Causal Inference Based on Directed Acyclic Graphical Models and the Randomization Distribution: A Probability-Sampling Approach

Joel E. Hanson

Abstract

This paper develops a method of design-based causal inference from a sample to a finite population. Causal effects are defined as finite population parameters based on the causal structure of the population. Only a non-parametric model for the joint distribution of the causal structure, which can be expressed as a directed acyclic graphical model, must be assumed. The finite population causal effects are expressed in two equivalent forms: (1) as stratified sampling type estimators with the strata determined by the confounders, and (2) as Horvitz-Thompson type estimators with the selection probabilities replaced by the propensity score. Both forms lead to sampling estimators that are subtle variations on common estimators, and we provide a delta method for computing their approximate variances.

Keywords: Randomization, Counterfactual, Confounding

1 Introduction: To Model or Not to Model ?

During the 1970's and 1980's, the question of whether to base descriptive population inferences on an assumed super-population parametric model or not has provoked heated, and at times, acrimonious debate in the sampling literature. One of the most cogent articles arguing against assuming a model as the basis for population inference was written in 1983 by Hansen, Madow, and Tepping [2]. They showed that even a slight misspecification in a model can lead to significant bias in estimating a population parameter, which could be estimated without bias using a consistent sampling estimator. Because the exact specification of any model for a population can never be known with certainty, they conclude that while design decisions may be guided and evaluated by models, "inferences concerning population characteristics should be made on the basis of the induced randomization, at least when samples are reasonably large." [2, page 776]. Evidence of a rapprochement between the two camps was evident by 1993 when a staunch advocate of model-dependent inference, T.M.F. Smith, conceded that due to the uncertainty in model specification inference based on the randomization distribution, "seems particularly appropriate for descriptive inferences in the public domain" [8, page 157]. However, Smith adamantly maintained that models were essential for analytic, or causal, inference. Interestingly, even though Hansen, Madow, and Tepping rejected statistical models for inference about population parameters, they also seem to accept the necessity of using statistical models for causal inference. In fact, they equate drawing causal inferences about a finite population with postulating a super population model for the "causal system" and estimating the parameters of that model [2, page 789].

In a recent (2002) review of the design-based vs. model-dependent inference debate, Graham Kalton also argued that analytic inferences required the assumption of a parametric statistical model [4]. This view is particularly surprising from someone who accepts model-dependent inference grudgingly and out of necessity:

> It is axiomatic that all models are false. The attraction of the design-base approach to survey inference for most *descriptive* estimation . . . is its avoidance of reliance on models. . . . My general approach to the use of model-dependent methods for *descriptive* estimation is to treat the model as a crutch, to be used only to the extent that the survey data cannot fully support the desired estimates. [emphasis added] [4, page 130]

This paper will adopt the same pragmatic philosophy with respect to the use of marginal structural models and other recent model-dependent methods in the causal analysis literature for the estimation of *causal* effects. However, we emphasize both that super-population parametric statistical models are not required for causal inference and that the parameters of such models are generally inadequate as estimates of causal effects. With a reasonably large sample, causal effects defined as finite population parameters can be consistently estimated based solely on the randomization distribution induced by the sampling design. When the sample size is not large enough, especially in some of the strata, a marginal structural model or some related method may need to be assumed. However, the population causal effect being estimated does not change whether or not any type of model is assumed. Thus, the decision to model or not to model depends only on the same practical considerations that arise when estimating descriptive parameters, and the risk of model misspecification bias need not be incurred to draw causal inferences when the sample is reasonably large.

Recent research in causal analysis has lead to the development of a formal nonparametric theory for evaluating causal effects using directed acyclic graphical models [5, 3] By relying on that research, in particular the results of Pearl and Robins, counterfactual causal effects are defined as finite population parameters based on the assumed causal structure of the population. Given a causal structure, which is also equivalent to specifying a directed acyclic graphical model, the causal effect of one variable on another can be derived by applying Pearl's definition of conditioning (intervening) on the causal variable. The resulting causal effect is expressed

as a function of observable finite population parameters, domain means and proportions. Even with an assumed complex time dependent causal structure, Robins non-parametric G-computation algorithm will yield well defined causal effects that depend only on observable finite population parameters. Hence, while a nonparametric model for the causal structure of the finite population must be assumed, none of the usual parametric or semi-parametric models need to be assumed.

Given the assumed causal structure, inference about the population causal effects can be based entirely on the randomization distribution induced by the sampling design. Because the causal assumptions are equivalent to specifying a directed acyclic graphical model for the variables in the population, the causal effects will have a much more intuitive definition than the usual parameters in log-linear or logistic regression models. We will see that the population causal effects can be represented as stratified sampling type estimators of domain means with the strata weights determined by the distribution of the confounders after conditioning (intervening) on the causal variable. Equivalently, the population causal effects can be represented as Horvitz-Thompson type estimators of domain means with the selection probabilities determined by the population propensity score—the probability of the causal variable given the confounders. Furthermore, both representations of causal effects lead naturally to unbiased sampling estimators, and thus inference about their sampling variability can be based entirely on the randomization distribution induced by the sampling design. Hence, probability-sampling methods are readily applicable to causal inference without parametric models.

Furthermore, common parametric statistical modes are generally inadequate to estimate causal effects. Including a confounding variable in a parametric model is in general not sufficient to control for the confounding [3]. Only in the simplest causal structures (for an example, see figure 1) can a parameter for the causal variable be interpreted as a causal effect, and then only when the model is linear in that parameter–i.e. no intersection terms with the causal variable. If interactions need to be included, then no single parameter is relevant to estimating the causal effect in the population. Furthermore, for all of the other causal structures considered in this paper, none of the common parametric models yield unbiased estimates of causal effects, even when linearity holds [3]. Hence, except for relatively simple causal structures with additional strong model-dependent assumptions, the usual parametric statistical models can not be relied upon to estimate causal effects.

Robins recently proposed a novel marginal structural model approach to causal inference which relies on weighting the observations by their propensity score and then fitting a parametric model to estimate the causal effect [3]. Without the modeling, this estimation method is equivalent to the Horvitz-Thompson type estimators discussed above. Hence, Robins method can be seen as a hybrid of the probability-sampling approach and the conventional parametric modeling approach. It also illustrates how fundamental the probability-sampling approach is to causal inference.

