

Computational Social Science: Sensitivity Analysis and Bounds

Your Name Here

MM/DD/YYYY

Sensitivity Analysis Rationale

So far in our explorations of observational studies we have considered various methods of accounting for measured confounders W as well as some cases in which we can account for unmeasured confounders U (e.g. if we have measured a suitable instrumental variable).

However, it is quite often the case that we are simply unable to control for unmeasured confounding in any meaningful way. Another strategy then is to quantify the *uncertainty* due to unmeasured confounding. Doing so is commonly referred to as Sensitivity Analysis. (Note: The term “Sensitivity Analysis” can be used much more broadly to refer to methods of varying *any* modeling assumptions, but here we use the more specific definition)

Recall that we already use methods of quantifying uncertainty due to random variation (e.g. standard error, confidence intervals, hypothesis testing). However, this *random* uncertainty is distinct from the *systematic* uncertainty due to unmeasured confounding.

Simulation

Let us again consider a version of our AspiTyleCedrin example. Much like in our Instrumental Variables lab, in this version both exposure to AspiTyleCedrin and the outcome of experiencing a migraine are affected by watching cable news, since AspiTyleCedrin ads are commonly shown on cable news channels, and stress from excessive cable news watching can trigger migraines. However, in this case we have not been able to measure an instrumental variable such as living near a pharmacy which sells AspiTyleCedrin.

Thus we have the following variables:

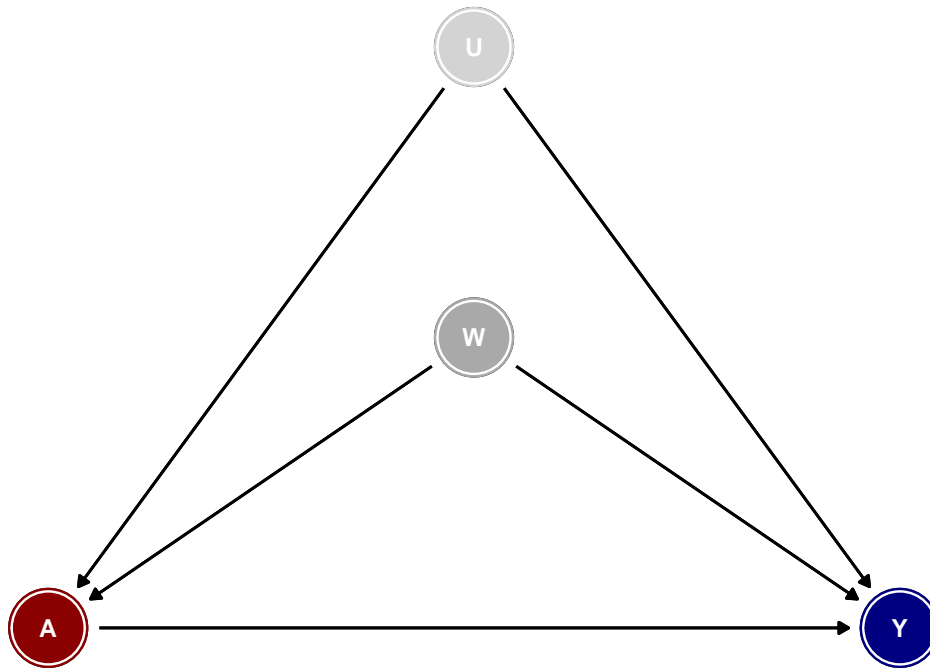
Endogenous variables:

- **A:** Treatment variable indicating whether the individual i took AspiTyleCedrin ($A_i = 1$) or not ($A_i = 0$).
- **Y:** Outcome variable indicating whether the individual experienced a migraine ($Y_{i_{obs}} = 1$) or not ($Y_{i_{obs}} = 0$).
- **W:** Variable representing sex assigned at birth, with $W = 0$ indicating AMAB (assigned male at birth), $W = 1$ indicating AFAB (assigned female at birth), and $W = 2$ indicating an X on the birth certificate, possibly representing an intersex individual or left blank.

