Experimental Methods: Lecture 8

Social Media and Experimental Design

Raymond Duch

June 16, 2022

University of Oxford

Road Map

- Treatment-adaptive design
- Text and causal inference
- Digital Trace collection

Treatment-adaptive designs

Types of Sequential Randomised Experiments

- Non-adaptive - assignment probabilities fixed
- Treatment-adapative - change based on number of subjects in treatment
- Covariate-adaptive - change based on covariate profiles of new and previous subjects
- Responsive-adaptive - change as function of previous units' outcomes

Treatment-adaptive designs

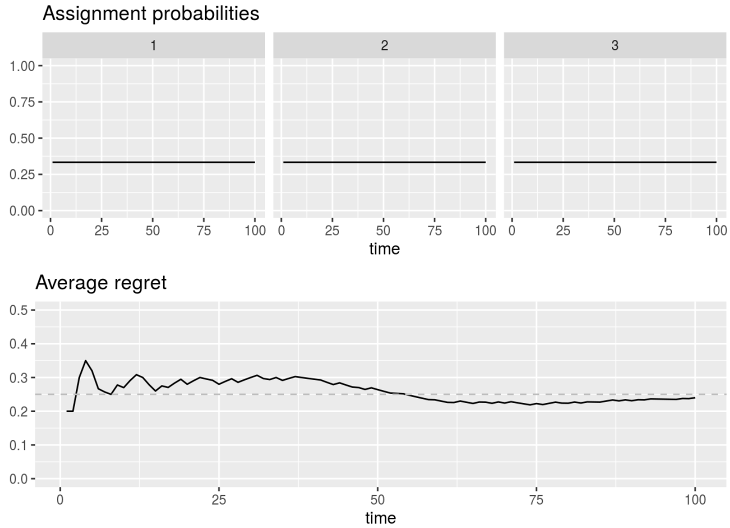
- *ATE* is not always quantity of interest
- Particularly online firms such as Google, Tiktok, FB, etc.
  - Randomly assign sampled users to different arms and dynamically re-orient sample based on which is more successful/more informative
  - Identify which of many will get the most clicks
- But also of interest to political scientists: Ballot initiatives and malfeasance information
- How do adaptive multi-arm trials work?

Regret

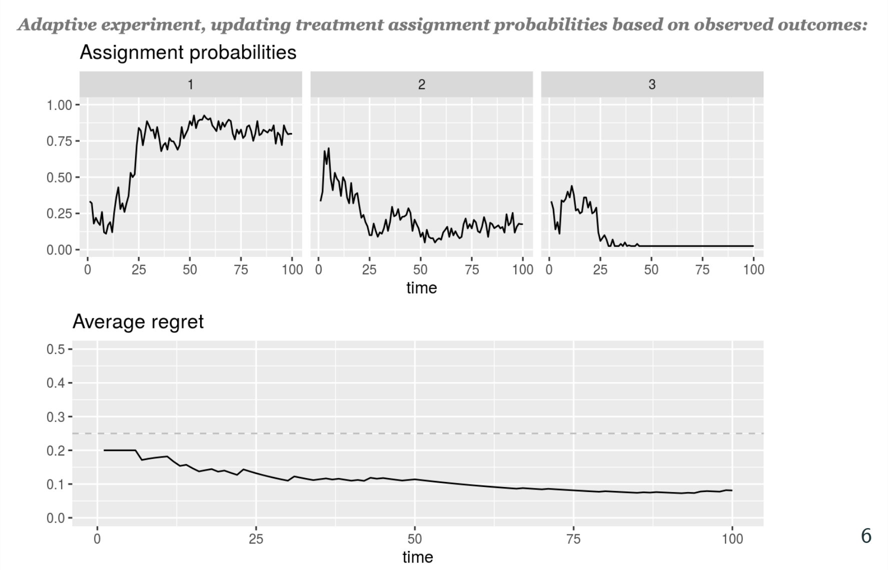
- the difference between the average outcomes we would have observed under optimal assignment and the average outcomes we actually observe under a given assignment algorithm
- Example
  - if the best prototype gives us a 90% click-through rate on average
  - a different prototype gives us a 40% rate on average
  - the regret from assigning the sub-optimal arm is 0.5

Regret: True arms 1 (.8) 2 (.6) 3 (.3)

Random experiment, with balanced assignment probabilities throughout:



Regret: True arms 1 (.8) 2 (.6) 3 (.3)



Treatment-adaptive design: Thompson Sampling

- When researchers are initially agnostic about the relative performance of the K arms, priors are distributed uniformly over parameter space, i.e. $\beta_{1,1}$
- In each t, treatment is randomly assigned according to probability of arms being best (= highest success rate)

$$P\left[\Theta_k = \max_k\{\Theta_1, ..., \Theta_K\} | (X_1^{n_{1,t}}, ..., X_K^{n_{K,t}})\right]$$

for K arms, vector of responses under treatment arm k observed up until and including t $X_k^{\{n_{k,t}\}}$ and Θ_k distributions of success rates

best_binomial_bandit

```
x=c(10,20,30,50)
n=c(100,102,120,130)
arm_probabilities =
best_binomial_bandit(x,n) print(arm_probabilities)
[1] 1.611266e-07 8.048293e-04 1.142867e-02 9.877663e-0

sum(arm_probabilities)
[1] 1
```

Treatment-adaptive design: Thompson Sampling

- Simulations to illustrate design and estimation
- Sample 100 observations for each of 10 periods, updating posterior probability of being best after each period, and assign treatment probabilities in the subsequent period accordingly
- In the first case, one arm has a true 0.20 probability of success, and the remaining 8 arms have a 0.10 probability of success