While marginal structural models may prove useful when the sample size is not large enough, especially in some strata, they are still subject to model misspecification bias. Hence, because of the dangers of model-dependent inference identified in the Hansen paper and confirmed in our simulations, this paper will argue for a model free approach to causal inference whenever possible. However, it must be emphasized that the fundamental assumption required for valid causal inference—a correctly specified directed acyclic graphical model—is still required when dispensing with parametric statistical models. Even though the model free approach allows finite population causal effects to be estimated with minimal bias in large samples, the validity of any causal inferences will still depend on the actual causal structure of the population.

Thus, it must be emphasized that the randomization distribution is only relevant to assessing the sampling variability of the estimate of any finite population parameter—including the causal effects defined above. The randomization allows us to assess how much an estimate is likely to change if another similar sample was drawn. Thus, if every value in the finite population where known, the causal effect could be computed without any sampling variability, yet the causal inference could still be in doubt if the causal assumptions are questioned. However, if the sample is not large enough or the design is flawed, the causal inference will be uncertain because the estimated value of the causal effect is uncertain. With model-dependent estimation the causal inference can be uncertain no matter how large the sample if the modeling assumptions are questioned.

In contrast, the model free approach has the virtue of yielding consistent estimates of population parameters with minimal assumptions in reasonably large samples, and consistent estimates of the finite population parameters are a necessary condition for causal inference. Hence, while emphasizing that causal inference always depend on causal assumptions, this paper will advocate a methodology that defines counterfactual causal effects in terms of finite population parameters and bases all inferences about those parameters on the randomization distribution induced by the sampling design.

2 Finite Population Causal Effects

In this section we rely on recent result in causal analysis to define counterfactual causal effects as finite population parameters, which can then be estimated using variations of common sampling estimators.

The N units of the finite population will be denoted $y_1, y_2, \ldots, y_N$. A sample of n units from the population will be denoted $y_1, y_2, \ldots, y_n$. As in most of our notation, we will follow Cochran [1] and denote population parameters by capital letters and sample estimates of those parameters by lower-case letters. Hence, the population mean and its sample estimate will be denoted, respectively: $\bar{Y} = \sum^N y_i$ and $\bar{y} = \sum^n y_i$.

Because stratified sampling is central to controlling confounding for causal inference, the following notation and its extensions will be used throughout the remainder of the article. The population of N units is divided into sub-populations of $N_1, N_2, \ldots, N_Z$ units, respectively. These non-overlapping

sub-populations partition the population, so that $N_1 + N_2 + \cdots + N_Z = N$.

The sub-populations are referred to as *strata*. The suffix z denotes the stratum and i the unit within the stratum. Thus, N_z denotes the total number of units within stratum z, n_z the total number of sampled units within stratum z, and $y_{i|z}$ the value of the ith unit in stratum z. The following notation refers to stratum z.

$$W_z = \frac{N_z}{N} \qquad \bar{Y}_z = \frac{1}{N_z}\sum_{i=1}^{N_z} y_{i|z}$$

$$\bar{y}_z = \frac{1}{n_z}\sum_{i=1}^{n_z} y_{i|z} \quad S_z^2 = \frac{1}{N_z - 1}\sum_{i=1}^{N_z}(y_{i|z} - \bar{Y}_z)^2$$

Because we will be identifying causal effects within strata, we will be comparing means and proportions of sub-populations within the strata and across the strata. A stratum z of N_z units is divided into sub-populations of $N_{z,1}, N_{z,2}, \ldots, N_{z,X}$ units, respectively. These non-overlapping sub-populations partition the stratum, so that $N_{z,1} + N_{z,2} + \cdots + N_{z,X} = N_z$.

These sub-populations extending across the strata are referred to as *domains*. Note that if the domains did not extend across the strata, then they would just create a finer stratification and no additional notation would be required. The suffix z denotes the stratum, x the domain and i the unit within the domain of the stratum. Thus, $N_{z,x}$ denotes the total number of units within stratum z and domain x, $n_{z,x}$ the total number of sampled units within stratum z and domain x, and $y_{i|z,x}$ the value of the ith unit in stratum z and domain x. The following notation all refers to stratum z and domain x.

$$\bar{Y}_{z,x} = \frac{1}{N_{z,x}}\sum_{i=1}^{N_{z,x}} y_{i|z,x} \quad S_{z,x}^2 = \frac{1}{N_{z,x} - 1}\sum_{i=1}^{N_{z,x}}(y_{i|z,x} - \bar{Y}_{z,x})^2$$

In order to think about the causal effect in the population, it will be helpful to define a joint distribution on all the population variables. The population causal effect will be defined in terms of a new post-intervention distribution of the variables in the population. In other words, the causal effect of a variable will be defined in terms of the distribution that would result if an intervention in the population occurred and all units were assigned a set value of the causal variable. The properties of this distribution can be expressed in terms of population parameters, which will allow causal effects to be expressed in terms of population parameters.

$$\mathbf{P}(Z = z) = W_z = \frac{N_z}{N} \qquad \mathbf{P}(Z = z \mid X = x) = W_{z|x}$$

$$\mathbf{E}(Y) = \bar{Y} = \sum_{i=1}^{N} y_i \qquad \mathbf{E}(Y \mid z) = \bar{Y}_z = \frac{1}{N_z}\sum_{i=1}^{N_z} y_{i|z}$$

$$\mathbf{E}(Y) = \sum_{i=1}^{Z} \mathbf{E}(Y \mid z)\mathbf{P}(z) = \sum_{i=1}^{Z} \bar{Y}_z \frac{N_z}{N}$$

$$\mathbf{E}(Y \mid x) = \sum_{i=1}^{Z} \mathbf{E}(Y \mid z, x)\mathbf{P}(z \mid x) = \sum_{i=1}^{Z} \bar{Y}_{z,x} \frac{N_{z,x}}{N_x}$$

When considering a graphical model for the joint distribution of the population, the question of confounding must all ways be addressed—are all the other relevant causal variables that are also associated with the treatment properly included in the population causal diagram? The goal of the study design is to include a sufficient number of variables in the causal model for a conclusion of "No unmeasured confounders" to be plausible.

Given a graphical model, population causal effects will be derived as stratified sampling type estimators with the strata determined by the confounders. The strata weights will be determined by the post-intervention distribution of the confounders. Within the strata, there is no longer any confounding of the causal effect, so the net causal effect will be the average over all the strata.

Hence, stratification for causal analysis serves to control for the confounding of the population causal effect, which is an entirely different purpose than stratification for the purpose of reducing the variance of a sample estimator—a common reason for stratification in sampling. However, as inference of population causal effects requires non-confounded and precise estimates of finite population parameters, stratification for both purposes should and will occur. We will, therefore, demonstrate that the population causal effect does not change if additional variables that are not confounders are stratified upon. In other words, while populating causal effects can be expressed in terms of stratified type estimators, the causal effect does not depend on the stratification—only the confounders.

