A Calibrated Bayes perspective, applied to adaptive survey design

Rod Little

UNIVERSITY OF MICHIGAN

Watch Out--

Scary Bayesians are coming…

I see you are “dipping your toes in the Bayesian waters”

…with Bayesian models for response propensity…

… but wait! Your analysis is still “design-based”

To get the most out of Bayes, you need to jump into the Bayesian ocean:

…base the entire inference for survey quantities on Bayesian models

Overview

- Design-based versus model-based survey inference
- Calibrated Bayes
- Some thoughts on Bayes and adaptive design

Survey estimation

- *Design-based* inference: population values are fixed, inference is based on probability distribution of sample selection. Obviously this assumes that we have a probability sample (or "quasi-randomization", where we pretend that we have one)
- *Model-based* inference: survey variables are assumed to come from a statistical model
- Probability sampling is not the basis for inference, but is useful for making the sample selection *ignorable*. (see e.g. Gelman et al., 2003; Little 2004)

Design vs model-based survey inference

- Two main variants of model-based inference:
  - *Superpopulation models*: Frequentist inference based on repeated samples from a "superpopulation" model
  - *Bayes*: add prior distribution for parameters; inference about finite population quantities or parameters based on posterior distribution
- A fascinating part of the more general debate about frequentist versus Bayesian inference in statistics at large:
  - Design-based inference is inherently frequentist
  - Purest form of model-based inference is Bayes

Limitations of design-based approach

- Inference is based on probability sampling, but true probability samples are harder and harder to come by:
  - Noncontact, nonresponse is increasing
  - Face-to-face interviews increasingly expensive
  - Can't do "big data" (e.g. internet, administrative data) from the design-based perspective
- Theory is basically asymptotic -- limited tools for small samples, e.g. small area estimation

Design-Based Approach Has Implicit Models

- Although not explicitly model-based, models are needed to motivate the choice of estimator
  - E.g. the HT estimator assumes an implicit HT model that y_i / π_i are "exchangeable" (iid conditional on parameters)
  - If implicit models are unreasonable, then the resulting inferences can be very poor in moderate samples (Basu's elephant being an extreme case)
- So models occur in design-based approach, as in the "model-assisted" paradigm

"Quasi"design-based inference

- Key feature of design-based approach is weights, inversely proportional to prob of inclusion
- Weights for selection, nonresponse, poststratification
- Modeling the inclusion propensities, using frequentist or Bayesian methods, leads to weights that are less variable, potentially increasing precision
- Inference remains essentially design-based – in my view; a full Bayesian analysis involves models for the survey variables
- Need terms to codify this distinction: maybe <u>weight modeling</u> and <u>prediction modeling</u>

Model-based approaches

- In *model-based*, or *model-dependent*, approaches, models are the basis for the entire inference: estimator, standard error, interval estimation
- Two main variants:
  - Superpopulation modeling
  - Bayesian (full probability) modeling
- Common theme is to predict non-sampled and nonresponding portion of the population, conditional on the sample and model
- Superpopulation models are super, but Bayes is better!

Parametric models

Usually prior distribution is specified via *parametric* models:

$$p(Y \mid Z) = \int p(Y \mid Z, \theta)\, p(\theta \mid Z)\, d\theta$$

$p(Y \mid Z, \theta)$ = parametric model, as in superpopulation approach

$p(\theta \mid Z)$ = prior distribution for θ

Inference about θ is then obtained from its posterior distribution, computed via Bayes' Theorem:

$$p(\theta \mid Y_{\text{inc}}, Z) =\propto p(\theta \mid Z) \times L(\theta \mid Y_{\text{inc}}, Z)$$

$$L(\theta \mid Y_{\text{inc}}, Z) = \text{Likelihood function}$$

That is: Posterior = Prior x Likelihood…

Posterior for θ leads to inference about population quantities by posterior predictive distribution

The model-based perspective- pros

- Flexible, unified approach for all survey problems
  - Models for nonresponse, response and matching errors, small area models, combining data sources, big data
  - Causal inference requires models
- Bayesian approach is not asymptotic, provides better small-sample inferences
- Probability sampling is justified as making sampling mechanism ignorable, improving robustness
  - Rubin's theory on ignorable selection/nonresponse is the right framework for assessing non-probability samples

The model-based perspective- cons

- Explicit dependence on the choice of model, which has subjective elements (but assumptions are explicit)
- Bad models provide bad answers – justifiable concerns about the effect of model misspecification
- Models are needed for all survey variables – need to understand the data, and potential for more complex computations
- Infrastructure: need personnel trained in statistical modeling

