TECHNICAL REPORT
R-492
December 2019

Generalizing Experimental Results by Leveraging Knowledge of Mechanisms

Carlos Cinelli and Judea Pearl*
University of California, Los Angeles
Departments of Statistics and Computer Science
carloscinelli@ucla.edu, judea@cs.ucla.edu

December 10, 2019

Abstract

We show how experimental results can be generalized across diverse populations by leveraging knowledge of mechanisms that produce the outcome of interest. We use Structural Causal Models (SCM) and a refined version of selection diagrams to represent such knowledge, and to decide whether it entails conditions that enable generalizations. We further provide bounds for the target effect when some of these conditions are violated. We conclude by demonstrating that the structural account offers a more reliable way of analyzing generalization than positing counterfactual consequences of the actual mechanisms.

1 Introduction

Generalizing results of randomized control trials (RCT) is critical in many empirical sciences and demands an understanding of the conditions under which such generalizations are feasible. When the mechanisms that determine the outcome differ between the study population and the target

*This research was supported in parts by grants from Defense Advanced Research Projects Agency [#W911NF-16-057], National Science Foundation [#IIS-1302448, #IIS-1527490, and #IIS-1704932], and Office of Naval Research [#N00014-17-S-B001].

population, generalization requires measuring the variables responsible for such differences or, if this is not possible, isolating them away by measuring other variables (Pearl and Bareinboim, 2014). Recent work (Huitfeldt et al., 2018, 2019; Huitfeldt, 2019) describes an interesting situation under which transportability across populations is feasible without such measurements. This feasibility, however, is not immediately inferable using a standard (non-parametric) selection diagram (Pearl and Bareinboim, 2014; Bareinboim and Pearl, 2016), because it relies on the invariance of only some components of the outcome mechanism, but not all.

In this paper, we use the theory of Structural Causal Models (SCM) (Pearl, 2009) to show how generalization in these settings can be modeled using ordinary structural equations, counterfactual logic and selection diagrams. We demonstrate that it requires two key assumptions: (i) the independence of causal factors that affect the outcome; and, (ii) *functional constraints* on how these factors interact to produce the outcome. We further extend the results in Huitfeldt et al. (2018) by relaxing the monotonicity assumption and providing bounds for the causal effect in the target domain. We conclude by demonstrating that the structural account offers a general and reliable scientific basis for analyzing the transportability of causal effects across populations.

2 Motivating example

To fix ideas, we borrow the "Russian Roulette" example from Huitfeldt (2019). Although stylized, this intuitive example illustrates the key features of the problem.

A Russian Roulette trial

Suppose the city of Los Angeles decides to run a randomized control trial (RCT) to assess the effect of playing "Russian Roulette" on mortality.[1] After running the experiment, the mayor of Los Angeles discovers that "Russian Roulette" is harmful: among those assigned to play Russian Roulette, 17.5% of the people died, as compared to only 1% among those who were not

[1]Russian Roulette consists of loading a bullet into a revolver, spinning the cylinder, pointing the gun at one's own head and then pulling the trigger. We do not recommend attempting this.

assigned to play the game (people can die due to other causes during the trial, for example, prior poor health conditions).

After hearing the news about the Los Angeles experiment, the mayor of New York City (a dictator) wonders what the overall mortality rate would be if the city forced everyone to play Russian Roulette. Currently, the practice of Russian Roulette is forbidden in New York, and its mortality rate is at 5% (4% higher than LA). The mayor thus asks the city's statistician for decide *whether* and *how* one could use the data from from Los Angeles to predict the mortality rate in New York, once the new policy is implemented.

Intuitively, it is clear that some form of prediction should be feasible. Mortality is a consequence of two "independent" processes (the game of Russian Roulette and prior health conditions of the individual), and while the first factor remains unaltered across cities, the second intensifies by a known amount (5% vs 1%). Moreover, we can safely assume that the two processes interact disjunctively, namely, that death occurs if and only if one the two processes takes effect. From these two assumptions and elementary probability theory, we can conclude that mortality in NYC would be 20.8% $(1/6 + 5\% - (1/6) \times 5\%)$. In section 3 we will use a structural model to formalize the notion of "independence" and, accordingly, how the data from NYC and LA should be combined and match our expectation. But before that, let us examine how this intuition clashes with the conclusion of a coarse analysis using selection diagrams.