Treatment-adaptive designs

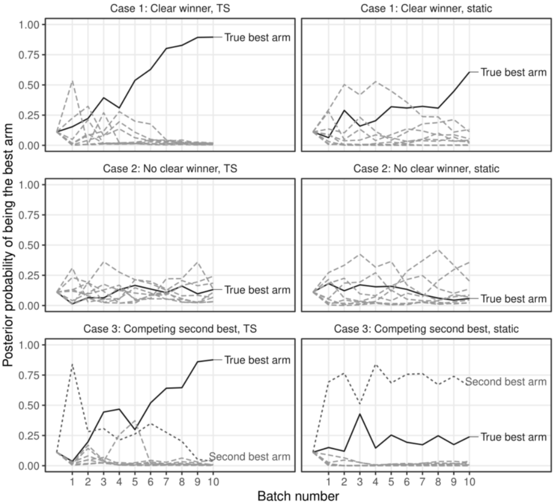
Table 1 Iterated Simulation Statistics

Design			RMSE		Coverage	
Assignment algorithm	**Case**	**Best arm selected**	**Best arm**	**ATE**	**Best arm**	**ATE**
TS	1: Clear winner	0.968	0.021	–	0.958	–
	2: No clear winner	0.193	0.033	–	0.880	–
	3: Competing second best	0.715	0.025	–	0.956	–
Static	1: Clear winner	0.909	0.031	0.038	0.941	0.949
	2: No clear winner	0.180	0.024	0.033	0.935	0.947
	3: Competing second best	0.635	0.031	0.038	0.940	0.945
TS,	1: Clear winner	0.956	0.023	0.029	0.957	0.952
Control-Augmented	2: No clear winner	0.174	0.034	0.041	0.879	0.886
	3: Competing second best	0.683	0.029	0.035	0.946	0.937

Note: Assignment algorithms are Thompson sampling (TS), balanced static design (Static), and control-augmented Thompson sampling (TS, Control-Augmented). "Best arm selected" column presents the portion of simulations under which the true best arm was selected. RMSE is average root mean squared error of the estimate of the mean of the true best arm, and the average treatment effect of the true best arm relative to the control. Coverage is with respect to 95% confidence intervals around the estimate. In all cases one of the inferior arms with a true success rate of 0.10 is selected as the control comparison.

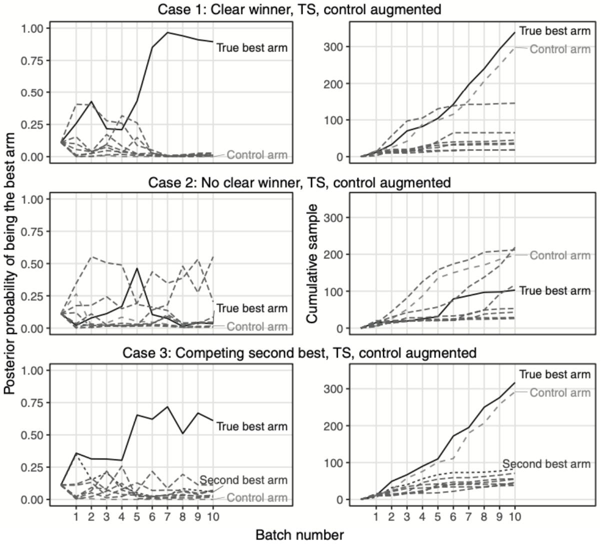
Treatment-adaptive designs

Figure 1: Simulated Posterior Probabilities Over Time, Thompson Sampling and Static Designs



Treatment-adaptive designs

FIGURE 2 Simulated Posterior Probabilities over Time and Cumulative Sample, Control-Augmented Adaptive Design



Molly Offer-Westort et al 2020

Table 2: Study One, Treatments and Outcome Measures

	Minimum Wage	Right to Work
Question Text	Imagine that the following ballot measure were up for a vote in your state. [**ballot measure text**]. If this measure were on the ballot in your state, would you vote in favor or against? [I would vote in favor of this measure; I would vote against this measure]	Imagine that the following ballot measure were up for a vote in your state. [**ballot measure text**]. If this measure were on the ballot in your state, would you vote in favor or against? [I would vote in favor of this measure; I would vote against this measure]
Proposal 1	The measure would: increase the minimum wage [**from {current}**] to {current + 1} per hour, adjusted annually for inflation, and provide that no more than $3.02 per hour in tip income may be used to offset the minimum wage of employees who regularly receive tips.	The measure would [**amend the State Constitution to**]: prohibit, as a condition of employment, forced membership in a labor organization (union) or forced payments of dues or fees, in full or pro-rata ("fair-share"), to a union. The measure will also make any activity which violates employees' rights provided by the bill illegal and ineffective and allow legal remedies for anyone injured as a result of another person violating or threatening to violate those employees' rights. The measure will not apply to union agreements entered into before the effective date of the measure, unless those agreements are amended or renewed after the effective date of the measure.
Proposal 2	The measure would: raise the minimum wage [**from {current}**] to {current + 1} per hour effective September 30th, 2021. Each September 30th thereafter, minimum wage shall increase by $1.00 per hour until the minimum wage reaches {current + 5} per hour on September 30th, 2026. From that point forward, future minimum wage increases shall revert to being adjusted annually for inflation starting September 30th, 2027.	The measure [**reads / would amend the State Constitution to read**]: The right of persons to work may not be denied or abridged on account of membership or nonmembership in any labor union or labor organization, and all contracts in negation or abrogation of such rights are hereby declared to be invalid, void, and unenforceable.
Proposal 3	The measure reads: Shall the minimum wage for adults over the age of 18 be raised [**from {current}**] to {current + 1} per hour by January 1, 2019?	The measure would [**amend the State Constitution to**]: ban any new employment contract that requires employee to resign from or belong to a union, pay union dues, or make other payment to a union. Required contributions to charity or other third party instead of payments to union are also banned. Employees must authorize payroll deduction to unions. Violations of the section is a misdemeanor.
Proposal 4	The measure would: raise the minimum wage [**from {current}**] to {current + 1} per hour worked if the employer provides health benefits, or {current + 2} per hour worked if the employer does not provide health benefits.	The measure [**reads / would amend the State Constitution to read**]: No person shall be deprived of life, liberty or property without due process of law. The right of persons to work shall not be denied or abridged on account of membership or nonmembership in any labor union, or labor organization.
Proposal 5	The measure would: raise the State minimum wage rate [**from {current}**] to at least {current + 1} per hour, and require annual increases in that rate if there are annual increases in the cost of living.	

Boldface text indicates randomly varied elements.