Finally, we will see that population causal effects can also be expressed equivalently as Horvitz-Thompson type estimators. This form will prove to be a natural link to a new model-dependent method for causal analysis recently proposed by Robins—marginal structural models [3]. This link could also be useful when a model is deemed necessary to adjust for non-response bias in sampling. In fact, all of the myriad model-dependent methods from causal analysis could be employed for this purpose. However, the focus of this paper is on the use of probability-sampling methods in causal inference in order to avoid the reliance on modeling assumptions.

2.1 Stratified Sampling Type Estimators

Given a distribution for the finite population and an assumed causal structure for the joint distribution, which can be represented as a directed acyclic graph, causal effects can be defined as finite population parameters. We will begin with Pearl's method, and later simply quote Robins' G-computation algorithm, which can be used to define causal effects in longitudinal samples where there are time-varying confounders.

Using Pearl's method, causal effects can be derived quite naturally using a graphical model for the joint distribution of the population and a new definition of conditioning (intervening) on the causal variable. Perl provides a rigorous justification for the following heuristic:

1. Specify a directed acyclic graphical model for the joint distribution of the relevant population variables.

2. Derive the joint distribution of the variables using the graphical structure.

3. Derive the joint distribution of the other variables after intervening on the causal variables by deleting all the probabilities or conditional probabilities of the causal variables.

4. Then, derive the distribution of the response variable after intervening on the causal variable by integrating out all of the confounding variables.

5. Finally, compute the expected value of the response variable using its distribution after intervening on the causal variables.

We will illustrate the method on the simple example from the last section. The graphical model is depicted in figure 1.
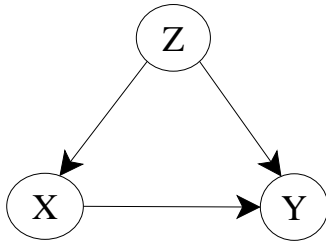

Figure 1: Z Confounds the Effect of X on Y

Given this joint distribution, we can derive the causal effect of x as follows:

$$\mathbf{P}(x, y, z) = \mathbf{P}(y \mid x, z)\,\mathbf{P}(x \mid z)\,\mathbf{P}(z)$$
$$\mathbf{P}(y, z \mid \check{x}) = \mathbf{P}(y \mid x, z)\,\mathbf{P}(z)$$
$$\mathbf{P}(y \mid \check{x}) = \sum_{z=1}^{Z}\mathbf{P}(y \mid x, z)\,\mathbf{P}(z)$$

Now we can use the distribution of Y after intervening on X to compute the average causal effect of X on Y as follows:

$$\mathbf{E}(Y \mid \check{x}) = \sum_{z=1}^{Z}\left(\sum_{i=1}^{N_{z,x}} y_{i|z,x}\,\mathbf{P}(y \mid x, z)\right)\mathbf{P}(z)$$

Now we simple translate this expression back into the familiar finite population notation using the distribution of the finite population to yield:

$$\mathbf{E}(Y \mid \check{x}) = \sum_{z=1}^{Z}\mathbf{E}(y \mid x, z)\,\mathbf{P}(z) = \sum_{z=1}^{Z}\frac{N_z}{N}\bar{Y}_{z,x} = \bar{Y}_{\check{x}}$$

Where $\bar{Y}_{\check{x}}$ has the same interpretation as before—the average value of y in the population if everyone where given x.

If $\frac{N_z}{N}$ is not known, an unbiased estimate is provided by $\frac{n_z}{n}$. Hence, in either case natural sample estimates are, respectively:

$$\bar{y}_{\check{x}} = \sum_{z=1}^{Z}\bar{y}_{z,x}\frac{N_z}{N} \quad \text{and} \quad \bar{y}_{\check{x}} = \sum_{z=1}^{Z}\bar{y}_{z,x}\frac{n_z}{n}$$

Both of these estimates are unbiased estimates of the population causal effect. The proofs in the both cases rely on the fact that, conditional on $n_{z,x}$, $\bar{y}_{z,x}$ is an unbiased estimate of $\bar{Y}_{z,x}$ for simple or stratified random sampling. When $\frac{N_z}{N}$ is not known, use the linearity of expectation and focus on $\mathbf{E}(\bar{y}_{z,x}\frac{n_z}{n})$, which can be written

$$\mathbf{E}\big(\,\mathbf{E}(\bar{y}_{z,x}\frac{n_z}{n} \mid n_{z,x})\big) = \bar{Y}_{z,x}\,\mathbf{E}(\frac{n_z}{n}) = \bar{Y}_{z,x}\frac{N_z}{N}$$

Hence, the finite population causal effect can be estimated without bias using a slight variation on the common stratified estimator for a domain mean. In the appendix we will derive the approximate variances for these estimators.

The causal effect based on the joint distribution represented by figure 1 can be estimated without confounding by a parametric model if the assumption of linearity holds. When there are interactions there is not a meaningful definition of causality associated with a single parameter. The situation is analogous to the problem of interpreting domain means that vary among the strata. However, model misspecification may be even more difficult to identify, especially in smaller samples, which is precisely when models are most relied upon.

Now we will consider the simplest causal structure where the effect can not be estimated by the usual parametric models even under the assumption of linearity. This situation is often referred to as time-varying confounders. In other words, Z is both a confounder of the effect of X_2 and on the causal pathway of X_1. Hence, none of the conventional methods of control work in this situation. For a thorough discussion see Robins refHernan00. The graphical model is depicted in figure 2.
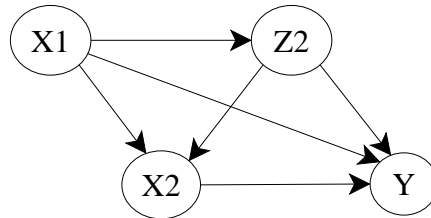

Figure 2: The Graphical Model for Two Causal Variables

Given this joint distribution, we can derive the causal effect of x as follows:

$$\mathbf{P}(x_1, x_2, y, z) = \mathbf{P}(y \mid x_1, x_2, z)\,\mathbf{P}(x_2 \mid z, x_1)\,\mathbf{P}(z \mid x_1)$$
$$\mathbf{P}(y, z \mid \check{x}_1, \check{x}_2) = \mathbf{P}(y \mid x_1, x_2, z)\,\mathbf{P}(z \mid x_1)$$
$$\mathbf{P}(y \mid \check{x}_1, \check{x}_2) = \sum_{z=1}^{Z}\mathbf{P}(y \mid x_1, x_2, z)\,\mathbf{P}(z \mid x_1)$$