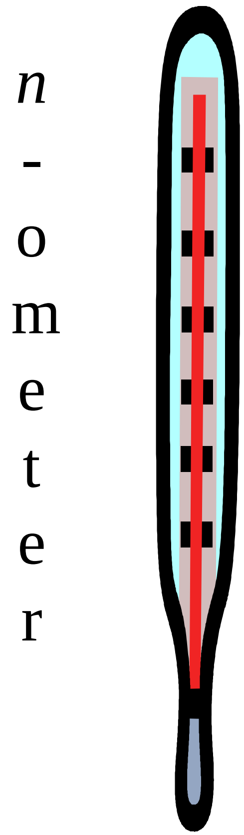
The current "status quo" -- design-model compromise

- Design-based for large samples, descriptive statistics
  - But may be *model assisted*, e.g. regression calibration:

$$\hat{T}_{\text{GREG}} = \sum_{i=1}^{N} \hat{y}_i + \sum_{i=1}^{N} I_i (y_i - \hat{y}_i) / \pi_i, \hat{y}_i = \text{ model prediction}$$

  - model estimates adjusted to protect against misspecification, (e.g. Särndal, Swensson and Wretman 1992).
- Model-based for small area estimation, nonresponse, time series,…
- Attempts to capitalize on best features of both paradigms… but … at the expense of "inferential schizophrenia" (Little 2012)?

Example: when is an area “s<u>mall</u>”?



Design-based inference

↕ -- ← n_0 = “Point of inferential schizophrenia”

Model-based inference

How do I choose n_0?

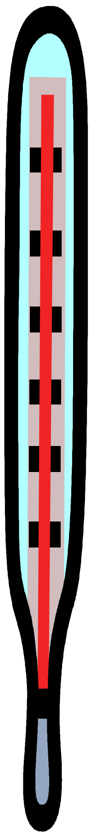
If $n_0 = 35$, should my entire statistical philosophy and inference be different when n=34 and n=36?

n=36, CI: [ ] (wider since based on direct estimate)

n=34, CI: [ ] (narrower since based on model)

Multilevel (hierarchical Bayes) models

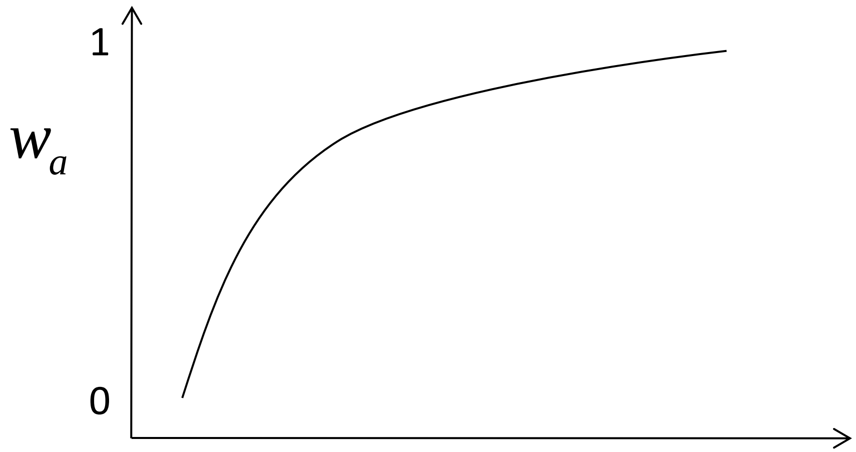
n-ometer



$$\tilde{\mu}_a = w_a \bar{y}_{\pi a} + (1 - w_a)\hat{\mu}_a$$

Model estimate

Direct estimate



Sample size *n*

Bayesian multilevel model estimates borrow strength increasingly from model as *n* decreases

Calibrated Bayes

- Frequentists should be Bayesian
  - Bayes is optimal under a correctly specified model
- Bayesians should be frequentist
  - We never know the model (and all models are wrong)
  - Inferences should be robust to misspecification, have good repeated sampling characteristics
- Calibrated Bayes (Box 1980, Rubin 1984, Little 2006, 2012, 2013)
  - Inference based on a Bayesian model
  - Model chosen to yield inferences that are well-calibrated in a frequentist sense
  - Aim for posterior credibility intervals that have (approximately) nominal frequentist coverage

Calibrated Bayes models for surveys should incorporate sample design features

- The "Calibrated" part of Calibrated Bayes implies:
- Generally weak priors that are dominated by the likelihood ("objective Bayes")
- Models that incorporate sampling design features:
  - Capture design weights and stratifying variables as covariates in the prediction model (e.g. Gelman 2007)
  - Clustering via hierarchical random effects models

