Augmentation

The purpose of **augmentation** is to provide nonparametric covariate adjustment for a marginal treatment contrast. An augmented estimator always has variance less than or equal to the original unadjusted estimator.

For treatment arms $k \in \{0, 1\}$, let Y_{ik} denote the outcome of subject i in arm k and X_{ik} a vector of covariates. Suppose $g(\theta_1, \theta_0)$ is a treatment contrast, and that the estimator of each θ_k admits an influence function expansion:

$$\hat{\theta}_k - \theta_k = \frac{1}{n_k}\sum_{i=1}^{n_k}\psi(Y_{ik}; \theta_k) + o_p(1) = \frac{1}{n_k}\sum_{i=1}^{n_k}\psi_{ik} + o_p(1)$$

The unadjusted treatment contrast is:

$$\hat{\Delta}_{\text{unadj}} = g(\hat{\theta}_1, \hat{\theta}_0). \tag{1.1}$$

The adjusted treatment contrast takes the form:

$$\hat{\Delta}_{\text{adj}} = g(\hat{\theta}_1, \hat{\theta}_0) - \omega'(\bar{X}_1 - \bar{X}_0),$$

where ω is a weight vector:

$$\omega = \Sigma_{XX}^{-1}\Sigma_{X\Delta}.$$

Σ_{XX} is the variance of $\bar{X}_1 - \bar{X}_0$:

$$\Sigma_{XX} = \mathbb{V}(\bar{X}_1 - \bar{X}_0) = \mathbb{V}(\bar{X}_1) + \mathbb{V}(\bar{X}_0).$$

Letting $\hat{\mu}_X$ denote the grand mean:

$$\hat{\mu}_X = \frac{1}{n_1 + n_0}\left\{\sum_{i=1}^{n_1} X_{i1} + \sum_{i=1}^{n_0} X_{i0}\right\},$$

Σ_{XX} may be estimated as:

$$\hat{\Sigma}_{XX} = \frac{1}{n_1^2}\sum_{i=1}^{n_1}(X_{i1} - \hat{\mu}_X)^{\otimes 2} + \frac{1}{n_0^2}\sum_{i=1}^{n_0}(X_{i0} - \hat{\mu}_X)^{\otimes 2}. \tag{1.2}$$

$\Sigma_{X\Delta}$ is the covariance between $\bar{X}_1 - \bar{X}_0$ and $g(\hat{\theta}_1, \hat{\theta}_0)$. To obtain this, consider the Taylor expansion of $g(\hat{\theta}_1, \hat{\theta}_0)$:

$$g(\hat{\theta}_1, \hat{\theta}_0) = g(\theta_1, \theta_0) + \dot{g}_1(\theta_1, \theta_0)(\hat{\theta}_1 - \theta_1) + \dot{g}_0(\theta_1, \theta_0)(\hat{\theta}_0 - \theta_0) + o_p(1).$$

Now:

$$
\begin{aligned}
\Sigma_{X\Delta} &= \mathbb{C}\{\bar{X}_1 - \bar{X}_0, g(\hat{\theta}_1, \hat{\theta}_0)\} = \mathbb{C}\{\bar{X}_1, g(\hat{\theta}_1, \hat{\theta}_0)\} - \mathbb{C}\{\bar{X}_0, g(\hat{\theta}_1, \hat{\theta}_0)\} \\
&= \dot{g}_1(\theta_1, \theta_0)\mathbb{C}(\bar{X}_1, \hat{\theta}_1 - \theta_1) - \dot{g}_0(\theta_1, \theta_0)\mathbb{C}(\bar{X}_0, \hat{\theta}_0 - \theta_0)
\end{aligned}
$$

Note that $\bar{X}_k$ only covaries with the effect estimate $\hat{\theta}_k$ from the same treatment arm.

Using the influence function expansions:

$$
\begin{aligned}
\Sigma_{X\Delta} &= \dot{g}_1(\theta_1, \theta_0) \cdot \frac{1}{n_1}\sum_{i=1}^{n_1}\mathbb{C}(\bar{X}_1, \psi_{i1}) - \dot{g}_0(\theta_1, \theta_0) \cdot \frac{1}{n_0}\sum_{i=1}^{n_0}\mathbb{C}(\bar{X}_0, \psi_{i0}) \\
&= \dot{g}_1(\theta_1, \theta_0) \cdot \frac{1}{n_1^2}\sum_{i=1}^{n_1}\mathbb{C}(X_{i1}, \psi_{i1}) - \dot{g}_0(\theta_1, \theta_0) \cdot \frac{1}{n_0^2}\sum_{i=1}^{n_0}\mathbb{C}(X_{i0}, \psi_{i0}).
\end{aligned}
$$

$\Sigma_{X\Delta}$ may be estimated as:

$$
\hat{\Sigma}_{X\Delta} = \dot{g}_1(\hat{\theta}_1, \hat{\theta}_0) \cdot \frac{1}{n_1^2}\sum_{i=1}^{n_1}(X_{i1} - \hat{\mu}_X)\hat{\psi}_{i1} - \dot{g}_0(\hat{\theta}_1, \hat{\theta}_0) \cdot \frac{1}{n_0^2}\sum_{i=1}^{n_0}(X_{i0} - \hat{\mu}_X)\hat{\psi}_{i0}. \tag{1.3}
$$

Overall, the covariate-adjusted estimator is:

$$
\hat{\Delta}_{\text{adj}} = g(\hat{\theta}_1, \hat{\theta}_0) - \hat{\Sigma}'_{X\Delta}\hat{\Sigma}^{-1}_{XX}(\bar{X}_1 - \bar{X}_0). \tag{1.4}
$$

The asymptotic variance is given by:

$$
\mathbb{V}(\hat{\Delta}_{\text{adj}}) = \Sigma_{\Delta\Delta} - \Sigma'_{X\Delta}\Sigma^{-1}_{XX}\Sigma_{X\Delta}. \tag{1.5}
$$

Here $\Sigma_{\Delta\Delta}$ is the variance of $g(\hat{\theta}_1, \hat{\theta}_0)$:

$$
\Sigma_{\Delta\Delta} = \dot{g}_1^2(\theta_1, \theta_0)\mathbb{V}(\hat{\theta}_1) + \dot{g}_0^2(\theta_1, \theta_0)\mathbb{V}(\hat{\theta}_0) + o_p(1),
$$

which may be estimated by:

$$
\hat{\Sigma}_{\Delta\Delta} = \dot{g}_1^2(\hat{\theta}_1, \hat{\theta}_0) \cdot \frac{1}{n_1^2}\sum_{i=1}^{n_1}\hat{\psi}_{i1}^2 + \dot{g}_0^2(\hat{\theta}_1, \hat{\theta}_0) \cdot \frac{1}{n_0^2}\sum_{i=1}^{n_0}\hat{\psi}_{i0}^2. \tag{1.6}
$$

Summary

The general procedure for nonparametric covariate adjustment via augmentation is:

1. Identify the treatment contrast of interest (1.1).
2. Calculate the influence function contributions of reach subject (ψ_{ik}).
3. Calculate the variance components Σ_{XX} (1.2), $\Sigma_{X\Delta}$ (1.3), $\Sigma_{\Delta\Delta}$ (1.6).
4. Calculate the covariate-adjusted estimator (1.4) and its variance (1.5).

References

1. Jian F, Tian L, Fu H, Hasegawa T, Wei LJ. Robust Alternatives to ANCOVA for Estimating the Treatment Effect via a Randomized Comparative Study. Journal of the American Statistical Association. 2019; 114:528, 1854-1864. doi.org/10.1080/01621459.2018.1527226.