Now we can use the distribution of Y after intervening on X_1 and X_2 to compute the average causal effect of X_1 and X_2 on Y as follows:

$$\mathbf{E}(Y \mid \check{x}_1, \check{x}_2) = \sum_{z=1}^{Z}\mathbf{E}(Y \mid x_1, x_2, z)\,\mathbf{P}(z \mid x_1)$$

Now we simple translate this expression back into the familiar finite population notation using the distribution of the finite population to yield:

$$\mathbf{E}(Y \mid \check{x}_1, \check{x}_2) = \sum_{z=1}^{Z} \frac{N_{z,x_1}}{N_{x_1}} \bar{Y}_{z,x_1,x_2}$$

Where $\bar{Y}_{\check{x}_1,\check{x}_2}$ has the same interpretation as before—the average value of y in the population if everyone where given x_1 and x_2.

If $\frac{N_{z,x_1}}{N_{x_1}}$ is not known, an unbiased estimate is provided by $\frac{n_{z,x_1}}{n_{x_1}}$. This follows by observing that, conditioning on n_{x_1}, n_{z,x_1} is a hyper-geometric random variable with expected value $= n_{x_1} \frac{N_{z,x_1}}{N_{x_1}}$ Hence,

$$\mathbf{E}(\frac{n_{z,x_1}}{n_{x_1}}) = \mathbf{E}(\mathbf{E}(n_{z,x_1} \mid n_{x_1})\frac{1}{n_{x_1}})) = \frac{N_{z,x_1}}{N_{x_1}}$$

Depending on whether $\frac{N_{z,x_1}}{N_{x_1}}$ is know or not, natural sample estimates are, respectively:

$$\bar{y}_{\check{x}_1,\check{x}_2} = \sum_{z=1}^{Z} \frac{N_{z,x}}{N_x} \bar{y}_{z,x_1,x_2} \quad \text{and} \quad \bar{y}_{\check{x}_1,\check{x}_2} = \sum_{z=1}^{Z} \frac{n_{z,x_1}}{n_{x_1}} \bar{y}_{z,x_1,x_2}$$

Both of these estimates are unbiased estimates of the population causal effect. The proofs in the both cases rely on the fact that, conditional on n_{z,x_1,x_2}, $\bar{y}_{z,x_1,x_2}$ is an unbiased estimate of $\bar{Y}_{z,x_1,x_2}$. In the latter case, the proof also utilizes the fact that $\frac{n_{z,x_1}}{n_{x_1}}$ is an unbiased estimate of $\frac{N_{z,x_1}}{N_{x_1}}$. Thus, the finite population causal effect can be estimated using a stratified estimator for a domain mean with the strata weights determined by the post-intervention distribution of the confounder. In the appendix we will derive the approximate variances for this estimator.

Now we will consider the most general causal structure addressed in this paper. For any type of longitudinal sampling involving causal inference this is the appropriate model for causal inference. All the causal variables are denoted by X_j; all of the confounders are denoted by Z_j; and the response is denoted by Y. Besides the graphical structure, we must assume that there are no unmeasured confounders. In other words, all the Z's that matter are included.

The causal effect for the graphical model depicted in figure 3 is known as Robin's G-computation Algorithm [3]. Note that many of the causal arrows have been suppressed for graphical clarity. Space permitting, there should be an arrow from every variable into Y and from every arrow with a smaller subscript into Z_j.

We will follow Robins and use the following notation $\bar{x}_k$ denotes $(x_1, x_2, \ldots, x_k)$, while the symbol $\check{}$ in $\check{\bar{x}}_k$ still denotes the causal effect under an intervention on $(x_1, x_2, \ldots, x_k)$. Robins G-computation algorithm expresses the average value of Y under an intervention on $\bar{x}_k$ in terms of common finite population parameters.

$$\mathbf{E}(Y \mid \check{\bar{x}}_k) = \sum_{\bar{z}} \mathbf{E}(Y \mid \bar{x}_k, \bar{z}_k) \mathbf{P}(z_k \mid \bar{z}_{k-1}, \bar{x}_{k-1}) \times \cdots \times \mathbf{P}(z_1)$$
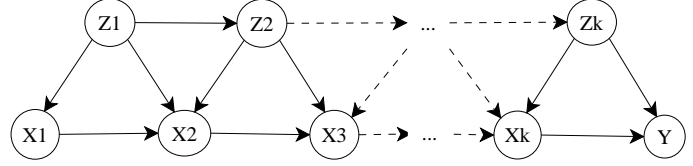

Figure 3: The Graphical Model for Robins G-Computation Algorithm

This formula could be derived using the method by Pearl just as for the two previous causal structures considered.

Using the finite population distribution, the causal effect of $\bar{x}_k$ on Y can be written as follows:

$$\bar{Y}_{\check{\bar{x}}_k} = \sum_{\bar{z}} \bar{Y}_{\bar{z}_k, \bar{x}_k} \frac{N_{\bar{z}_k, \bar{x}_{k-1}}}{N_{\bar{z}_{k-1}, \bar{x}_{k-1}}} \times \cdots \times \frac{N_{z_1}}{N}$$

Even when all the $\frac{N_{\bar{z}_k, \bar{x}_{k-1}}}{N_{\bar{z}_{k-1}, \bar{x}_{k-1}}}$ are unknown then $\bar{Y}_{\check{\bar{x}}_k}$ can be estimated without bias by

$$\bar{y}_{\check{\bar{x}}_k} = \sum_{\bar{z}} \bar{y}_{\bar{z}_k, \bar{x}_k} \frac{n_{\bar{z}_k, \bar{x}_{k-1}}}{n_{\bar{z}_{k-1}, \bar{x}_{k-1}}} \times \cdots \times \frac{n_{z_1}}{n}$$