Ballot Initiative Ex. from Offer-Westort et al 2020

scifile.tex × | rankings.R × | Chile Corruption MRP.Rmd × | Corona_Italy_cleaning.R × | Analysis.R × | Corona_italy_anal

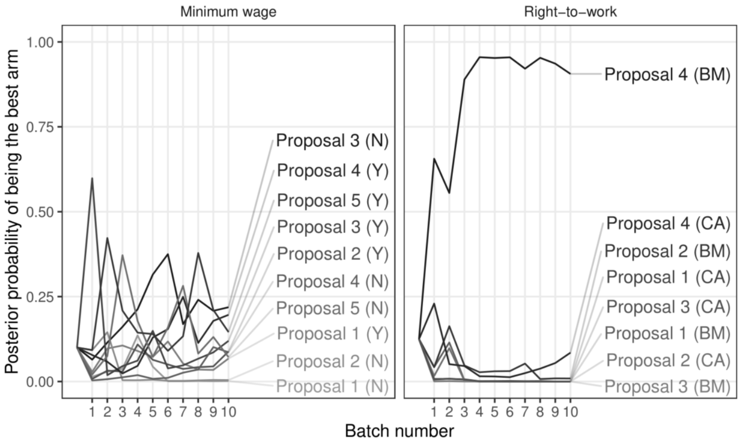
Filter

	id	Y_mw	Y_rtw	Z_mw	Z_rtw	weights_rtw	prob_rtw	weights_mw	prob_mw	batch
1	1	1	1	Proposal 4 (N)	Proposal 2 (CA)	8.000000	0.12500	10.000000	0.10000	1
2	2	1	1	Proposal 4 (N)	Proposal 4 (BM)	8.000000	0.12500	10.000000	0.10000	1
3	3	1	0	Proposal 3 (N)	Proposal 1 (CA)	8.000000	0.12500	10.000000	0.10000	1
4	4	1	0	Proposal 2 (N)	Proposal 3 (BM)	8.000000	0.12500	10.000000	0.10000	1
5	5	1	1	Proposal 2 (N)	Proposal 2 (BM)	8.000000	0.12500	10.000000	0.10000	1
6	6	1	1	Proposal 3 (N)	Proposal 4 (BM)	8.000000	0.12500	10.000000	0.10000	1
7	7	1	1	Proposal 1 (N)	Proposal 3 (BM)	8.000000	0.12500	10.000000	0.10000	1
8	8	1	1	Proposal 2 (Y)	Proposal 3 (BM)	8.000000	0.12500	10.000000	0.10000	1
9	9	1	1	Proposal 4 (Y)	Proposal 3 (BM)	8.000000	0.12500	10.000000	0.10000	1
10	10	1	1	Proposal 3 (N)	Proposal 1 (CA)	8.000000	0.12500	10.000000	0.10000	1
11	11	1	1	Proposal 3 (Y)	Proposal 1 (CA)	8.000000	0.12500	10.000000	0.10000	1
12	12	1	1	Proposal 5 (Y)	Proposal 4 (CA)	8.000000	0.12500	10.000000	0.10000	1
13	13	0	1	Proposal 5 (Y)	Proposal 2 (BM)	8.000000	0.12500	10.000000	0.10000	1
14	14	1	1	Proposal 5 (Y)	Proposal 3 (CA)	8.000000	0.12500	10.000000	0.10000	1
15	15	1	1	Proposal 1 (Y)	Proposal 1 (BM)	8.000000	0.12500	10.000000	0.10000	1
16	16	1	1	Proposal 3 (Y)	Proposal 4 (CA)	8.000000	0.12500	10.000000	0.10000	1
17	17	1	0	Proposal 1 (Y)	Proposal 2 (CA)	8.000000	0.12500	10.000000	0.10000	1
18	18	1	1	Proposal 5 (Y)	Proposal 1 (BM)	8.000000	0.12500	10.000000	0.10000	1

Showing 1 to 18 of 1,000 entries, 10 total columns

Treatment-adaptive designs

Figure 3: Study One, Overtime Posterior Probabilities



Adaptive experimentation tutorial

- Molly Offer-Westort, Vitor Hadad, Susan Athey
- `https://mollyow.shinyapps.io/adaptive/`

Causal Inference with Text (Fong and Grimmer 2021

Lottery versus Triage Vignette

Le Pen : "Je

Lottery Vignette

Q14.1.

Consider the following situation: a major clinic has developed plans for allocating limited supplies of the COVID-19 vaccine. It would like to vaccinate all 1000 nurses who work in the clinic. There will only be 500 vaccines available for the 1000 nurses who work in the clinic. Because of the limited supply, the vaccine will be allocated by a **lottery**. The names of the 1000 nurses will be put into a large container and shuffled. 500 names will be randomly selected from the container. These 500 randomly selected names will receive the vaccine.

Do you agree or disagree that this **lottery** method is an appropriate way to allocate the scarce vaccines to the nurses? Use the slider to indicate whether you disagree or agree: 0 means strongly disagree and 100 means strongly agree.

Strongly Disagree Strongly Agree

0 10 20 30 40 50 60 70 80 90 100

Lottery Allocation?

Lottery versus Triage Vignette

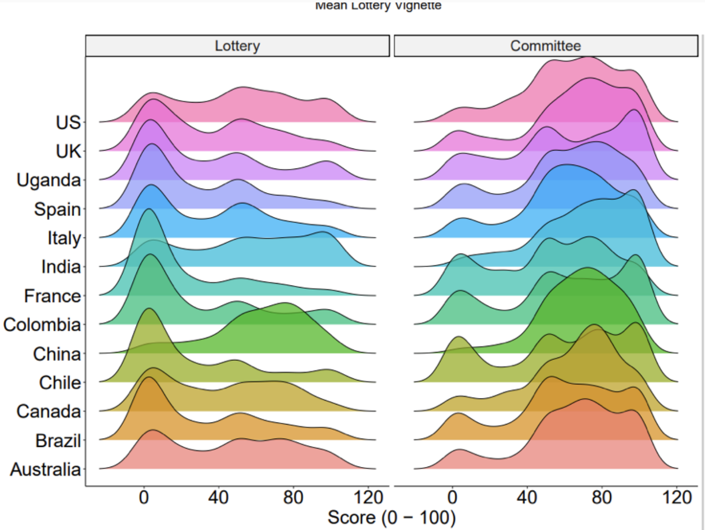
Q481.

Consider the following situation: a major clinic has developed plans for allocating limited supplies of the COVID-19 vaccine. It would like to vaccinate all 1000 nurses who work in the clinic. There will only be 500 vaccines available for the 1000 nurses who work in the clinic. Because of the limited supply, the vaccine will be allocated by an **independent committee of expert physicians**, who have no personal connection with the clinic and do not know any of the nurses involved. The medical histories of the 1000 nurses will be provided to the committee of experts. The committee of expert physicians will select the 500 nurses who they believe will most benefit from the vaccine.