An "impossibility" result

Selection diagrams are causal diagrams enriched with "selection nodes" S, usually represented by square nodes (■). These new nodes are used by the analyst to indicate which *local mechanisms* are suspected to differ between two environments (in our example, the mortality mechanism is suspected to differ between Los Angeles and New York). More importantly, the absence of a selection node pointing to a variable represents the *assumption* that the local mechanism responsible for assigning the value to that variable is the same in the two populations (Pearl, 1995, 2009; Pearl and Bareinboim, 2014; Bareinboim and Pearl, 2016).

To build our selection diagram, we need to introduce some notation. The population of Los Angeles will be denoted by Π and that of New York by Π^*. The random variable Y stands for mortality, with events $Y = 1$ denoting "death" and $Y = 0$ denoting "survival"; the random variable X stands for

the "treatment" assignment, with events $X = 1$ denoting "play Russian Roulette" and $X = 0$ denoting "not play Russian Roulette". The random variable Y_x denotes the potential response of Y when the treatment X is experimentally set to x. Thus, mathematically, the findings of the RCT can be translated to $P(Y_1 = 1) = 17.5\%$ and $P(Y_0 = 1) = 1\%$, and the available data from New York is $P^*(Y_0 = 1) = 5\%$. Our task is to estimate $P^*(Y_1 = 1)$.

(a) Coarse causal diagram (b) Coarse selection diagram

Figure 1: Coarse causal (a) and selection (b) diagrams of the Russian Roulette trial. The presence of $S \to Y$ in (b) correctly prohibits the naive transportation of the interventional distribution $P(Y_x)$ from the source Π (Los Angeles) to the target environment Π^* (New York).

The coarsest causal diagram of the Russian Roulette trial comprises only the treatment X and the outcome Y, as shown in Figure 1a. To move from the causal diagram to the selection diagram, we need to think of what may differ between LA and NYC. Since we already know from the data that $P(Y_0 = 1) \neq P^*(Y_0 = 1)$, we suspect there are differences in the way mortality is determined in the two cities (for example, people in New York may be in poorer health conditions, or the air quality may be worse). Thus, the selection diagram must contain a selection node S pointing to the mortality variable Y to indicate this disparity, as shown in Figure 1b.

Graphically, checking whether a causal relationship is transportable from one environment to another involves checking for d-separation (Pearl, 2009) of the source of disparity (the selection node S) from our target quantity. The presence of the selection node pointing directly into Y prevents the separation of S from Y, and leads us to conclude that transportability is impossible without further assumptions. On the other hand, the intuition that led us to predict the new mortality rate in NYC tells us that such assumptions, once formalized, could license transportability. This intuition, as we discussed, was based on two assumptions that are not shown in the coarse selection diagram of Figure 1, and, consequently, it fails to reveal a way of transporting the results of the trial. The diagram represents only the

existence of a disparity between LA and NYC, not its nature, and *correctly* warns us that, absent further assumptions, we are not authorized to make any generalization.

3 Building the structural model

We now explicate formally what we know about the game of "Russian Roulette" and health factors, and show how this knowledge renders transportability possible.

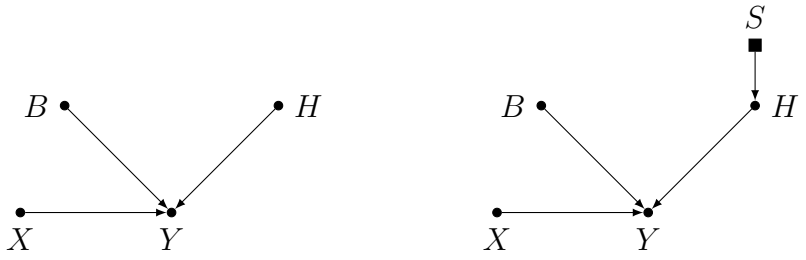
Prior health conditions *versus* physical mechanism