The proof is similar to proof of unbiasedness for the previous two causal estimator considered. Conditional on $n_{\bar{z}_k, \bar{x}_k}$, $\bar{y}_{\bar{z}_k, \bar{x}_k}$ is an unbiased estimate of $\bar{Y}_{\bar{z}_k, \bar{x}_k}$. Hence, the proof relies on repeated application of the law of iterated expectation. After conditioning on $n_{\bar{z}_k, \bar{x}_k}$, we are left with the problem of computing

$$\mathbf{E}\left(\frac{n_{\bar{z}_k, \bar{x}_{k-1}}}{n_{\bar{z}_{k-1}, \bar{x}_{k-1}}} \frac{n_{\bar{z}_{k-1}, \bar{x}_{k-2}}}{n_{\bar{z}_{k-2}, \bar{x}_{k-2}}} \times \cdots \times \frac{n_{z_1}}{n}\right)$$

We begin by conditioning on $n_{\bar{z}_{k-1}, \bar{x}_{k-1}}$, for then every fraction is constant but the numerator of the first which, for fixed x_k, is distributed as a hyper-geometric random variable with $\mathbf{E}(n_{\bar{z}_k, \bar{x}_{k-1}} \mid n_{\bar{z}_{k-1}, \bar{x}_{k-1}}) = n_{\bar{z}_{k-1}, \bar{x}_{k-1}} \frac{N_{\bar{z}_k, \bar{x}_{k-1}}}{N_{\bar{z}_{k-1}, \bar{x}_{k-1}}}$. Thus, after conditioning the expectation above is reduced to the following

$$\frac{N_{\bar{z}_k, \bar{x}_{k-1}}}{N_{\bar{z}_{k-1}, \bar{x}_{k-1}}} \mathbf{E}\left(\frac{n_{\bar{z}_{k-1}, \bar{x}_{k-2}}}{n_{\bar{z}_{k-2}, \bar{x}_{k-2}}} \frac{n_{\bar{z}_{k-2}, \bar{x}_{k-3}}}{n_{\bar{z}_{k-3}, \bar{x}_{k-3}}} \times \cdots \times \frac{n_{z_1}}{n}\right)$$

Hence, now condition on $n_{\bar{z}_{k-2}, \bar{x}_{k-2}}$ and repeat the argument above. Continue until $k = 1$.

The same style of computing by successive conditioning can be used to derive the approximate variance of the G-computation algorithm estimator, $\bar{y}_{\check{\bar{x}}_k}$. The details will be sketched in the appendix.

In all of the estimators considered in this section, we considered strata where the weights were unknown. This would be senseless for non-causal estimates because the estimator would simply collapse back to the non-stratified form. However, for causal estimators stratification is essential to control for confounding, but in the next section we shall see that stratification on variables that are not confounders will also lead to causal estimators collapsing.

2.2 Horvitz-Thompson Type Estimators and Marginal Structural Models

Now we will give the motivation for the model-dependent causal estimates. First consider rewriting the causal effect for the simple causal structure depicted in figure 1 as follows

$$
\begin{aligned}
\mathbf{E}(Y \mid \breve{x}) &= \sum_{z=1}^{Z} \sum_{i=1}^{N_{z,x}} y_{i|z,x} \frac{1}{N_{z,x}} \frac{N_z}{N} = \frac{1}{N} \sum_{z}^{Z} \sum_{i=1}^{N_{z,x}} \frac{y_{i|z,x}}{\frac{N_{z,x}}{N_z}} \\
&= \frac{1}{N} \sum_{z}^{Z} \sum_{i=1}^{N_{z,x}} \frac{y_{i|z,x}}{\pi_{x|z}} = \frac{1}{N} \left(\sum_{i=1}^{N_x} \sum_{z}^{Z} \frac{y_{i|z,x}}{\pi_{x|z}} \right) \\
&= \frac{1}{N} \left(\sum_{i=1}^{N_x} \frac{y_{i|x}}{\pi_{x|i}} \right)
\end{aligned}
$$

where $\pi_{x|z} = \frac{N_{z,x}}{N_z} = \mathbf{P}(x \mid z)$ is the proportion of units in stratum z that are also in stratum x and $\pi_{x|i} = \pi_{x|z(i)}$ is just the same quantity with the dependence on z suppressed. Hence, the population causal effect is in the form of a Hansen-Hurwitz type estimator, which can be interpreted as the average value of Y in the population if everyone were assigned x. Weighting by the reciprocal of the probability of being in stratum x—also know as the propensity score—serves to project the value of $y_i \mid z, x$ to all of the units in stratum z that are not also in x. For example, if $\pi_{x|z} = 1/2$ for every x and y, then each $y_{i|x}$ will be counted twice in computing the average for all those units in stratum x. If $X \perp\!\!\!\perp Z$ then $\pi_{x|z} = \frac{N_x}{N}$ then the formula simply reduces to the domain mean of x, for in this case Z is not a confounder.

Notice that we are summing over all the values of $y_{i|x}$. This is because units with the same value of y_i can have different values of z and hence different propensity scores. Thus, writing the formula in the form of a Horvitz-Thompson estimator would have required keeping the double sum because the dependence on z could not have been suppressed within $z(i)$

The Hansen-Hurwitz type estimator makes it clear that the causal effect of x_1 compared to x_0 can be estimated simply as

$$
\frac{1}{N} \left(\sum_{i=1}^{N_{x_1}} \frac{y_{i|x_1}}{\pi_{x_1|i}} - \sum_{i=1}^{N_{x_0}} \frac{y_{i|x_0}}{\pi_{x_0|i}} \right)
$$

To motivate Robins' IPWT method, notice that this expression can be thought of as comparing simple averages in a counterfactual population that was obtained from the original population by inverse-propensity-score weighting. In this counterfactual population the effect of X on Y is no longer confounded and so can be considered as a simple difference in domain averages.

The astute reader will notice that the estimator is not quite in the right form for a domain mean. Shouldn't the estimator be divided by N_x, not N? We will see that the answer is "No". In the counterfactual population the sample size of both domain x_0 and x_1 is N. Let $'$ denote the new population, so that we have

$$
N'_{z,x} = \frac{N_x}{N_{z,x}} N_{z,x} = N_z \qquad N'_x = \sum_{z} N'_{z,x} = N
$$

$$
N'_z = \sum_{x} N'_{z,x} = 2N_z \qquad N' = \sum_{x} N'_x = 2N
$$

Notice that in general, the 2 in the formula above will be replaced by the number of levels of x—the effect of each level is estimated as if we were able to intervene and assign that level to the entire population, which is also why the number of levels multiplies the sample size. There is one counterfactual population for each level of x. Also, the distribution of the confounder does not change in this counterfactual population $\mathbf{P}(Z' = z) = \mathbf{P}(Z = z) \quad \forall z \in Z$. Furthermore, the distribution of $Y \mid Z, X$ does not change since $\mathbf{P}(Y' = y \mid Z' = z, X' = x) = \frac{N'_{y,z,x}}{N'_{z,x}} = \frac{N_{y,z,x}}{N_{z,x}}$