Do you agree or disagree that this **expert committee** method is an appropriate way to allocate the scarce vaccines to the nurses? Use the slider to indicate whether you disagree or agree: 0 means strongly disagree and 100 means strongly agree.

	Strongly Disagree										Strongly Agree	
	0	10	20	30	40	50	60	70	80	90	100	
Expert Committee Allocation?												

Lottery versus Triage Vignette



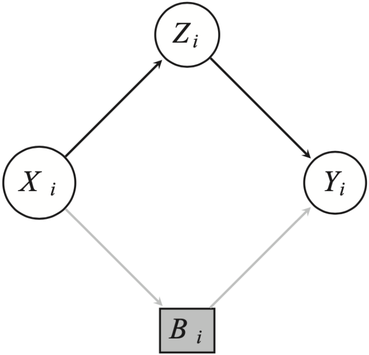
A Common Problem

Table 1 Extensive Use of Latent Treatments

Experiment type	Count	Latent	Aliased	Single vignette	Both
Survey experiment	29	100%	66%	97%	62%
Field experiment	13	92	54	77	46
Conjoint experiment	5	100	0	100	0
Lab experiment	4	100	75	100	75

Acyclit

FIGURE 1 Directed Acyclic Graph for Causal Text Diagram



Note: The text, $\boldsymbol{X}_i$, causes both the latent treatment of interest, Z_i, and the unmeasured latent treatments, $\boldsymbol{B}_i$. These latent treatments, in turn, cause the outcome, Y_i.

Hong Kong

TABLE 4 Hong Kong Experiment Treatments

	December 2019	October 2020
Intercept	64.23	69.03
	(3.14)	(1.07)
Commitment	5.23	2.68
	(1.74)	(1.23)
Bravery	−0.72	1.85
	(1.82)	(1.38)
Mistreatment	0.97	0.14
	(1.77)	(1.39)
Flags	0.04	−2.12
	(1.81)	(1.41)
Threat	−2.50	−2.07
	(1.86)	(1.36)
Economy	−0.44	−0.94
	(1.84)	(1.35)
Violation	−0.98	0.75
	(1.81)	(1.38)
N	1,983	2,072

Note: Results come from a linear model, in which the outcome is the degree to which the respondent agrees the U.S. government should help Hong Kong. The left column refers to the original experiment and the right column refers to a replication experiment.

Digital Trace and Machine Learning with Post-Stratification to Measure Public Opinion (Cerina and Duch 2020)

Digital Vote Overview

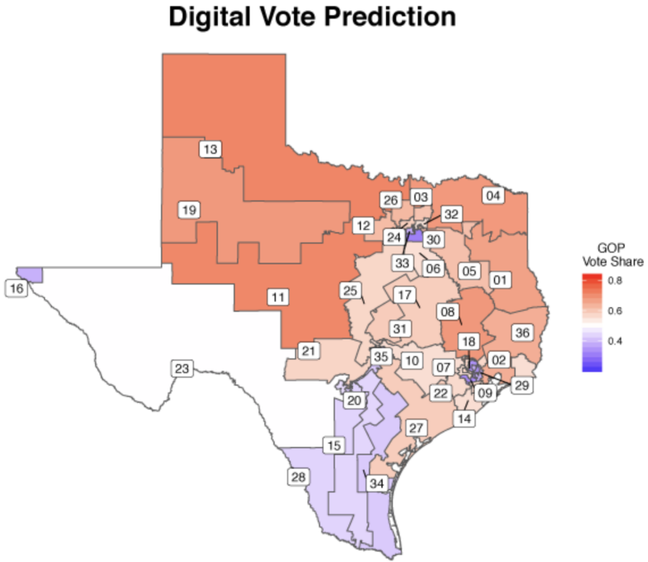
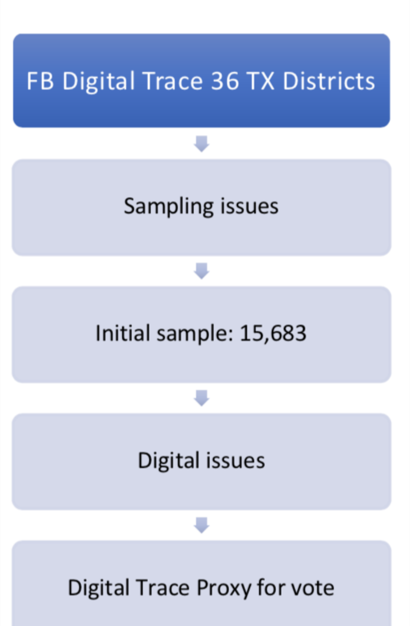
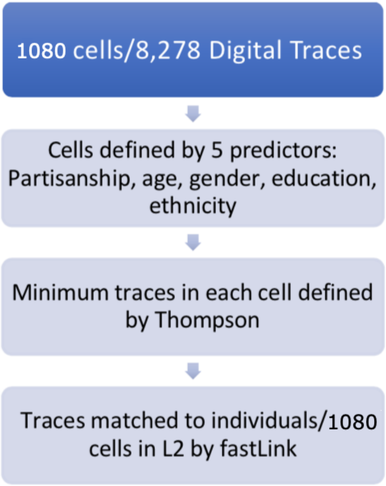


Figure 1: Digital Vote predicted GOP Percent Vote Share.