People can die during the trial from two different causes—either due to the "treatment" itself (if the participant is unlucky), or due to other "natural causes," such as prior poor health conditions. Therefore, we start the refinement of our model by defining two extra random variables, B and H: (i) B denotes "bad luck" when playing Russian Roulette, and represents a match ($B = 1$) or mismatch ($B = 0$) between the trigger and the location of the bullet in the cylinder; (ii) and H denotes *all* other health factors producing death ($H = 1$) or survival ($H = 0$). Accordingly, our causal diagram will contain two new edges, $H \to Y$ and $B \to Y$, since both "health conditions" and "bad luck" are key determinants of mortality Y. The updated causal diagram is shown Figure 2a. Note the absence of a directed or bidirected edge between H and B, which encodes our assumption that these two mechanisms are activated independently of each other.[2]

The new model helps us see more clearly the commonalities and disparities between LA and NYC. First, since there is a multitude of factors that can affect prior health conditions, and those are likely to differ between the two cities (as suggested by the observed difference $P(Y_0 = 1) \neq P^*(Y_0 = 1)$), we again introduce a selection node pointing to H. Moreover, to encode the assumption that the probability of "bad luck" occurring is the same in both cities, we do not connect B to a selection node.[3] The new selection diagram

[2]The arrow $X \to Y$ comprises, of course, many intermediate mechanisms (such as loading the gun, spinning the cylinder, pulling the trigger) that are not modeled explicitly.

[3]Note that, although reasonable, one cannot take this assumption for granted—it could be the case that revolvers used for Russian Roulette in New York have a different number of chambers than those used in Los Angeles. *The absence of a selection node* pointing to



(a) Causal diagram (b) Selection diagram

Figure 2: New causal (a) and selection (b) diagrams explicitly including the variables "health conditions" (H) and "bad luck" (B) when playing Russian Roulette. Here the analyst asserts (using the selection node S) that H may differ between LA and NYC, but assumes that the mechanism triggering B is the same between the two cities. Also important is the absence of a directed edge or a bidirected edge between H and B.

is shown in Figure 2b.

The diagram of Figure 2b now guides us toward leveraging the data obtained in LA to make predictions in NYC. If we can find a way to *block the source of disparity originating from* H, we would be left with the invariant physical mechanism shared by both cities. However, since H is unobserved, blockage is impossible without further assumptions. We now ask ourselves whether such assumptions, in the form of *functional constraints on the structural equation of* Y, are implied by our understanding of how the mechanisms operate.

Leveraging functional constraints

Our understanding that mortality is caused by *either one* of the two processes (prior health conditions or bad luck in the game), dictates the following *functional specification* of Y,

$$Y = H \vee (X \wedge B) \tag{1}$$

Where $\vee$ denotes the logical "or" operator, and $\wedge$ denotes the logical "and" operator. Like any structural equation, Equation 1 defines the potential outcomes Y_0 and Y_1 (Pearl, 2009, Ch.7) which we may now find useful to

B encodes the *assumption* that this is not the case.

encode explicitly. Its first implication is that $Y_0 = H$ and $Y_1 = H \vee B = Y_0 \vee B$. This tells us that, once we know the potential response of units under no treatment (Y_0) we do not need to know anything else about their previous health condition (H) to determine the value of Y_1—B would suffice.[4] We can represent this fact in a modified selection diagram, in which the potential outcomes are now also shown explicitly (Figure 3). The diagram reveals that Y_0 blocks the source of health disparities between the two populations, and we conclude therefore that $P(Y_1|Y_0) = P^*(Y_1|Y_0)$, an important feature of invariance.

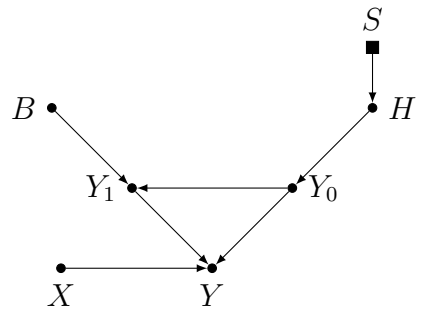


Figure 3: Selection diagram explicitly showing the potential outcomes Y_0 and Y_1 as implied by the functional constraints. Note that $Y_1 \perp\!\!\!\perp S \mid Y_0$.