Exogenous variables:

- **U:** Unmeasured confounding variable, cable news watching, which affects the exposure A and the outcome Y ,

And our DAG is as follows:



Simulate the dataset:

```
n = 1e4 # Number of individuals

# NOTE: Again, don't worry too much about how we're creating this dataset,
# this is just an example.

df <- data.frame(U = rbernoulli(n, p = 0.34),
                 W = sample(0:2, size = n, replace = TRUE,
                           prob = c(0.49,0.50,0.01))
)

df <- df %>%
  mutate(Y_0 = as.numeric(rbernoulli(n,
                                     p = (0.87 + 0.035*(W > 0) +
                                           0.05*(U > 0)))),
         Y_1 = as.numeric(rbernoulli(n,
                                     p = (0.34 + 0.035*(W > 0) +
                                           0.3*(U > 0)))),
         A = as.numeric(rbernoulli(n,
                                     p = (0.03 + 0.06*(W > 0) + 0.21*(U == 1))),
         ITE = Y_1 - Y_0,
         Y = as.numeric((A & Y_1) | (!A & Y_0))
)

head(df)

##      U W Y_0 Y_1 A ITE Y
## 1 FALSE 1   1   1 0  0 1
## 2 FALSE 1   1   0 0 -1 1
## 3  TRUE 0   1   0 1 -1 0
```

```
## 4 TRUE 1 0 1 0 1 0
## 5 FALSE 1 1 0 0 -1 1
## 6 FALSE 0 1 1 0 0 1
```

```
summary(df)
```

```
##      U              W              Y_0              Y_1
## Mode :logical   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## FALSE:6618     1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000
## TRUE :3382     Median :1.0000   Median :1.0000   Median :0.0000
##              Mean   :0.5202   Mean   :0.8997   Mean   :0.4643
##              3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##              Max.   :2.0000   Max.   :1.0000   Max.   :1.0000
##      A              ITE              Y
## Min.   :0.0000   Min.   :-1.0000   Min.   :0.0000
## 1st Qu.:0.0000   1st Qu.: -1.0000   1st Qu.:1.0000
## Median :0.0000   Median : 0.0000   Median :1.0000
## Mean   :0.1311   Mean   :-0.4354   Mean   :0.8517
## 3rd Qu.:0.0000   3rd Qu.: 0.0000   3rd Qu.:1.0000
## Max.   :1.0000   Max.   : 1.0000   Max.   :1.0000
```

```
ATE_true <- mean(df$ITE)
ATT_true <- mean((df %>% filter(A == 1))$ITE)

df <- df %>% select(A, W, Y)
```

Recall that the Average Treatment Effect (ATE) is the average difference in the pair of potential outcomes averaged over the entire population of interest (at a particular moment in time), or rather, it is just the average (or expected value) of the individual-level treatment effect.

$$\text{ATE} = E[Y_i(1) - Y_i(0)]$$

Question 1: Use the `group_by()` and `summarize()` functions to find the estimated average treatment effect using the following formula. Compare this result to the actual ATE (saved as `ATE_true`).

$$\hat{\text{ATE}} = E[Y_i|A_i = 1, W_i = w_i] - E[Y_i|A_i = 0, W_i = w_i]$$

```
est <- df %>% group_by(A) %>%
  summarise(E_Y = mean(Y))
```

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```
ATE_crude <- est$E_Y[2] - est$E_Y[1]
```

$$\text{ATE} = -0.4354$$

$$\hat{\text{ATE}} \approx -0.32444$$

The following code chunk uses the R package `ATE` to estimate the ATE in a similar manner, but also allows us to calculate confidence intervals and p-values for this estimate:

```
ate_out <- ATE(df$Y, df$A, (df %>% select(W)))
summary(ate_out)
```

```
## Call:
## ATE(Y = df$Y, Ti = df$A, X = (df %>% select(W)))
##
##           Estimate      StdErr 95%.Lower 95%.Upper Z.value  p.value
## E[Y(1)]  0.5691920  0.0140665  0.5416222  0.5967618  40.464 < 2.2e-16 ***
## E[Y(0)]  0.8949198  0.0032814  0.8884884  0.9013511 272.727 < 2.2e-16 ***
## ATE      -0.3257277  0.0144438 -0.3540370 -0.2974185 -22.551 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 2: Report the estimate calculated by the `ATE()` function call above, as well as the 95% confidence interval and p-value.

$$\begin{aligned}\hat{ATE} &\approx -0.3257277 \\ 95\% \text{ CI}_{\hat{ATE}} &\approx [-0.354037, -0.2974185] \\ \text{p-value} &\approx 1.2996299 \times 10^{-112}\end{aligned}$$

Question 3: What would you conclude from the information you reported in the previous question if you did not know the true ATE? From what you do know the true ATE and the DAG, do you think the confidence interval appropriately captures the level of uncertainty of that estimate? Explain.