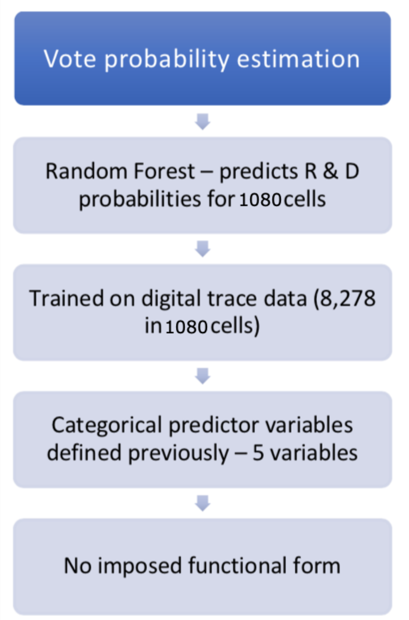
Digital Vote Overview



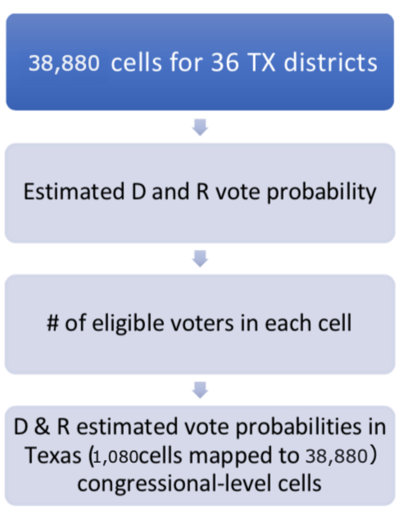
Digital Vote Overview



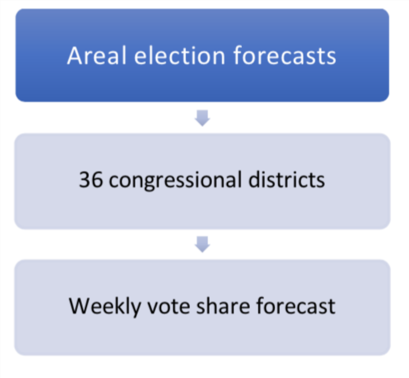
Digital Vote Overview



Digital Vote Overview



Digital Vote Overview



Sampling

Sampling

- Tune sample-size to capture differences in prevalence of voting-groups;

- theoretical sampling distribution:

$$\boldsymbol{n} \sim \text{Multinomial}(p_{1,r=1}, \ldots, p_{G,r=1}, p_{1,r=0}, \ldots, p_{G,r=0}, N); \quad (1)$$

- *worst-case-scenario* sampling for a multinomial distribution following the recommendations of Thompson[Thompson, 1987];

- digital sample is non-probability, but we use theoretical probability distribution as benchmark to guide our power calculation.

Sampling

- find N such that we have probability of at least 0.9 that all estimates of the multinational parameters are within 0.025 of the population proportions;
- ensures sample will contain representatives from groups which are 2.5% of voting population or higher;
- $N = 1610$ per week[Thompson, 1987]

Facebook Sampling

Sampling

- monitor 7 *weeks-to-election-day*;
- assume each multinomial parameter is time-independent;
- sample N for every week within our monitoring frame, for a total of $N = 11270$;
- effective weekly sample size must be further inflated to account for sample loss by poor matching with the voter registry;
- parameter independence is likely true only for swing-voters who are affected by the election-cycle - too harsh for other categories.

Matching

Matching

- get rich voter profiles by matching our social media sample to the state-wide Texas L2 voter registration file;
- individual-level data for over 13 million registered Texas residents, ranging from voting history to socio-economic and demographic characteristics;
- use `R` package `fastLink`[Enamorado et al., 2018] to match our virtual sample to the voter registration list, based on names, *sex* and *city of residency*.

Matching

Threshold	0.75	0.85	0.9	0.95	0.99
Match Count	8739	8394	8384	8283	7322
Match Rate (%)	54.59	52.89	52.84	52.24	46.31
FDR (%)	2.03	1.18	1.16	1.08	0.81

Table 1: `fastLink` output summary table.

- From $15,683$ digital traces in the sample, we successfully match $8,278$;
- success rate of 52.8%;
- these come from $4,475$ distinct registered users;
- `fastlink` reveals $1,417$ borderline cases.

Matching

Conditional on being a registered partisan:

- $\mathrm{P(Dig.\ Rep.|Reg.\ Rep.)} = 0.82$
- $\mathrm{P(Dig.\ Dem.|Reg.\ Dem.)} = 0.82$

Conditional on having attended a partisan primary:

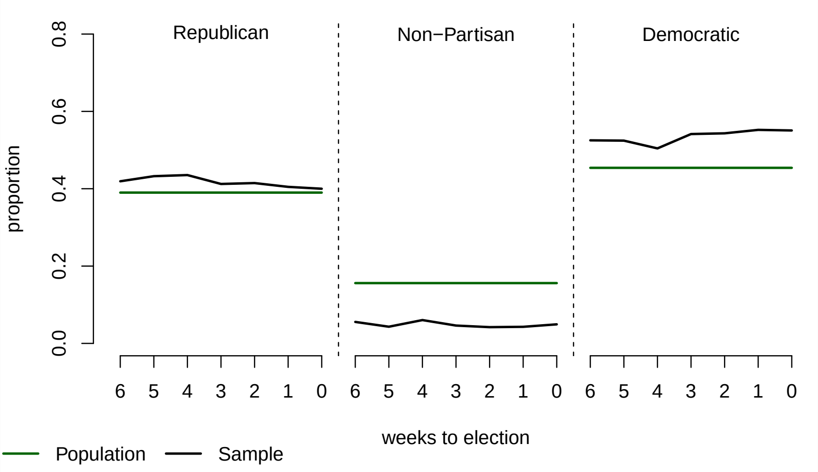
- $\mathrm{P(Dig.\ Rep.|Rep.\ Primary)} = 0.86$
- $\mathrm{P(Dig.\ Dem.|Dem.\ Primary)} = 0.91$

We are comfortable concluding that digital traces represent revealed preferences.

Matching

- final sample size is $8,278$ digital traces;
- slightly below the Thompson number;
- holding probability constant at 0.9, powered to estimate voter categories larger than 2.9% percent of the total population.

Sample v. Population



Vote-Share Estimation

Vote-Share Estimation

Three steps to obtaining area-estimates of vote-share:

- identification of voter characteristics from L2 file to use in vote choice prediction model (voter-categories);
- estimation, at the voter-category level, of vote choice conditional on the probability of vote turnout;
- weighting predictions by cell counts and summing over the area of interest to recover estimates of support.

Vote-Share Estimation

- select just under 50 variables from VR file for turnout prediction at the individual level;
- Missing values in $\boldsymbol{X}$ are imputed with a random-forest multiple-imputation strategy implemented via the packages `ranger`[Wright and Ziegler, 2015] and `missForest`[Stekhoven and Bühlmann, 2011];
- further select subset of variables for vote-choice model Z - we are not powered to include all the variables as we do for turnout;
- define the voter category C_g, for categories $g = 1, ..., G$ as a unique realization of the set of variables which compose $\boldsymbol{Z}$, i.e. $C_g = \{Z_1 = z_1, ..., Z_m = z_m\}$.