Finally in this new population it is easy to see that there is no longer confounding since

$$
\mathbf{P}(X') = \frac{N}{2N} = \frac{1}{2} \qquad \text{and} \qquad \mathbf{P}(X' \mid Z') = \frac{N_z}{2N_z} = \frac{1}{2}
$$

Ironically because there is no confounding in the counterfactual population, parametric models for that population can be used for causal inference. This was the great insight behind the marginal structural models approach advocated by Robins and his colleagues. cite In fact, this insight extends to any statistical comparison between $\mathbf{P}(Y'|X' = x_1)$ and $\mathbf{P}(Y'|X' = x_1)$, but unfortunately the estimation of standard errors is complicated because even with simple random sampling the counterfactual population has repeat measurements on all units with $\pi_{x_0|z(i)} < 1$, which is hopefully all units since otherwise some $\pi_{x_0|z(i)} = 0$. Robins has proposed a GEE approach to handling this problem that can be implemented in most statistical software.

So far we have only discussed the simplest causal structure, however we can consider more complicated causal estimates since the Hansen-Hurwitz form derived above can also be derived for Robin's G-computation algorithm. Hence, the same approach above applied to figure 3 yields,

$$
\mathbf{E}(Y \mid \breve{x}_k) = \frac{1}{N} \sum_{i=1}^{N_{\bar{x}_k}} \frac{y_{i|\bar{x}_k}}{\pi_{\bar{x}_k|\bar{z}_k(i)}}
$$

where $\pi_{\bar{x}_k|\bar{z}_k(i)} = \mathbf{P}(x_k \mid \bar{z}_k, \bar{x}_{k-1}) \mathbf{P}(x_{k-1} \mid \bar{z}_{k-1}, \bar{x}_{k-2}) \times \cdots \times \mathbf{P}(x_1 \mid z_1)$ Thus, marginal structural models fit on these counterfactual populations will also have a causal interpretation.

For example, consider a density function f_θ, then θ can be defined as a finite population causal parameter using the pseudo-log-likelihood

$$
\ell(\theta) = \sum_{k=1}^{K} \sum_{i=1}^{N_{\bar{x}_k}} \frac{\log(f_\theta(y_{i|\bar{x}_k}))}{\pi_{\bar{x}_k|\bar{z}_k(i)}}
$$

Where θ solves $\frac{\partial \ell(\theta)}{\partial \theta} = 0$.

Finally, note that by writing the population causal effect in this manner, the form of the estimator with any unequal probability of selection sampling is readily apparent. The sample estimate would have the identical form except that the propensity score would be multiplied by the probability of selection.

3 Proofs of Approximate Variances

We begin by considering the simplest causal estimators corresponding to figure 1

$$\bar{y}_{\tilde{x}} = \sum_{z=1}^{Z} \frac{N_z}{N} \bar{y}_{z,x} \quad \text{and} \quad \bar{y}_{\tilde{x}} = \sum_{z=1}^{Z} \frac{n_z}{n} \bar{y}_{z,x}$$

In the first equation above we assume that $\frac{N_z}{N}$ is known, which is the usual case in stratified sampling. If $n_{z,x}$ is known then the variance is straightforward to compute, so we will only consider the case that $n_{z,x}$ is unknown. However there are still two sub-cases to consider. If n_z is known then to the order $\frac{1}{n^2}$ we get

$$\mathbf{V}(\bar{y}_{\tilde{x}}) \approx \sum_z (1 - f_z) W_z^2 \frac{S_{z,x}^2}{\pi_{x|z} n_z} \left(1 + \frac{1 - \pi_{x|z}}{\pi_{x|z} n_z} \right)$$

where $\pi_{x|z} = \mathbf{P}(x \mid z) = \frac{N_{z,x}}{N_z}$ is the propensity score as before. Note that $\pi_{x|z} n_z$ is the expected sample size in strata z, and if $\pi_{x|z} = 1$ then the formula reduces to the usual formula for stratified sampling.

When n_z is not known then to the order $\frac{1}{n^2}$ we get

$$\mathbf{V}(\bar{y}_{\tilde{x}}) \approx (1 - f) \sum_z W_z \frac{S_{z,x}^2}{\pi_{x,z} n} \left(1 + \frac{1 - \pi_{x,z}}{n \pi_{x,z}} \right)$$

where $\pi_{x,z} = \mathbf{P}(x, z) = \frac{N_{z,x}}{N}$. Note that when $N_{z,x} = N_z$ (i.e. every unit in strata z has x) then the formula reduces to the formula for the variance of the post-stratification estimator [9, page 124].

Now we consider a case that does not have a counterpart in traditional sampling theory—unknown strata weights. It would be meaningless to estimate the strata weights from the same sample that the domain means were estimated from because the estimator would just collapse back to the form for simple random sampling. In fact that will only happen with the causal estimator whenever the causal variable x occurs for every unit in the population, for then there is no counterfactual estimate necessary and we are just computing the variance of $\bar{y}$ Hence, when $W_z = \frac{N_z}{N}$ is not known, then to the order $\frac{1}{n^2}$ we get

$$\mathbf{V}(\bar{y}_{\tilde{x}}) \approx (1 - f) \sum_z W_z^2 \frac{S_{z,x}^2}{\pi_{x,z} n} \left(1 + \frac{1 - \pi_{x|z}}{n \pi_{x,z}} \right)$$

Notice that the formula will collapse to the variance under simple random sampling precisely when $N_{z,x} = N_z$.