A second implication of Equation 1 is that the treatment effect is *monotonic*, that is $Y_1 \geq Y_0$ for all individuals. This, in turn, implies $P(Y_1 = 1|Y_0 = 1) = 1$, which Huitfeldt et al. (2018) showed to be sufficient for identifying $P(Y_1 = 1 \mid Y_0 = 0)$. Indeed, by the law of total probability,

$$P(Y_1 = 1) = P(Y_1 = 1 \mid Y_0 = 1)P(Y_0 = 1) + P(Y_1 = 1 \mid Y_0 = 0)P(Y_0 = 0)$$

The quantities $P(Y_0 = 0)$ and $P(Y_0 = 1)$ are given from the RCT (1% and 99% respectively) and, due to monotonicity, $P(Y_1 = 1 \mid Y_0 = 1) = 1$. Thus, we have:

$$P(Y_1 = 1 \mid Y_0 = 0) = \frac{P(Y_1 = 1) - P(Y_0 = 1)}{P(Y_0 = 0)} = \frac{17.5\% - 1\%}{99\%} = 1/6$$

[4]Here $Y_0 = H$ for simplicity, but this need not be the case. The same argument would hold, for instance, if we define H to be a random variable with arbitrary cardinality and $Y = g(H) \vee (X \wedge B)$, where $g(H) \in \{0, 1\}$.

That is, the counterfactual quantity $P(Y_1 = 1 \mid Y_0 = 0)$ (the share of people who would die if forced to play Russian Roulette, among those who would not have died if not forced to do so) not surprisingly, equals 1/6 (the probability of having "bad luck" in the game of Russian Roulette using a revolver with 6 chambers)—a somewhat convoluted way of expressing our physical understanding of the problem.

So far we have established that $P(Y_1|Y_0) = P^*(Y_1|Y_0)$, and that $P(Y_1 = 1|Y_0 = 1) = 1$ and $P(Y_1 = 1 \mid Y_0 = 0) = 1/6$. Combining these results with the data from NYC under the "no-Russian-Roulette" regime (that is, $P^*(Y_0 = 1) = 5\%$ and $P^*(Y_0 = 0) = 95\%$) we can finally evaluate our target quantity $P^*(Y_1 = 1)$,

$$\begin{aligned} P^*(Y_1 = 1) &= P^*(Y_1 = 1 \mid Y_0 = 1)P^*(Y_0 = 1) + P^*(Y_1 = 1 \mid Y_0 = 0)P^*(Y_0 = 0) \\ &= P(Y_1 = 1 \mid Y_0 = 0)(5\%) + P(Y_1 = 1 \mid Y_0 = 0)(95\%) \\ &= (1)(5\%) + (1/6)(95\%) = 20.83\% \end{aligned}$$

Which matches the intuitive answer obtained in Section 2.

Relaxing monotonicity

A key step for obtaining a point estimate for $P^*(Y_1 = 1)$ was the monotonicity property, which emanates from the functional form of Equation 1. This property holds trivially in our example of the Russian Roulette, when Y represents death, but it may not hold for other outcomes or, more generally, it may not hold in contexts beyond our stylized example. Remarkably, however, and contrary to the cautionary note found in Huitfeldt et al. (2018), even in the absence of monotonicity, one can still predict the transported causal effect, albeit in the form of a *bound*. The next theorem shows that the counterfactual independence $Y_1 \perp\!\!\!\perp S \mid Y_0$ by itself is strong enough for bounding the causal effect in the target domain.