Vote-Share Estimation

- Estimate the joint probability of a voter-category supporting the Republican candidate and turning out on election-day;
- decompose the problem similarly to Lauderdale et al.[Lauderdale et al., 2017]:

$$\mathrm{P}_g\left(R=1, T=1|C\right)=\mathrm{P}_g\left(R=1|T=1, C\right)\times\mathrm{P}_g\left(T=1|C\right) \tag{2}$$

.

Vote-Share Estimation

- train a random forest on the expanded (long) turnout dataset $\boldsymbol{X}^{+}$;

- output a prediction probability of turning out for each member of the voting population:

$$\hat{\mathrm{P}}_h\left(T = 1|\boldsymbol{x}\right) = \varphi^T\left(\boldsymbol{x}_h\right); \tag{3}$$

- 1) extract turnout probabilities from matched sample $\hat{\mathrm{P}}_s\left(T = 1|\boldsymbol{x}\right)$;

- 2) average within voter categories identified by $\boldsymbol{C}$ to obtain category-level estimates of turnout probabilities:

$$\mathrm{P}_g\left(T = 1|C\right). \tag{4}$$

Vote-Share Estimation

- calculate global empirical error distribution using the MSPE1 procedure[Lu, 2017];
- Normal distribution does not characterize the empirical distribution perfectly; it is still useful to obtain reasonable prediction intervals:

$$\mathrm{P}_{h}\left(T = 1|\boldsymbol{x}\right) \sim \mathrm{N}\left(\varphi^{T}\left(\boldsymbol{x}_{h}^{\mathsf{I}}\right), (\hat{\sigma}^{T}_{\mathrm{RMSE1}})^{2}\right). \tag{5}$$

Likelihood Estimation

Likelihood Estimation

- train a probability machine as above to estimate the probability of voting Republican, conditional on the individual turning out and the set of their voter characteristics:

$$\hat{\mathrm{P}}_s\left(R=1|T=1,\boldsymbol{z}\right)=\varphi^R\left(\boldsymbol{z}_s|\hat{\mathrm{P}}_s\left(T=1|\boldsymbol{x}\right)\right); \tag{6}$$

- output category-level predictions; for category g such that $s \in g$ if $\boldsymbol{z}_s = C_g$:

$$\hat{\mathrm{P}}_g\left(R=1|T=1,C\right)=\varphi^R\left(C_g\right). \tag{7}$$

- estimate the global error via MSPE1 and approximate distribution with Normal density:

Area Estimation

Area Estimation

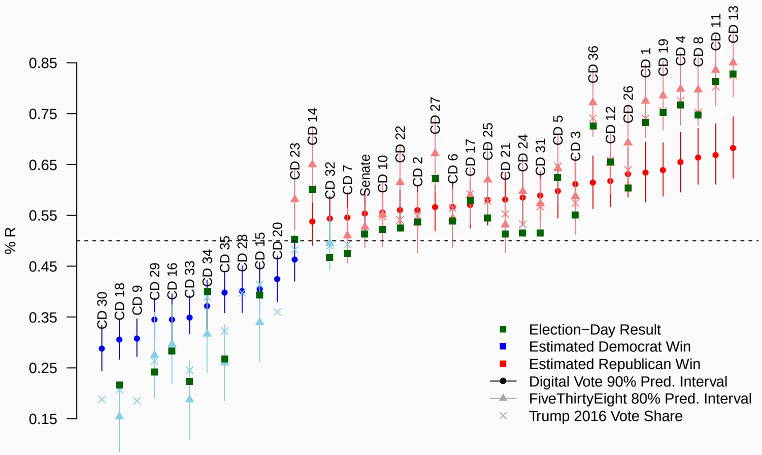
Senate Election (similar for Congress):

- State cell counts: $Q_g = \sum_h 1(\boldsymbol{z}_h = C_g)$;
- Republican support Senate estimate:

$$V_{gw}^{R} = \frac{\sum_g \mathrm{P}_g \left(R = 1, T = 1 | C, L = 1, W = w\right) \times Q_g}{\sum_g \mathrm{P}_g \left(T = 1 | C, L = 1, W = w\right) \times Q_g}; \tag{9}$$

Results

Results

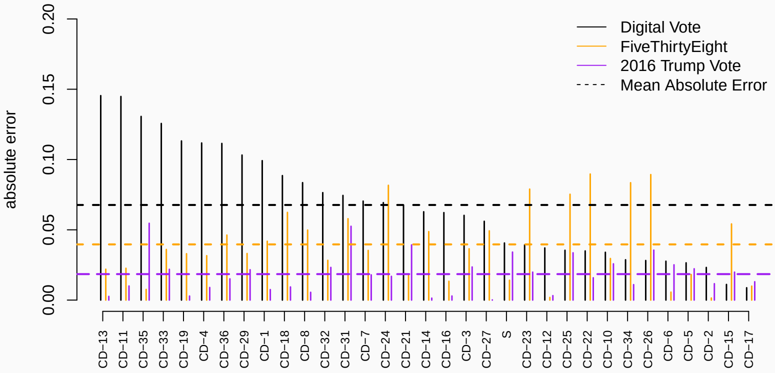


0.85
0.75
0.65
0.55
0.45
0.35
0.25
0.15
% R
CD 30
CD 18
CD 9
CD 29
CD 16
CD 33
CD 34
CD 35
CD 28
CD 15
CD 20
CD 23
CD 14
CD 32
CD 7
Senate
CD 10
CD 22
CD 2
CD 27
CD 6
CD 17
CD 25
CD 21
CD 24
CD 31
CD 5
CD 3
CD 36
CD 12
CD 26
CD 1
CD 19
CD 4
CD 8
CD 11
CD 13
Election−Day Result
Estimated Democrat Win
Estimated Republican Win
Digital Vote 90% Pred. Interval
FiveThirtyEight 80% Pred. Interval
Trump 2016 Vote Share

Results

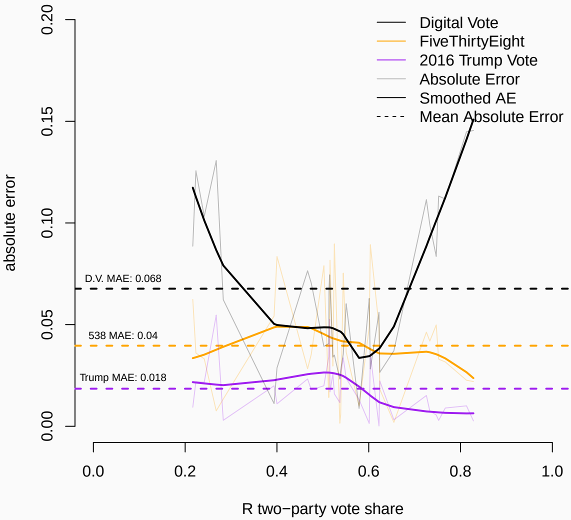
- we produce ‘hypothetical’ estimates for all districts;
- no race-specific effects due to systematic non-response;
- there is evident attenuation bias;
- directionally, our estimates are correct for all but 3 districts (7 and 32 and 23), and correct for the senate;
- on absolute error `FiveThirtyEight` outperforms.