These results show a very narrow confidence interval and a tiny p-value, which would generally lead us to be very confident in this estimate. However, we can see the true ATE value is not included in this CI, and furthermore we know from the DAG that there is unmeasured confounding U that we have not controlled for.

This is a stark reminder that our interpretation of measures such as p-values, CIs, etc. are only valid if our assumptions are correct! Here we know they are not.

The CI and p-value calculated above account only for the random uncertainty (and systematic uncertainty from measured confounding W). Now we will consider different methods of quantifying the uncertainty due to the *unmeasured confounding* U .

Manski Bounds

One method for quantifying the uncertainty due to the unmeasured confounding U is by quantifying logical bounds upon necessary parameters and propagating those bounds through to the parameter of interest.

First, let us consider the ATE in more detail. We can rewrite the formula from above as:

$$ATE = E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)] = \mu_t - \mu_c$$

where μ_t is simply the average outcome under the counterfactual in which everyone received the treatment ($A = 1$ for everyone) and μ_c is simply the average outcome under the counterfactual in which no one received the treatment ($A = 0$ for everyone). NOTE: For simplicity we will ignore W for now.

Note that by the definition of expectation we can further break down these μ values like so:

$$\begin{aligned}\mu_t &= P(A = 1) \cdot E[Y_i(1)|A = 1] + P(A = 0) \cdot E[Y_i(1)|A = 0] \\ \mu_c &= P(A = 1) \cdot E[Y_i(0)|A = 1] + P(A = 0) \cdot E[Y_i(0)|A = 0]\end{aligned}$$

Or:

$$\begin{aligned}\mu_t &= p \cdot \mu_{t,1} + (1 - p) \cdot \mu_{t,0} \\ \mu_c &= p \cdot \mu_{c,1} + (1 - p) \cdot \mu_{c,0}\end{aligned}$$

where:

- p is the probability of treatment.
- $\mu_{t,1}$ is the average outcome of treatment among those who actually receive the treatment.
- $\mu_{t,0}$ is the average outcome of treatment among those who *do not* receive the treatment.
- $\mu_{c,1}$ is the average outcome of *not receiving* treatment among those who actually receive the treatment.
- $\mu_{c,0}$ is the average outcome of *not receiving* treatment among those who *do not* receive the treatment.

In practice, we can reasonably estimate p , $\mu_{t,1}$ and $\mu_{c,0}$, but *not* $\mu_{t,0}$ and $\mu_{c,1}$.

Question 4: Explain why the previous statement is true.

We can estimate p from the proportion of treated individuals in our dataset. Furthermore, our observed data does contain the outcome of treatment among individuals receiving the treatment and the outcome of no treatment among individuals receiving the treatment, but it is impossible for us to observe what would happen to individuals under the opposite treatment condition than actually happened.

However, we do know that since the outcome Y is binary, $\mu_{t,0}$ and $\mu_{c,1}$ must by definition lie inside the interval $[0,1]$. This knowledge allows us to construct bounds on μ_t and μ_c like so:

$$\begin{aligned}\mu_t &\in [p \cdot \mu_{t,1}, p \cdot \mu_{t,1} + (1 - p)] \\ \mu_c &\in [(1 - p) \cdot \mu_{c,0}, (1 - p) \cdot \mu_{c,0} + p]\end{aligned}$$

Question 5: Show how the bounds for μ_t and μ_c above follow from bounds of $[0,1]$ on $\mu_{t,0}$ and $\mu_{c,1}$.