THEOREM 1. *Let* $P(Y_i = j) := p_{ij}$ *and* $P^*(Y_i = j) := p^*_{ij}$. *If* $Y_1 \perp\!\!\!\perp S \mid Y_0$, *then* p^*_{11} *is bounded by,*

$$p^*_{11} \geq \min\left\{\left(1 - \frac{p^*_{01}}{p_{01}}\right) b_L + \left(\frac{p^*_{01}}{p_{01}}\right) p_{11}, \left(1 - \frac{p^*_{01}}{p_{01}}\right) b_U + \left(\frac{p^*_{01}}{p_{01}}\right) p_{11}\right\}$$

and,

$$p^*_{11} \leq \max\left\{\left(1-\frac{p^*_{01}}{p_{01}}\right)b_L + \left(\frac{p^*_{01}}{p_{01}}\right)p_{11}, \left(1-\frac{p^*_{01}}{p_{01}}\right)b_U + \left(\frac{p^*_{01}}{p_{01}}\right)p_{11}\right\}$$

Where, $b_L = \max\left\{0, \frac{p_{11}-p_{01}}{p_{00}}\right\}$ *and* $b_U = \min\left\{\frac{p_{11}}{p_{00}}, 1\right\}$.

Proof. The bounds are obtained by solving a linear optimization problem, as detailed in Appendix A. □

For example, Theorem 1 assures us that, regardless of whether monotonicity holds, $P^*(Y_1 = 1)$ must lie between,

$$16.8\% \leq P^*(Y_1 = 1) \leq 20.8\%$$

In other words, the results of the trial in LA tells us that implementing the policy in NYC would cause *at least* an increase of $16.8\% - 5\% = 11.8\%$ and *at most* an increase of $20.8\% - 5\% = 15.8\%$ in mortality. In many practical settings, bounds as informative as those are enough for policy decisions.

4 Discussion

Having worked through our stylized example, we can now discuss its lessons to real world settings, where functional constraints are quite common. Suppose a new treatment (e.g., a drug) is under development, which is currently unavailable to the general public, and for which we have experimental data determining its effect on mortality for a specific study population. Our task is to predict the causal effect of the treatment in the general population, *knowing* that there are structural disparities between the two. For example, we see that mortality among the placebo-taking subjects differs substantially from mortality in the general population.

The non-parametric solution to this problem requires "blocking" the variables responsible for the disparity. The first lesson that emerges from our example is that certain functional constraints may allow such blocking, albeit not by direct measurements. This may occur when the outcome is a product of independent causes, only some of which are carriers of disparities, and when the outcome produced under the "no-treatment" condition is sufficient to block these sources of disparity.

While one may summarize the main "identification assumption" in terms of the counterfactual independence $Y_1 \perp\!\!\!\perp S \mid Y_0$, we did not commence the analysis by positing this or any "identification assumption". Instead, we explicated our understanding of the problem directly in a structural model, and the necessary counterfactual independence emerged naturally as a *logical consequence of the structure*. If some of those modeling assumptions happen to be challenged (say unobserved confounding between H and B) we should refrain from positing that $Y_1 \perp\!\!\!\perp S \mid Y_0$, and the diagram would vividly warn us against such assumptions.[5]

Thus, the second lesson that emerges from our example is that judging whether the counterfactual independence $Y_1 \perp\!\!\!\perp S \mid Y_0$ is plausible in a specific application requires scientific knowledge (e.g., biological, physical, etc), and the tools of structural modeling facilitate examining both the plausibility of the assumed knowledge and its logical ramifications. To make the point clearer, we now apply the modeling strategy of the last section in an example described in Huitfeldt et al. (2018, page 11):

> Consider a team of investigators who are interested in the effect of antibiotic treatment on mortality in patients with a specific bacterial infection (...) the investigators believe that the response to this antibiotic is completely determined by an unmeasured bacterial gene, such that only those who are infected with a bacterial strain with this gene respond to treatment. The prevalence of this bacterial gene is equal between populations, because the populations share the same bacterial ecosystem (...) if the investigators further believe that the gene for susceptibility reduces the mortality in the presence of antibiotics, but has no effect in the absence of antibiotics, they will conclude that G may be equal between populations.

Here the conclusion that G may be equal between populations is equivalent to claiming $Y_1 \perp\!\!\!\perp S \mid Y_0$. But is the description above sufficient for substantiating this claim? Figure 4 shows two models compatible with the description, yet leading to two opposite conclusions.

Let the variable A represent the binary treatment (antibiotic), Y represent the binary outcome (mortality), BG stand for the presence or absence

[5] In those cases a sensitivity analysis might still be possible, and one could investigate how big a departure from the original model would be necessary to invalidate the main conclusions (see, e.g., Cinelli et al. (2019); Cinelli and Hazlett (2020)).