Results



absolute error
0.00
0.05
0.10
0.15
0.20
Digital Vote
FiveThirtyEight
2016 Trump Vote
Mean Absolute Error
CD−13
CD−11
CD−35
CD−33
CD−19
CD−4
CD−36
CD−29
CD−1
CD−18
CD−8
CD−32
CD−31
CD−7
CD−24
CD−21
CD−14
CD−16
CD−3
CD−27
S
CD−23
CD−12
CD−25
CD−22
CD−10
CD−34
CD−26
CD−6
CD−5
CD−2
CD−15
CD−17

Results


Digital Vote
FiveThirtyEight
2016 Trump Vote
Absolute Error
Smoothed AE
Mean Absolute Error
D.V. MAE: 0.068
538 MAE: 0.04
Trump MAE: 0.018
absolute error
0.00
0.05
0.10
0.15
0.20
0.0
0.2
0.4
0.6
0.8
1.0
R two−party vote share

Sub-Category Predictions

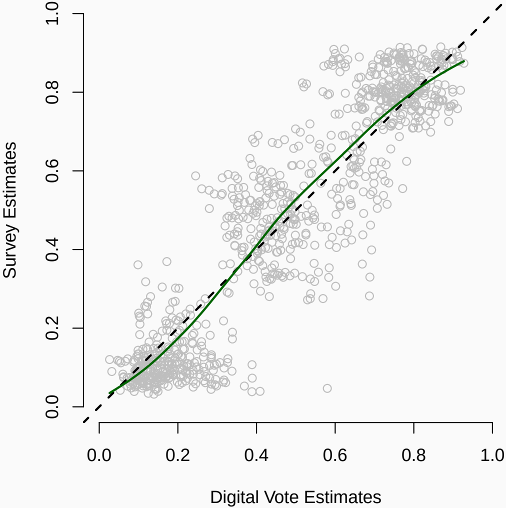


Figure 1: Voter sub-categories comparisons: Surveys v. Digital V...

Conclusions

Conclusions

- our method is a novel variation on MRP, which exploits Random Forests to gain generalizability (via capturing non-linear relationships) and computing speed;
- our sampling strategy suggests careful sampling on social media with respect to politics can be quite representative of the population of interest;
- digital traces are shown to be highly correlated with partisanship, and hence valid as measures of political support.

Social Media Treatment + Behavioral Outcomes

Online Political Information [Enrique et al., 2019]

- Outcome: Vote for the incumbent party in Mexico municipal elections
- Treatment: Online malfeasance video campaign
- Design: Saturation

Information Treatment via Social Media

- 61 million person experiment [Bond et al., 2012]
- Twitter India Vaccines: [Alatas et al., 2019]
- Information Architecture Malfeasance: [Duch and Torres, 2021]
- Vaccine Compliance and Incentives: (Duch et al 2021)

[Enrique et al., 2019] Design: Saturation

- Randomly assign municipalities to 20% % 80% saturation
- Block random assignment:
    - 128 Municipalities
    - 42 Blocks – 3 municipalities in each block have similar incumbent governing parties
- Within each block the municipalities are randomly assigned to three treatments:
    - Control: No FB ad
    - Low saturation: FB information ad
    - High saturation: FB information ad
- Within each municipality:
    - 5 equally-populated segments
    - 20% saturation - 1 in 5 assigned FB info
    - 80% saturation - 4 in 5 assigned FB info

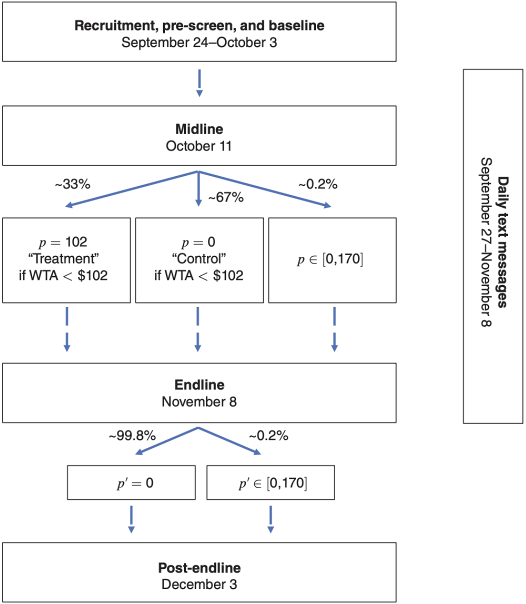
Design: Saturation

- Estimation issues: [Baird et al., 2018]
- Twitter India Vaccines: [Alatas et al., 2019]
- Information Architecture Malfeasance: [Duch and Torres, 2021]
- Vaccine Compliance and Incentives: (Duch et al 2021)

Welfare Effects of FB [Allcott et al., 2020]

- Outcome: WTA
- Treatment: FB Deactivation
- Design: Online recruitment and random assignment to T and C

Welfare Effects of FB [Allcott et al., 2020]

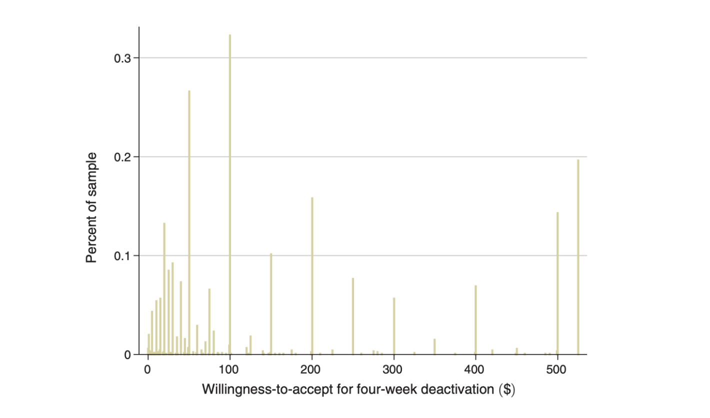


Welfare Effects of FB [Allcott et al., 2020]