$$\begin{aligned}
LL_{\mu_t} &= \mu_t, [\mu_{t,0} = 0] \\
&= p \cdot \mu_{t,1} + (1 - p) \cdot 0 \\
&= p \cdot \mu_{t,1} \\
UL_{\mu_t} &= \mu_t, [\mu_{t,0} = 1] \\
&= p \cdot \mu_{t,1} + (1 - p) \cdot 1 \\
&= p \cdot \mu_{t,1} + (1 - p) \\
LL_{\mu_c} &= \mu_c, [\mu_{c,1} = 0] \\
&= p \cdot 0 + (1 - p) \cdot \mu_{c,0} \\
&= (1 - p) \cdot \mu_{c,0} \\
UL_{\mu_c} &= \mu_c, [\mu_{c,1} = 1] \\
&= p \cdot 1 + (1 - p) \cdot \mu_{c,0} \\
&= p + (1 - p) \cdot \mu_{c,0}
\end{aligned}$$

We can then use these bounds of μ_t and μ_c in the formula for the ATE to get bounds for the ATE itself.

$$\begin{aligned}
ATE &= \mu_t - \mu_c \\
LL_{ATE} &= LL_{\mu_t} - UL_{\mu_c} \\
&= [p \cdot \mu_{t,1}] - [p + (1 - p) \cdot \mu_{c,0}] \\
UL_{ATE} &= UL_{\mu_t} - LL_{\mu_c} \\
&= [p \cdot \mu_{t,1} + (1 - p)] - [(1 - p) \cdot \mu_{c,0}]
\end{aligned}$$

Question 6: Why are the bounds for the ATE calculated in this way? For example, why not find the difference of both lower bounds and both upper bounds, respectively?

The ATE is the *difference* in the average outcome among the treatment and control, so this difference will be smallest when the average outcome of the treatment is high and average outcome of the control is low (because they will be closer to each other), and vice versa for the upper bound.

We now have a formula for the ATE using only p , $\mu_{t,1}$ and $\mu_{c,0}$, which we said earlier can be reasonably estimated from our data. Thus we can use plug-in estimators for these values to estimate the bounds for the ATE, like so:

$$\begin{aligned}
\hat{LL}_{ATE} &= [\hat{p} \cdot \hat{\mu}_{t,1}] - [\hat{p} + (1 - \hat{p}) \cdot \hat{\mu}_{c,0}] \\
\hat{UL}_{ATE} &= [\hat{p} \cdot \hat{\mu}_{t,1} + (1 - \hat{p})] - [(1 - \hat{p}) \cdot \hat{\mu}_{c,0}]
\end{aligned}$$

Question 7: Calculate these bounds on the ATE using the data.

```

p_hat <- nrow(df %>% filter(A == 1)) / nrow(df)
mu.t1_hat <- df %>% filter(A == 1) %>% summarise(mean = mean(Y)) %>% pull(mean)
mu.c0_hat <- df %>% filter(A == 0) %>% summarise(mean = mean(Y)) %>% pull(mean)

LL_ATE <- p_hat*mu.t1_hat - (p_hat + (1 - p_hat)*mu.c0_hat)
UL_ATE <- p_hat*mu.t1_hat + (1 - p_hat) - (1 - p_hat)*mu.c0_hat

```

$$ATE \in [-0.8334, 0.1666]$$

Note that by definition this interval must contain zero and must have a width of one. Do not confuse this with a confidence interval! Also note that this does not include the random variation associated with our estimates.

Question 8: Compare this interval to the true ATE.

The interval does indeed include the true ATE value. What's more, while the interval crosses zero, it clearly has more negative range than positive, which is encouraging considering the true value is in fact negative.

Rosenbaum Sensitivity Analysis

While Manski Bounds are useful to estimate a “worst case” range of possible values, since zero is necessarily included they are not especially informative.

It is useful, then, to be able to use outside knowledge or other reasonable assumptions about possible ranges of values to possibly restrict this interval.