We will sketch the proof for this last case. Begin by conditioning on $n_{z,x} \forall z \in Z$ and observe that for $n_{z,x} > 0$ (which we assume is a set with probability very close to 1)

$$\mathbf{V}(\bar{y}_{\tilde{x}} \mid n_{z,x} \forall z \in Z) = \sum_z \left(\frac{n_z}{n} \right)^2 \frac{S_{z,x}^2}{n_{z,x}} \left(1 - \frac{n_{z,x}}{N_{z,x}} \right)$$

Now integrate over the $n_{z,x} > 0$ (the variance is zero otherwise) to yield

$$\mathbf{V}(\bar{y}_{\tilde{x}}) = \sum_z \frac{1}{n^2} S_{z,x}^2 \left(\mathbf{E} \left(\frac{n_z^2}{n_{z,x}} \right) - \mathbf{E} \left(\frac{n_z^2}{N_{z,x}} \right) \right)$$

We will use a well-known Taylor series approximation to estimate the first expectation. To begin we condition on n_z and note that then $n_{z,x}$ has a hyper-geometric distribution with mean $\frac{N_{z,x}}{N_z}$. Then we use a second order Taylor series approximation of $\frac{1}{x}$ to yield

$$\mathbf{E} \left(\frac{n_z^2}{n_{z,x}} \mid n_z \right) \approx n_z^2 \left(\frac{1}{n_z} \frac{N_z}{N_{z,x}} + \frac{1}{n_z^2} \frac{N_z}{N_{z,x}} \left[\frac{N_z}{N_{z,x}} - 1 \right] \left(1 - \frac{n_z}{N_z} \right) \right)$$

Now integrate out the n_z to get

$$\mathbf{E} \left(\frac{n_z^2}{n_{z,x}} \right) \approx n \frac{N_z}{N} \frac{N_z}{N_{z,x}} + \frac{N_z}{N_{z,x}} \left[\frac{N_z}{N_{z,x}} - 1 \right] \left(1 - \frac{n}{N} \right)$$

Substituting this expression for the first expectation along with the following simple approximation for the second expectation, $\mathbf{E} \left(\frac{n_z}{n} \right)^2 \approx \left(\frac{N_z}{N} \right)^2 \left(1 + \frac{1}{n} \right)$, into the variance equation above yields

$$\mathbf{V}(\bar{y}_{\tilde{x}}) = \sum_z \frac{N_z}{N_{z,x}} S_{z,x}^2 \left[\frac{N_z}{N} \frac{1 - f}{n} + \left(\frac{N_z}{N_{z,x}} - 1 \right) \left(\frac{1 - f}{n^2} \right) \right]$$

$$= (1 - f) \sum_z W_z^2 \frac{S_{z,x}^2}{\pi_{z,x} n} \left(1 + \frac{1 - \pi_{x|z}}{\pi_{z,x} n} \right)$$

In the general case , after conditioning on $n_{\bar{z}_k, \bar{x}_k}$ to derive an expression for the variance, we are left with the problem of computing

$$\mathbf{E} \left(\frac{1}{n_{\bar{z}_k, \bar{x}_k}} \left(\frac{n_{\bar{z}_k, \bar{x}_{k-1}}}{n_{\bar{z}_{k-1}, \bar{x}_{k-1}}} \right)^2 \left(\frac{n_{\bar{z}_{k-1}, \bar{x}_{k-2}}}{n_{\bar{z}_{k-2}, \bar{x}_{k-2}}} \right)^2 \times \cdots \times \left(\frac{n_{z_1}}{n} \right)^2 \right)$$

We will focus on the first two terms in the expectation and show how to reduce the problem by one term. The method we use can be repeated until no random terms remain. We will need the following approximation, to the order n^{-2},

$$\mathbf{E} \left(\frac{1}{n_{\bar{z}_k, \bar{x}_k}} \mid n_{\bar{z}_k, \bar{x}_{k-1}} \right) \approx$$

$$\frac{1}{n_{\bar{z}_k, \bar{x}_{k-1}}} \frac{N_{\bar{z}_k, \bar{x}_{k-1}}}{N_{\bar{z}_k, \bar{x}_k}} + \left(\frac{1}{n_{\bar{z}_k, \bar{x}_{k-1}}} \right)^2 \frac{N_{\bar{z}_k, \bar{x}_{k-1}}}{N_{\bar{z}_k, \bar{x}_k}} \left[\frac{N_{\bar{z}_k, \bar{x}_{k-1}}}{N_{\bar{z}_k, \bar{x}_k}} - 1 \right]$$

which also holds with k replaced by $k - 1$, of course.

And we need the simple approximation, also to the order n^{-2},

$$\mathbf{E} \left(\frac{1}{(n_{\bar{z}_{k-1}, \bar{x}_{k-1}})^2} \mid n_{\bar{z}_{k-1}, \bar{x}_{k-2}} \right) \approx$$

$$\left(\frac{1}{n_{\bar{z}_{k-1}, \bar{x}_{k-2}}} \right)^2 \left(\frac{N_{\bar{z}_{k-1}, \bar{x}_{k-2}}}{N_{\bar{z}_{k-1}, \bar{x}_{k-1}}} \right)^2$$

Hence, now we are ready to estimate the first two terms in the expectation above. We begin by conditioning on $n_{\bar{z}_k,\bar{x}_{k-1}}$, for then every fraction is constant but the first which, for fixed x_k, is distributed as a hyper-geometric random variable with $\mathbf{E}(n_{\bar{z}_k,\bar{x}_k} \mid n_{\bar{z}_k,\bar{x}_{k-1}}) = n_{\bar{z}_k,\bar{x}_{k-1}} \frac{N_{\bar{z}_k,\bar{x}_k}}{N_{\bar{z}_k,\bar{x}_{k-1}}}$. Thus,

$$\mathbf{E}\left(\frac{(n_{\bar{z}_k,\bar{x}_{k-1}})^2}{n_{\bar{z}_k,\bar{x}_k}} \mid n_{\bar{z}_k,\bar{x}_{k-1}}\right) \approx n_{\bar{z}_k,\bar{x}_{k-1}} \frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_k,\bar{x}_k}} + $$
$$\frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_k,\bar{x}_k}}\left[\frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_k,\bar{x}_k}} - 1\right]$$

Next condition on $n_{\bar{z}_{k-1},\bar{x}_{k-1}}$, and compute

$$\mathbf{E}(n_{\bar{z}_k,\bar{x}_{k-1}} \mid n_{\bar{z}_{k-1},\bar{x}_{k-1}}) = n_{\bar{z}_{k-1},\bar{x}_{k-1}} \frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_{k-1},\bar{x}_{k-1}}}$$

Thus, by successive conditioning

$$\mathbf{E}\left(\frac{1}{n_{\bar{z}_k,\bar{x}_k}}\left(\frac{(n_{\bar{z}_k,\bar{x}_{k-1}})}{n_{\bar{z}_{k-1},\bar{x}_{k-1}}}\right)^2 \mid n_{\bar{z}_{k-1},\bar{x}_{k-1}}\right) \approx$$
$$\frac{1}{n_{\bar{z}_{k-1},\bar{x}_{k-1}}}\frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_{k-1},\bar{x}_{k-1}}}\frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_k,\bar{x}_k}}$$
$$+ \left(\frac{1}{n_{\bar{z}_{k-1},\bar{x}_{k-1}}}\right)^2 \frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_k,\bar{x}_k}}\left[\frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_k,\bar{x}_k}} - 1\right]$$