TABLE 1—SAMPLE SIZES

Phase	Sample size
Recruitment and baseline	N = 1,892,191 were shown ads N = 32,201 clicked on ads N = 22,324 completed pre-screen survey N = 20,959 were from United States and born between 1900 and 2000 N = 17,335 had 15 < daily Facebook minutes ≤ 600 N = 7,455 consented to participate N = 3,910 finished baseline N = 2,897 had valid baseline and were randomized, of which:
Midline	N = 2,897 began midline N = 2,743 received a price offer, of which: N = 1,661 were in impact evaluation sample
Endline	N = 2,710 began endline N = 2,684 finished endline, of which: N = 1,637 were in impact evaluation sample
Post-endline	N = 2,067 reported Facebook mobile app use, of which: N = 1,219 were in impact evaluation sample

Welfare Effects of FB [Allcott et al., 2020]



Welfare Effects of FB [Allcott et al., 2020]

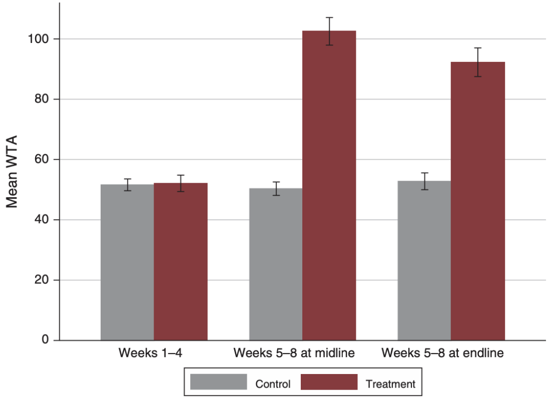
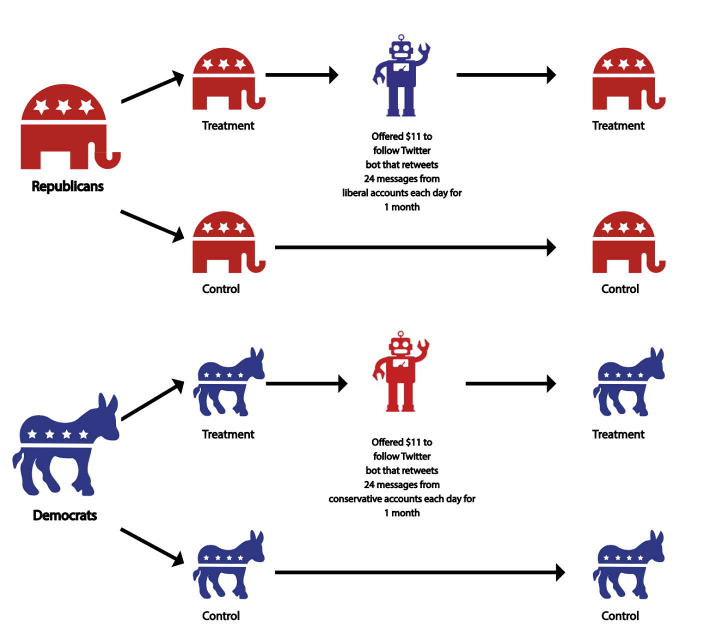


FIGURE 12. AVERAGE VALUATION OF FACEBOOK IN TREATMENT AND CONTROL

Embedded Experiments

Polarization Twitter [Bail et al., 2018]



Embedded Experiments: Examples

- [Coppock et al., 2016]; When Treatments are Tweets
- our sampling strategy suggests careful sampling on social media with respect to politics can be quite representative of the population of interest;

Treatment Delivery: WhatsApp

Twitter Data Collection

References i

Alatas, V., Chandrasekhar, A. G., Mobius, M., Olken, B. A., and Paladines, C. (2019).
When celebrities speak: A nationwide twitter experiment promoting vaccination in indonesia.
Working Paper 25589, National Bureau of Economic Research.

Allcott, H., Braghieri, L., Eichmeyer, S., and Gentzkow, M. (2020).
The welfare effects of social media.
American Economic Review, 110(3):629–76.

References ii

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., and Volfovsky, A. (2018).
Exposure to opposing views on social media can increase political polarization.
Proceedings of the National Academy of Sciences, 115(37):9216–9221.

References iii

Baird, S., Bohren, J. A., McIntosh, C., and Ozler, B. (2018).
Optimal design of experiments in the presence of interference.
The Review of Economics and Statistics, 100(5):844–860.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., and Fowler, J. H. (2012).
A 61-million-person experiment in social influence and political mobilization.
Nature, 489:295 EP –.

References iv

- Coppock, A., Guess, A., and Ternovski, J. (2016).
  When treatments are tweets: A network mobilization experiment over twitter.
  Political Behavior, 38(1):105–128.
- Duch, R. and Torres, F. (2021).
  Information architecture for corruption messaging.
  Working Paper.
- Enamorado, T., Fifield, B., and Imai, K. (2018).
  Using a probabilistic model to assist merging of large-scale administrative records.

References v

- Enrique, J. R., Larreguy, H., Marshall, J., and Simpser, A. (2019).
  Online political information: Facebook ad saturation and electoral accountability in mexico.
  Working paper.
- Lauderdale, B. E., Bailey, D., Blumenau, Y. J., and Rivers, D. (2017).
  Model-based pre-election polling for national and sub-national outcomes in the us and uk.
  Technical report, Working paper.

References vi

Lu, B. (2017).
Constructing Prediction Intervals for Random Forests.
PhD thesis, Pomona College.

Stekhoven, D. J. and Bühlmann, P. (2011).
Missforest—non-parametric missing value imputation for mixed-type data.
Bioinformatics, 28(1):112–118.

Thompson, S. K. (1987).
Sample size for estimating multinomial proportions.
The American Statistician, 41(1):42–46.

References vii

Wright, M. N. and Ziegler, A. (2015).
Ranger: a fast implementation of random forests for high dimensional data in c++ and r.
arXiv preprint arXiv:1508.04409.