y_{a^0a^1}=1)\} = \alpha + \beta_0 a^0 + \beta_1 a^1$$

- The model is
  - 'marginal', as it gives marginal, over (L^0, L^1), causal effects
  - 'structural', as a synonym for 'causal'
- The parameters in a MSM model have simple interpretations in terms of causal effects

Interpretation of β_0

$$\text{logit}\{\Pr(Y_{a^0 a^1} = 1)\} = \alpha + \beta_0 a^0 + \beta_1 a^1$$

$$\begin{aligned}\beta_0 &= \text{logit}\{\Pr(Y_{1a^1} = 1)\} - \text{logit}\{\Pr(Y_{0a^1} = 1)\} \\ &= \log\left\{\frac{\Pr(Y_{1a^1} = 1)}{\Pr(Y_{1a^1} = 0)} \Big/ \frac{\Pr(Y_{0a^1} = 1)}{\Pr(Y_{0a^1} = 0)}\right\}\end{aligned}$$

- β_0 is the increase in log odds of being infection free, if everybody is given AZT at $t = 0$ as compared to if nobody is given AZT at $t = 0$, when everybody is given the same level of AZT at $t = 1$
  - the direct effect of A^0, not mediated through A^1

Interpretation of β_1

$$\text{logit}\{\Pr(Y_{a^0 a^1} = 1)\} = \alpha + \beta_0 a^0 + \beta_1 a^1$$

$$\begin{aligned}\beta_1 &= \text{logit}\{\Pr(Y_{a^0 1} = 1)\} - \text{logit}\{\Pr(Y_{a^0 0} = 1)\} \\ &= \log\left\{\frac{\Pr(Y_{a^0 1} = 1)}{\Pr(Y_{a^0 1} = 0)} \Big/ \frac{\Pr(Y_{a^0 0} = 1)}{\Pr(Y_{a^0 0} = 0)}\right\}\end{aligned}$$

- β_1 is the increase in log-odds of being infection free, if everybody is given AZT at $t = 1$ as compared to if nobody is given AZT at $t = 1$, when everybody is given the same level of AZT at $t = 0$

Fitting of MSMs

- A MSM can be fitted with a regression model for the exposure, together with IPW estimation
- Under sequential exchangeability, IPW gives unbiased estimates of the model parameters

IPW

- **Step 1**: Fit a regression model for the exposure at each time point, given the observed past up to that time point, e.g.

$$\text{logit}\{\Pr(A^0 = 1|L^0)\} = \alpha' + \gamma_0' L^0$$

$$\text{logit}\{\Pr(A^1 = 1|L^0, A^0, L^1)\} = \alpha'' + \beta_0'' A^0 + \gamma_0'' L^0 + \gamma_1'' L^1$$

- **Step 2**: For each subject, use the fitted exposure model to estimate a subject-specific weight

$$W = 1/\left\{\hat{\Pr}(A^0|L^0)\hat{\Pr}(A^1|L^0, A^0, L^1)\right\}$$

- **Step 3**: Fit the MSM using **weighted** regression, as if it would have been a model for $\Pr(Y = 1|A^0, A^1)$, e.g.

$$\text{logit}\{\Pr(Y = 1|A^0, A^1)\} = \alpha + \beta_0 A^0 + \beta_1 A^1$$

In R

```
> #step 1
> fit0=glm(formula=A0~L0,family=binomial,data=aids)
> fit1=glm(formula=A1~L0+A0+L1,family=binomial,
   data=aids)
> #step 2
> pred0=predict(object=fit0,type="respons")
> pred1=predict(object=fit1,type="respons")
> w0=1/(aids$A0*pred0+(1-aids$A0)*(1-pred0))
> w1=1/(aids$A1*pred1+(1-aids$A1)*(1-pred1))
> w=w0*w1
```


In R, cont'd

```
> #step 3
> summary(glm(formula=Y~A0+A1,family=binomial,
     data=aids,weights=w))

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.69471    0.06328   10.98   <2e-16 ***
A0           0.95105    0.07835   12.14   <2e-16 ***
A1           1.40305    0.10076   13.93   <2e-16 ***
```

- *Interpretation?*

Standard errors

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.69471    0.06328   10.98   <2e-16 ***
A0           0.95105    0.07835   12.14   <2e-16 ***
A1           1.40305    0.10076   13.93   <2e-16 ***
```

- The obtained standard errors are wrong, since they assume that the weights are known and not estimated
- Correct standard errors can be obtained with some additional programming
  - sandwich formula
  - bootstrap

Why IPW works

- The unadjusted regression model

$$\text{logit}\{\Pr(Y = 1|A^0, A^1)\} = \alpha + \beta_0 A^0 + \beta_1 A^1$$

  produces the marginal causal effect of (A^0, A^1) on Y if exposed and unexposed are exchangeable
- This is in fact true in the weighted sample
  - assuming that sequential exchangeability holds in the unweighted sample
- The weighting procedure eliminates all confounding by (L^0, L^1)

Outline

Stabilized weights

- The IPW estimates are unbiased
- However, they are often highly unstable
  - in particular if the exposure is continuous
- The IPW estimates can be stabilized by using stabilized weights

$$SW = \frac{\hat{\Pr}(A^0)\hat{\Pr}(A^1|A^0)}{\hat{\Pr}(A^0|L^0)\hat{\Pr}(A^1|L^0, A^0, L^1)}$$

  - requires regression models for the numerator as well

More complex outcomes

- In real studies
  - outcomes are often measured repeatedly
  - the survival time (often censored) is often the main target of analysis
- MSMs and IPW can be used for repeated outcomes and survival outcomes as well
  - analysis and interpretation get more complex
  - beyond the scope of this course

Doubly robust estimation

- To estimate marginal effects, we can use either
  - an outcome model, or
  - an exposure model
- It is possible to combine both models into a **doubly robust** estimator
  - unbiased if either model is correct, not necessarily both
  - two chances of valid inference instead of only one
  - beyond the scope of this course

Summary

- Standard regression models are models for the outcome
- By default, outcome models give conditional causal effects
- Marginal causal effects can be obtained by standardization
  - either with an outcome model, or with an exposure model