The Rosenbaum-Rubin approach incorporates a bit more complexity than do Manski bounds. First, we imagine the unmeasured confounding U as binary with some probability q , where:

$$q = P(U_i = 1) = 1 - P(U_i = 0)$$

We then specify models for the relationships between the unmeasured confounding U and the other variables: treatment assignment A , outcome under treatment $Y(1)$, and outcome under no treatment $Y(0)$. Note that we are still ignoring the measured confounder W for now. Since A , $Y(1)$, and $Y(0)$ are all binary, we can specify logistic models for these relationships, like so:

$$\begin{aligned} \text{logit}[P(A_i = 1|U_i = u)] &= \gamma_0 + \gamma_1 \cdot u \\ \text{logit}[P(Y_i(1) = 1|U_i = u)] &= \alpha_0 + \alpha_1 \cdot u \\ \text{logit}[P(Y_i(0) = 1|U_i = u)] &= \beta_0 + \beta_1 \cdot u \end{aligned}$$

Or:

$$\begin{aligned}
P(A_i = 1|U_i = u) &= \text{logit}^{-1}[\gamma_0 + \gamma_1 \cdot u] \\
&= \frac{e^{\gamma_0 + \gamma_1 \cdot u}}{1 + e^{\gamma_0 + \gamma_1 \cdot u}} \\
P(Y_i(1) = 1|U_i = u) &= \text{logit}^{-1}[\alpha_0 + \alpha_1 \cdot u] \\
&= \frac{e^{\alpha_0 + \alpha_1 \cdot u}}{1 + e^{\alpha_0 + \alpha_1 \cdot u}} \\
P(Y_i(0) = 1|U_i = u) &= \text{logit}^{-1}[\beta_0 + \beta_1 \cdot u] \\
&= \frac{e^{\beta_0 + \beta_1 \cdot u}}{1 + e^{\beta_0 + \beta_1 \cdot u}}
\end{aligned}$$

Now we recall that we can write the ATE as:

$$\begin{aligned}
ATE &= \mu_t - \mu_c \\
&= [p \cdot \mu_{t,1} + (1 - p) \cdot \mu_{t,0}] - [p \cdot \mu_{c,1} + (1 - p) \cdot \mu_{c,0}] \\
&= p \cdot (\mu_{t,1} - \mu_{c,1}) + (1 - p) \cdot (\mu_{t,0} - \mu_{c,0})
\end{aligned}$$

Following the derivations in Chapter 22 of Imbens/Rubin, we can write each of these parameters (p , $\mu_{t,1}$, $\mu_{t,0}$, $\mu_{c,1}$, $\mu_{c,0}$) in terms of the parameters seen above (q , γ_0 , γ_1 , α_0 , α_1 , β_0 , β_1).

Chapter 22.4 details the translations of our reasonably estimable values p , $\mu_{t,1}$, and $\mu_{c,0}$, and shows that these relationships can allow us to find estimate values for γ_0 , α_0 , and β_0 .

If we then postulate values for q , γ_1 , α_1 , and β_1 , we can then estimate values for $\mu_{t,0}$ and $\mu_{c,1}$. Our sensitivity analysis then comes down to the decisions we make for postulating q , γ_1 , α_1 , and β_1 .

For example, if we fix $q = p$ and let $\gamma_1 \rightarrow \infty$, $\alpha_1 \rightarrow \infty$, and $\beta_1 \rightarrow \infty$, we find:

$$ATE = p \cdot \mu_{t,1} - p - (1 - p) \cdot \mu_{c,0}$$

which is equivalent to the lower limit derived from the Manski bounds! Furthermore, if we fix $q = p$ and let $\gamma_1 \rightarrow \infty$, $\alpha_1 \rightarrow -\infty$, and $\beta_1 \rightarrow \infty$, we find:

$$ATE = p \cdot \mu_{t,1} + (1 - p) - (1 - p) \cdot \mu_{c,0}$$

which is equivalent to the upper limit derived from the Manski bounds! Therefore we can see Manski bounds as simply a special case of this type of sensitivity analysis.

References

- Imbens/Rubin Chapter 22
- Gustafson P, McCandless LC. Probabilistic approaches to better quantifying the results of epidemiologic studies. Int J Environ Res Public Health. 2010 Apr;7(4):1520-39. doi: 10.3390/ijerph7041520. Epub 2010 Apr 1. PMID: 20617044; PMCID: PMC2872335.
- Bradshaw, Patrick. Public Health 250C slides Weeks 10 & 11