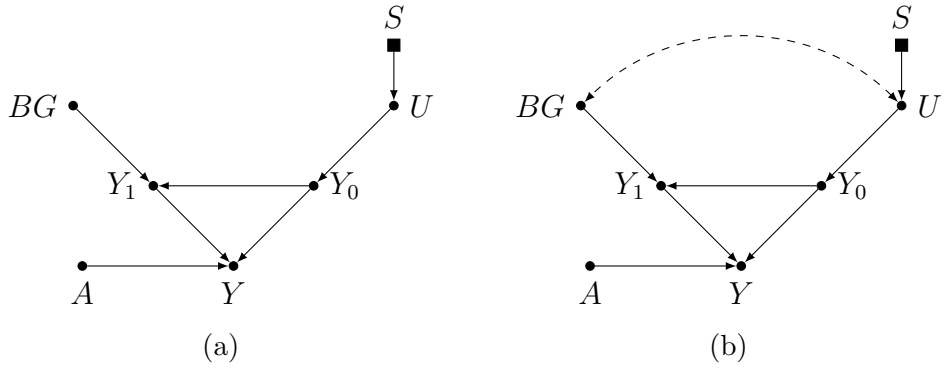


Figure 4: Two selection diagrams compatible with the verbal description of Huitfeldt et al. (2018, page 11). Yet, model (a) implies $Y_1 \perp\!\!\!\perp S \mid Y_0$, and model (b) implies the opposite; conditioning on Y_0 opens the colliding path $S \to U \leftrightarrow BG \to Y_1$.

of the "bacterial gene" and finally let U be a binary variable that summarizes all other factors that may cause death ($Y = 1$). The description of the problem suggests the functional specification,

$$Y = U \wedge (\neg A \vee \neg BG) \tag{2}$$

showing the antibiotics and the bacterial gene both helping to *reduce* mortality ($\neg$ denotes the logical "not"). Equation 2 entails the potential outcomes $Y_0 = U$ and $Y_1 = U \wedge (\neg BG) = Y_0 \wedge (\neg BG)$, which are explicitly shown in both diagrams as dictated by the functional specification. Moreover, in both models the prevalence of the bacterial gene BG is equal between populations (i.e., $BG \perp\!\!\!\perp S$). In the model of Figure 4a, as in our previous analysis, we indeed conclude that $Y_1 \perp\!\!\!\perp S \mid Y_0$, and that $P^*(Y_1)$ is transportable. However, in the model of Figure 4b, there is an unmeasured confounder between BG and U. Conditioning on Y_0 (a child of a collider) opens the colliding path $S \to U \leftrightarrow BG \to Y_1$, thus not licensing the independence $Y_1 \perp\!\!\!\perp S \mid Y_0$.

Surely, one could supplement the verbal description of the problem with other counterfactual assumptions such as $BG \perp\!\!\!\perp S \mid Y_0$. While this would rule out the scenario in Figure 4b, it raises the question of opaqueness; how can we judge whether such counterfactual statement is plausible? The answer is, again, to derive it from scientific (structural) knowledge. Attempting to skip the structural model, by articulating assumptions directly in terms of a counterfactual independence, leaves the researcher vulnerable to unnecessary

judgmental errors that can be avoided with simple and reliable graphical operations. In sum, when leveraging functional constraints, the basic tools and principles that guide non-parametric identification remain applicable, perhaps even indispensable.

References

Bareinboim, E. and Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352.

Cinelli, C. and Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Cinelli, C., Kumor, D., Chen, B., Pearl, J., and Bareinboim, E. (2019). Sensitivity analysis of linear structural causal models. *International Conference on Machine Learning*.

Huitfeldt, A. (2019). Effect heterogeneity and external validity in medicine. Available in: https://www.lesswrong.com/posts/wwbrvumMWhDfeo652/.

Huitfeldt, A., Goldstein, A., and Swanson, S. A. (2018). The choice of effect measure for binary outcomes: Introducing counterfactual outcome state transition parameters. *Epidemiologic methods*, 7(1).