Full model for *Y* and *I*

$$p(Y, I \mid Z, \theta, \phi) = p(Y \mid Z, \theta)\; p(I \mid Y, Z, \phi)$$

Model for Population

Model for Inclusion

- Full posterior distribution of parameters (hard):

$$p(\theta, \phi \mid Y_{\text{obs}}, Z, I) \propto p(\theta, \phi \mid Z) L(\theta, \phi \mid Y_{\text{obs}}, Z, I)$$

- Posterior distribution ignoring the inclusion mechanism (easier):

$$p(\theta \mid Y_{\text{obs}}, Z) \propto p(\theta \mid Z) L(\theta \mid Y_{\text{obs}}, Z)$$

- When the full posterior reduces to this simpler posterior, the inclusion mechanism is called *ignorable* for Bayesian inference (Rubin 1976)

Conditions when inclusion mechanism can be ignored

- Two general and simple sufficient conditions for ignoring the data-collection mechanism are:

Inclusion at Random (IAR):

$$p(I \mid Y, Z, \phi) = p(I \mid Y_{\text{obs}}, Z, \phi) \text{ for all } Y.$$

Bayesian Distinctness:

$$p(\theta, \phi \mid Z) = p(\theta \mid Z)\, p(\phi \mid Z)$$

- Ignorability is specific to the survey variable *Y,* unlike probability sampling, which guarantees ignorability for any outcome
- In adaptive design, Y_{obs} can include paradata or survey data from earlier waves

Bayes and responsive design

- Predictive Bayes modeling has more potential for gains in efficiency than Bayesian weight modeling
  - Need to model survey variables!
  - Specifically, model relationship of survey variables with weights (as covariates)

Example: subsampling callbacks

- Elliott and Little (2000 JASA) assessed subsampling callbacks for National Comorbidity Study (NCS)
- "Our analysis suggests that randomly dropping a subset of late callbacks will save resources whenever (a) the per call back or per interview cost is increasing, or (b) the probability of a successful interview attempt is decreasing… In general, it appears that surveys with constant or modestly increasing callback costs, such as the 1991 NCS, yield trivial savings, whereas surveys that change mode from postal to telephone or face-to-face interview, such as the U.S. Census Bureau's ACS, yield substantial savings."

Example: subsampling callbacks

- "… our approach yields conservative estimates of efficiency gains from subsampling, in the sense that calculations have **assumed design-based inference** for population means, with **weights included to compensate for differential probabilities of selection**. If modeling assumptions are made about the **distributions of outcomes across callback strata**, then different subsampling schemes might be optimal"
  Elliott and Little (2000)

Example: weighting for nonresponse

$\text{corr}^2(X, Y)$

$\text{corr}^2(X, R)$		Low	High
	Low	bias ---,var ---	bias ---, var $\Downarrow$
	High	bias ---, var $\uparrow$	bias $\Downarrow$, var $\Downarrow$

Too often weighting adjustments put us here … Modeling of relationship between weights and the outcomes is needed to get us out of this square!

We need good predictors of Y – but we focus on predictors of R…

Example: Penalized Spline of Propensity Prediction (PSPP)

- PSPP (Little & An 2004, Zhang & Little 2009, 2011).
- Regression imputation that is
  - Non-parametric (spline) on the propensity to respond
  - Parametric on other covariates
- Exploits the key property of the propensity score that conditional on the propensity score and assuming missing at random, missingness of Y does not depend on other covariates
- This property leads to a model-based version of double robustness (as in GREG).
- Does very well in simulation studies

Penalized Spline of Propensity model

Estimate: Y^{*}=logit (Pr(R=1|$X_1, \ldots, X_p$))

Impute using the regression model:

$$(Y \mid Y^{*}, X_1, \ldots, X_p; \beta) \sim$$

$$N(s(Y^{*}) + g(Y^{*}, X_2, \ldots, X_p; \beta), \sigma^2)$$

- Nonparametric part
- Needs to be correctly specified
- We choose penalized spline

- Parametric part
- Misspecification does not lead to bias
- Increases precision
- X_1 excluded to prevent multi-collinearity

Missing Not at Random Models

- Difficult problem, since information to fit non-MAR is limited and highly dependent on assumptions
- Sensitivity analysis is preferred approach – though this form of analysis is not appealing to consumers of statistics, who want clear answers

An MNAR model: Proxy Pattern-Mixture Analysis

$x_i = x(z_i) =$ best predictor of y_i given covariates z_i

(estimated on respondents, and scaled to same variance as y_i)

$[y_i, x_i \mid r_i = r] \sim G(\mu^{(r)}, \Sigma^{(r)})$

$\Pr(r_i = 1 \mid x_i, y_i) = g\left(y_i^*(\lambda)\right),\; y_i^*(\lambda) = x_i + \lambda y_i$