Now we condition on $n_{\bar{z}_{k-1},\bar{x}_{k-2}}$ and use the approximations above to yield

$$\mathbf{E}\left(\frac{1}{n_{\bar{z}_k,\bar{x}_k}}\left(\frac{(n_{\bar{z}_k,\bar{x}_{k-1}})}{n_{\bar{z}_{k-1},\bar{x}_{k-1}}}\right)^2 \mid n_{\bar{z}_{k-1},\bar{x}_{k-2}}\right) \approx$$
$$\left(\frac{1}{n_{\bar{z}_{k-1},\bar{x}_{k-2}}}\frac{N_{\bar{z}_{k-1},\bar{x}_{k-2}}}{N_{\bar{z}_{k-1},\bar{x}_{k-1}}}\right)\frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_{k-1},\bar{x}_{k-1}}}\frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_k,\bar{x}_k}}$$
$$+ \left(\left(\frac{1}{n_{\bar{z}_{k-1},\bar{x}_{k-2}}}\right)^2 \frac{N_{\bar{z}_{k-1},\bar{x}_{k-2}}}{N_{\bar{z}_{k-1},\bar{x}_{k-1}}}\left[\frac{N_{\bar{z}_{k-1},\bar{x}_{k-2}}}{N_{\bar{z}_{k-1},\bar{x}_{k-1}}} - 1\right]\right)$$
$$\times \frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_{k-1},\bar{x}_{k-1}}}\frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_k,\bar{x}_k}}$$
$$+ \left(\frac{1}{n_{\bar{z}_{k-1},\bar{x}_{k-2}}}\right)^2 \left(\frac{N_{\bar{z}_{k-1},\bar{x}_{k-2}}}{N_{\bar{z}_{k-1},\bar{x}_{k-1}}}\right)^2 \frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_k,\bar{x}_k}}\left[\frac{N_{\bar{z}_k,\bar{x}_{k-1}}}{N_{\bar{z}_k,\bar{x}_k}} - 1\right]$$

At this point, we are back where we started except that k has been reduced by one, so if we continue we will be able to estimate the entire expectation and hence the variance of any G-computation causal estimator from a stratified or simple random sample.

We will illustrate the method on the relatively simple causal estimator for the graphical model depicted in figure 2. We saw that a causal estimator in that case was given by

$$\bar{y}_{\bar{x}_1,\bar{x}_2} = \sum_{z=1}^{Z} \frac{n_{z,x_1}}{n_{x_1}}\bar{y}_{z,x_1,x_2}$$

Hence, in order to approximate the variance we must be able to estimate $\mathbf{E}\left(\frac{1}{n_{\bar{z}_2,\bar{x}_2}}\left(\frac{(n_{\bar{z}_2,\bar{x}_1})}{n_{\bar{z}_1,\bar{x}_1}}\right)^2 \mid n_{\bar{z}_1,\bar{x}_0}\right)$ realizing that $z = $

z_2, $n_{\bar{z}_1,\bar{x}_0} = n$, $n_{\bar{z}_1,\bar{x}_1} = n_{x_1}$, and so on. Hence making these substitutions we get

$$\left(\frac{1}{n}\frac{N}{N_{\bar{x}_1}} + \left(\frac{1}{n}\right)^2 \frac{N}{N_{\bar{x}_1}}\left[\frac{N}{N_{\bar{x}_1}} - 1\right]\right)\frac{N_{\bar{z},\bar{x}_1}}{N_{\bar{x}_1}}\frac{N_{\bar{z},\bar{x}_1}}{N_{\bar{z},\bar{x}_2}}$$
$$+ \left(\frac{1}{n}\right)^2 \left(\frac{N}{N_{\bar{x}_1}}\right)^2 \frac{N_{\bar{z},\bar{x}_1}}{N_{\bar{z},\bar{x}_2}}\left[\frac{N_{\bar{z},\bar{x}_1}}{N_{\bar{z},\bar{x}_2}} - 1\right]$$

Substituting into the variance equation yields the following approximate variance

$$\mathbf{V}(\bar{y}_{\bar{x}_2}) \approx (1-f)\sum_z W_{z|x_1}^2 \frac{S_{z,\bar{x}_2}^2}{\pi_{\bar{x}_2,z}n}\left(1 + \frac{1 - \pi_{x_1}}{\pi_{x_1}n} + \frac{1 - \pi_{x_2|z,x_1}}{\pi_{x_2|z,x_1}n}\right)$$

Finally, it should be noted that this method of proof will not work for unequal probability sampling because after conditioning the distribution of the remaining sample sizes and domain means depends on the sample. Hence, the conditional variance formula used above no longer holds. Though a Taylor series expansion of the estimator itself could be employed in the simpler cases, the details are much more complicated because covariance terms must be estimated as well. However, for the Horvitz-Thompson type estimators considered in this paper, we saw that they had equivalent representations as stratified sampling type estimators. Thus, as long as the sampling method is simple or stratified random sampling and the sample sizes are large, these approximate variance formulas can be used for both types of estimators.

References

[1] Cochran, W. (1977), *Sampling Techniques* (3rd ed.), New York: Wiley.

[2] Hansen, M., Madow, W., and Tepping, B. (1983), "An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Surveys," *Journal of the American Statistical Association,* 78, 776-793.

[3] Hernan, M., Brumback, B., and Robins, J. (2000), "Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men," *Epidemiology,* 11, 561-570.

[4] Kalton, G. (2002), "Models in the Practice of Survey Sampling (Revisited)," *Journal of Official Statistics,* 18, 129-154.

[5] Pearl, J. (1995), "Causal Diagrams for Empirical Research," *Biometrika,* 82, 669-688.

[6] Skinner, C.J., Holt, D., and Smith, T.M.F. (ed.) (1989) *Analysis of Complex Surveys,* New York: Wiley.

[7] Smith, T.M.F. (1991), "Post-Stratification," *The Statistician,* 40, 315-323.

[8] Smith, T.M.F. (1993), "Populations and Selections: Limitations of Statistics," *Journal of the Royal Statistical Society,* Ser. A, 156, 144-166.

[9] Thompson, S. K. (2002), *Sampling* (2nd ed.), New York: Wiley.

[10] Thompson, S. K., and Seber, G. (1996), *Adaptive Sampling,* New York: Wiley.