Huitfeldt, A., Swanson, S. A., Stensrud, M. J., and Suzuki, E. (2019). Effect heterogeneity and variable selection for standardizing causal effects to a target population. *European journal of epidemiology*, pages 1–11.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.

Pearl, J. (2009). *Causality*. Cambridge university press.

Pearl, J. and Bareinboim, E. (2014). External validity: From do-calculus to transportability across populations. *Statistical Science*, 29(4):579–595.

A Appendix

Here we show how to obtain the bounds of Theorem 1. To simplify notation, let $P(Y_i = j) := p_{ij}$, $P^*(Y_i = j) := p^*_{ij}$, $P^*(Y_1 = 1|Y_0 = 1) = P(Y_1 = 1|Y_0 = 1) = a$ and $P^*(Y_1 = 1|Y_0 = 0) = P(Y_1 = 1|Y_0 = 0) = b$. The target function to be optimized is p^*_{11}, which can be written as,

$$p^*_{11} = ap^*_{01} + b(1 - p^*_{01}) \tag{3}$$

Our goal is to pick a and b such that it maximizes (or minimizes) Eq. 3 subject to the following constraints: (i) a and b need to be between zero and one (since a and b need to be valid probabilities); and, (ii) a and b must conform to the observed results of the trial in the source domain, that is, $p_{11} = ap_{01} + b(1 - p_{01})$. Thus, our optimization problem is,

$$\max_{a,b} \; ap^*_{01} + b(1 - p^*_{01}) \quad \text{s.t.} \quad ap_{01} + b(1 - p_{01}) = p_{11},\; 0 \leq a \leq 1,\; 0 \leq b \leq 1$$

To simplify the problem, we can use the equality constraint $p_{11} = ap_{01} + b(1 - p_{01})$ to eliminate one of the variables. For instance, writing a in terms of b gives us,

$$a = \frac{p_{11} - b(1 - p_{01})}{p_{01}} \tag{4}$$

Which results in a new target function,

$$ap^*_{01} + b(1 - p^*_{01}) = \left(\frac{p_{11} - b(1 - p_{01})}{p_{01}}\right) p^*_{01} + b(1 - p^*_{01}) \tag{5}$$

$$= \left(1 - \frac{p^*_{01}}{p_{01}}\right) b + \left(\frac{p^*_{01}}{p_{01}}\right) p_{11} \tag{6}$$

Since $0 \leq a \leq 1$, the substitution also results in additional constraints on b,

$$\frac{p_{11} - p_{01}}{p_{00}} \leq b \leq \frac{p_{11}}{p_{00}} \tag{7}$$

Thus, define the lower and upper bounds on b as $b_L = \max\left\{0, \frac{p_{11}-p_{01}}{p_{00}}\right\}$ and $b_U = \min\left\{\frac{p_{11}}{p_{00}}, 1\right\}$. Our new maximization problem can be written as,

$$\max_b \quad \left(1 - \frac{p^*_{01}}{p_{01}}\right) b + \left(\frac{p^*_{01}}{p_{01}}\right) p_{11} \qquad \text{s.t.} \qquad b_L \leq b \leq b_U \tag{8}$$

Since the target function is linear, the maximum occurs at the extreme points of b. Thus, we have that,

$$p^*_{11} \leq \max\left\{\left(1 - \frac{p^*_{01}}{p_{01}}\right) b_L + \left(\frac{p^*_{01}}{p_{01}}\right) p_{11}, \left(1 - \frac{p^*_{01}}{p_{01}}\right) b_U + \left(\frac{p^*_{01}}{p_{01}}\right) p_{11}\right\} \tag{9}$$

Analogously, for the minimization problem, we must have,

$$p^*_{11} \geq \min\left\{\left(1 - \frac{p^*_{01}}{p_{01}}\right) b_L + \left(\frac{p^*_{01}}{p_{01}}\right) p_{11}, \left(1 - \frac{p^*_{01}}{p_{01}}\right) b_U + \left(\frac{p^*_{01}}{p_{01}}\right) p_{11}\right\} \tag{10}$$