MAR: $\lambda = 0$, MNAR: $\lambda \neq 0$ (Andridge and Little 2011)

$[y_i$ indep $r_i \mid y_i^*(\lambda)]$, which identifies the model for given λ

$g()$ is arbitrary, unspecified

Sensitivity analysis for different choices of λ (e.g. $0, 1, \infty$)

If x_i is a noisy measure of y_i, it may be plausible to assume $\lambda = \infty$ leading to method for adjustment for predictors with measurement error (West and Little, *Applied Statistics* 2013)

Indices of potential absolute bias (PAB) for a mean

$\lambda = \infty$ leads to following measures of bias for mean of Y:

- Let $\hat{\rho} > 0$ be the estimated correlation between X and Y, based on the sample data.
- Let $\overline{x}$ denote the sample mean of X from the administrative data and $\overline{x}_R, \overline{y}_R$ be the means of X and Y from the respondents.
- Define the *unadjusted potential absolute bias* (PABU) as

$$\text{PABU} = \left| \overline{x} - \overline{x}_R \right| / \hat{\rho}$$

- Define the *adjusted potential absolute bias* (PABA) as

$$\text{PABA} = \left| \overline{x} - \overline{x}_R \right| (1 - \hat{\rho}^2) / \hat{\rho}$$

Bayes and responsive design

- Develop priors based on previous surveys
  - Design-based approach ignores (or treats informally) information from previous surveys
  - Bayes can use prior surveys as "meta-data" to inform decisions for current survey
  - Priors can accommodate down-weighting of previous survey information: e.g. "power" priors (Chen and Ibrahim 2000 Stat Science)
  - Bayesian power calculations – neglected topic, particularly in sample survey context

Bayesian updating

- Bayes rule is natural … <u>the</u> theorem … for sequential decision-making:

$$D_k = \text{ data at stage } k$$

$$p(\theta \mid D_0, D_1, ..., D_k) \propto p(\theta \mid D_0, D_1, ..., D_{k-1}) L(\theta \mid D_k)$$

- Selection is ignorable for likelihood inference, if design at any stage depends on data before that stage
- Basis for sequential treatment allocation in clinical trials –which models the outcomes!
- Relationship between outcomes and propensity (e.g. PSPP) can be modeled and updated from prior stages

Conclusion

- I view Bayesian modeling as a natural framework for developing responsive design and analysis
- No free lunch: models make assumptions
- But assumptions are explicit and can be evaluated and criticized.

References 1

Box, G.E.P. (1980), Sampling and Bayes inference in scientific modeling and robustness (with discussion), *JRSSA*, 143, 383-430.

Joyce, P.M., Malec, D., Little, R.J., Gilary, A., Navarro, A. and Asiala, M.E. (2014). Statistical Modeling Methodology for the Voting Rights Act Section 203 Language Assistance Determinations. *JASA,* 109, 36-47.

Gelman, A. (2007). Struggles with survey weighting and regression modeling. *Statist. Sci.,* 22, 2, 153-164 (with discussion and rejoinder).

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2003), *Bayesian Data Analysis*, 2nd. edition. New York: CRC Press.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. *JRSSB*, 17, 269-278.

Horvitz, D.G. & Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *JASA,* 47, 663-685.

Little, R.J.A. (2004). To Model or Not to Model? Competing Modes of Inference for Finite Population Sampling. *JASA,* 99, 546-556.

References 2

Little, R.J.A. (2006). Calibrated Bayes: A Bayes/frequentist roadmap. *Am. Statist.*, 60, 3, 213-223

_____ (2012). Calibrated Bayes: an alternative inferential paradigm for official statistics (with discussion and rejoinder). *JOS*, 28, 3, 309-372.

_____ (2013). Survey Sampling: Past Controversies, Current Orthodoxies, and Future Paradigms. In *Past, Present and Future of Statistical Science*, COPSS 50th Anniversary Volume, X. Lin, D. L. Banks, C. Genest, G. Molenberghs, D.W. Scott, and J.-L. Wang, eds. CRC Press.

Rubin, DB (1984), Bayesianly justifiable and relevant frequency calculations for the applied statistician, *Annals Statist.* 12, 1151-1172.

Särndal, C.-E., Swensson, B. & Wretman, J.H. (1992), *Model Assisted Survey Sampling*, Springer Verlag: New York.

Zheng, H. & Little, R.J. (2005). Inference for the population total from probability-proportional-to-size samples based on predictions from a penalized spline nonparametric model. *JOS*, 21, 1-20.