Notes

Xingyu Chen[1]

November 12, 2025

Contents

1 Tpyst symbols

Tpyst symbols.

2 Random effects model for election polls

Andrew Gelman's recent blog post, Polls & Betting Odds & Nonsampling Errors & Win Probabilities & Vote Margins, discusses how to estimate winning probabilities based on polling data. A ChatGPT summary is available here: chatgpt-history.

He presents three examples: the New Jersey governor race, the Virginia governor race, and the New York City mayoral race. Individual polls are often unreliable, so systematic differences between polls must be considered. A simple random-effects model can capture this phenomenon:

$$Y_{i,t} = \theta + \eta_{i,t} + \varepsilon_i$$

Here, $Y_{i,t}$ is the t-th poll for the i-th media, θ is the true underlying support level, ε_i represents the systematic error for poll i, and $\eta_{i,t}$ represents the sampling error of the poll.

When focusing on a single media's polls (fixing $i = i_0$), Gelman uses historical information to estimate ε_t, assuming $\varepsilon_t \sim \mathcal{N}(0, 2\%)$. He then uses polling data of media i_0 to estimate the sample variance $\sigma_{i_0}^2$ of η_{i,i_0}.

Let $\overline{Y_{i_0}} = \frac{1}{n}\sum_{i=1}^{n} Y_{i,i_0}$. Under his Bayesian viewpoint, the posterior distribution of θ is approximated by:

$$\theta \mid Y \sim \mathcal{N}\left(\overline{Y_{i_0}}, \sqrt{\sigma_{i_0}^2 + 2\%}\right)$$

The winning probability can then be computed as:

$$\Pr(\theta > 0.5 \mid Y).$$

3 Writing

3.1 Good title: The unreasonable effectiveness of XX in/of YY

Good title form The unreasonable effectiveness of mathematics in the natural sciences by Eugene Wigner in 1960 [1].

Richard Hamming in computer science, "The Unreasonable Effectiveness of Mathematics".

Arthur Lesk in molecular biology, "The Unreasonable Effectiveness of Mathematics in Molecular Biology"

Peter Norvig in artificial intelligence, "The Unreasonable Effectiveness of Data"

Vela Velupillai in economics, "The Unreasonable Ineffectiveness of Mathematics in Economics"

Terrence Joseph Sejnowski in Artificial Intelligence: The Unreasonable Effectiveness of Deep Learning in Artificial Intelligence".

3.2 Good title: Rebel With a Cause

Form the famous movie Rebel Without a Cause.

The special issue Volume 8, Issue 2, 2022 Issue of *Observational Studies* titleed Rebel With a Cause

4 On the undistinguishable or identification of statistical models

Based on talk with Ruiqi Zhang, Lin Liu and Chatgpt.

The undistinguishable means you can not distinguish two models form data. Which will corresponds to 3 cases.

One and two are both in the modeling stage. When you consider one model P_θ, this corresponds to non-identifiable model. When you consider two models, this corresponds to two models sharing some common distributions.

The third case is in the hypothesis testing stage, two models can not be distinguished by data since they are too close.

In a word, two models $\mathcal{P}_1$ and $\mathcal{P}_2$ as two distribution calsses are undistinguishable if $\mathcal{P}_1 \cap \mathcal{P}_2 \neq \emptyset$ or they are too close under some metric.

4.1 On latent structure models

The potential outcome model is an example of latent structure model. The observed random variable is determined by some unobservable/latent variable is in this calss.

Definition 4.1. *The observed random variable X is determined by a high dimensional latent variable Z by a map $X = f(Z)$.*

***Example* 4.2.** The observed random variable is (A, Y) determined by three latent variable $(A, Y(0), Y(1)), A \in \{0, 1\}, Y = AY(1) + (1 - A)Y(0)$, consider two submodels:

- $\mathcal{P}_1$: $Y(1) - Y(0) = 0, Y(1) \perp A, Y(0) \perp A$
- $\mathcal{P}_2$: $Y(1) - Y(0) = Z \neq 0$, but $Y(1) \overset{d}{=} Y(0), Y(1) \perp A, Y(0) \perp A$.

The model 2 is not empty, take

$$Y(0), \varepsilon \sim \mathcal{N}(0, 1), Y(0) \perp \varepsilon, Z = -\frac{1}{2}Y(0) + \frac{\sqrt{3}}{2}\varepsilon$$

then $Y(1) \sim \mathcal{N}(0, 1)$ and $Y(1) \overset{d}{=} Y(0)$.

On the observed data level, we can not distinguish these two models since they both have the same conditional distribution of $Y|A$, therefore they are undistinguishable in modeling stage.

This example is the reason why the sharp null hypothesis can not be tested in randomized experiment, and also the joint distribution of $(Y(0), Y(1))$ is not identifiable.

[2] talk about the identification of joint distribution of potential outcomes under some assumptions.

5 On the cluster analysis

6 Causal Inference

6.1 Causal Example

6.1.1 Yule–Simpson's Paradox

6.1.1.1 Healty worker effect

6.1.2 Smoke cause lung cancer

6.1.2.1 Mediation analysis

6.1.2.2 Sensitivity analysis

6.2 Identification

6.2.1 G-formula

6.2.2 Joint distribution of potential outcomes

[2] consider a case with multiple randomized controlled trials(RCTs), where data are (G, A, Y), G is the indicator of RCTs, A is the treatment, Y is the outcome.

Under consistency, positivity, and exchangeability. Adding one assumption called "transportability":

$$Y(1) \perp G \mid Y(0)$$

We can then identify the conditional distribution $Y(1) \mid Y(0)$.

$$\begin{aligned}\Pr(Y(1)=b \mid G=g) &= \sum_a \Pr(Y(1)=b, Y(0)=a \mid G=g) \Pr(Y(0)=a \mid G=g) \\ &= \sum_a \Pr(Y(1)=b \mid Y(0)=a) \Pr(Y(0)=a \mid G=g)\end{aligned}$$

Here $\Pr(Y(1)=b \mid G=g)$ and $\Pr(Y(0)=a \mid G=g)$ can be identified form data by the consistency assumption, using them to solve the above equation system, we can identify $\Pr(Y(1)=b \mid Y(0)=a)$.

6.3 The equivalence between DAG and potential outcome framework

6.3.1 The equivalence between nonparametric structural equation model(NPSEM) and potential outcome framework

6.3.2 The equivalence between SWIG and FFRCISTG

6.4 ADMG : the inequility constraints

6.4.1 Bell inequality

6.5 Instrumental variable

[3]

- data are (X, A, Z, Y) only assume consistency, positivity, unconfounder, exculusion, no monotonicity, provide a bound estimation on ATE.
- The identification assumption of IV:

"Critically, under the four assumptions introduced in the previous section, the ATE is not point identified. Analysts typically take one of two approaches for point identification. The first approach invokes some type of homogeneity assumptions and places various restrictions on how the effects of A and Z vary from unit to unit in the study population. See Hernan and Robins (2019) and Wang and Tchetgen Tchetgen (2018) for prominent examples. However, homogeneity assumptions are often implausible or difficult to verify in specific applications. The second approach invokes an assumption known as monotonicity, which has the following form: $A(z = 1) \geq A(z = 0)$, i.e., if $A(z = 0) = 1$ then $A(z = 1) = 1$ (Imbens and Angrist, 1994). Under monotonicity, the target estimand is no longer the ATE, but instead is the local average treatment effect (LATE):"

- The lower bound and upper bound is not a differentiable functional, thus an assumption is invoked to make the bound functional differentiable and thus have inference function to faster convergence rate.

7 Semiparametric theory

7.1 Parametric theory

7.1.1 Efficiency

7.1.1.1 Convolution theorem

7.2 nonparametric theory

7.3 Regularity

7.4 Influence function

7.4.1 Numerical calcaulation of influence function

[4]

7.4.2 Von mise representation

7.4.3 Tangent space

S8 in [5]

Formulation in [6]

7.4.4 Higher order influence function

Bibliography

[1] E. P. Wigner and others, "The unreasonable effectiveness of mathematics in the natural sciences," *Mathematics and science*, vol. 13, pp. 1–14, 1990.

[2] P. Wu and X. Mao, "The Promises of Multiple Experiments: Identifying Joint Distribution of Potential Outcomes," *arXiv preprint arXiv:2504.20470*, 2025.

[3] A. W. Levis, M. Bonvini, Z. Zeng, L. Keele, and E. H. Kennedy, "Covariate-assisted bounds on causal effects with instrumental variables," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, p. qkaf28, 2025.

[4] Y. Mukhin, "Kernel von Mises Formula of the Influence Function," in *The Thirty-ninth Annual Conference on Neural Information Processing Systems*,

[5] E. Graham, M. Carone, and A. Rotnitzky, "Towards a Unified Theory for Semiparametric Data Fusion with Individual-Level Data," *arXiv preprint arXiv:2409.09973*, 2024.

[6] A. van der Vaart, "On differentiable functionals," *The Annals of Statistics*, vol. 19, no. 1, pp. 178–204, 1991.