Weight Smoothing with Laplace Prior and Its Application in GLM Model

Xi Xia[1]
Michael Elliott[1,2]

[1]Department of Biostatistics, [2]Survey Methodology Program, University of Michigan

National Cancer Institute Grant R01-CA129101

November 4, 2013

Weighting in Complex Survey Design

- When target quantity of interest is correlated with probabilities of inclusion, applying weights inverse to probabilities of inclusion in estimation is common measure to eliminate or reduce bias.
- Some examples are the Horvitz-Thompson estimators of population total and mean:

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \pi_i^{-1} Y_i$$

$$\hat{\mu}_{HT} = N^{-1} \sum_{i=1}^{n} \pi_i^{-1} Y_i$$

- When data are not closely associated with probability of inclusion, incorporating weights increases the variance of estimation due to extra variability in weights.

Weight Trimming

- A common approach to cope with inflated estimation variance is weight trimming or winsorization (Potter 1990, Kish 1992, Alexander et al. 1997)
- Concept: To limit the variability in weights by trimming extreme weights down to a threshold, and redistributing trimmed values among others.
- Target: To reduce variance at cost of increased bias, lead to overall reduction in RMSE.
- Examples: NAEP (Potter 1988), Empirical MSE(Cox and McGrath 1981), Exponential Distribution Method (Chowdbury et al. 2007)

Bayesian Inference Approach

- Treat unobserved sample (Y_{nob}) as missing, and build model($P(y|\theta)$) that captures underlying data pattern.
- To estimate quantity of interest $Q(Y)$, e.g population mean or slope, from marginal posterior predictive distribution (Ericson 1969, Holt and Smith 1979, Little 1993):
  $p(Q(Y)|y) = \int f(Q(Y)|\theta)p(\theta|y)d\theta = \frac{\int f(Q(Y)|\theta)f(y|\theta)p(\theta)d\theta}{\int f(y|\theta)p(\theta)d\theta}$
- Under ignorable sampling design ($p(I|Y,\phi) = p(I|Y_{obs},\phi)$), $p(Y_{nob}|Y_{obs},I) = p(Y_{nob}|Y_{obs})$, allowing inference about $Q(Y)$ without explicitly modeling the sampling inclusion parameter I. (Ericson 1969, Holt and Smith 1979, Little 1993, Rubin 1987, Skinner et al. 1989)
  - Sensible models in still need to account for the sample design in both the likelihood and prior model structure.

Incorporating Unequal Probabilities of inclusion

- Pool samples with same or similar probabilities of inclusion in strata, index by h=1,...H, and re-assign weight as $w_h = N_h/n_h$, where n_h=sample size in weight stratum h, and N_h=population size in weight stratum h.
- Model data by:

$$y_{hi}|\theta_h \ f(y_{hi};\theta_h), i = 1, ...N_h$$

  for all elements in hth inclusion stratum, and θ_h allows for interaction between model parameter(s) and inclusion stratum h.
- Noninformative prior on θ_h represents a fully-weighted analysis on expectation of the posterior predictive distribution of Q(Y).

Weight Smoothing

- Follows the idea of modeling parameter and stratum interaction, but treat strata means as random effects in a hierarchical model to achieve shrinkage estimator between fully-weighted estimate and unweight estimate
- Corresponding hierarchical model:

$$Y_{hi} \overset{iid}{\sim} N(\mu_h, \sigma^2)$$
$$\mu \sim N_H(\phi, G)$$

where $\mu = (\mu_1, ...\mu_H)$, $\phi = (\phi_1, ...\phi_H)$, and $h = 1, ..., H$ indexes different "weight strata" defined

- The posterior mean of the population mean is derived as:

$$E(\bar{Y}|y) = \sum_{h=1}^{H} [n_h \bar{y}_h + (N_h - n_h)\hat{\mu}_h]/N$$

Weight Smoothing for Generalized Linear Models

- To extend weight smoothing model to GLM:
- Basic form of GLM:
$f(y_i|\theta_i,\phi) = \exp\left[\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i,\phi)\right]$
- Link Function:
$g(E(y_i|\theta_i)) = g(\mu_i) = g(b'(\theta_i)) = \eta_i = x_i^T\beta$
- Random effect β:
$(\beta_1^T,...\beta_H^T)^T|\beta^*, G \sim N_{HP}(\beta^*, G)$
- Population Quantity B approximated by:
$\sum_{h=1}^{H} W_h \sum_{i=1}^{n_h} \frac{(\hat{y}_{hi} - g^{-1}(\mu_i(\hat{B})))x_{hi}}{V(\mu_{hi}(\hat{B}))g'(\mu_{hi}(\hat{B}))} = 0$

Laplace Prior

- Inspired by the choice of Laplace prior in Bayesian LASSO(Park & Casella 2008), we apply Laplace prior in weight smoothing model.
- Comparison between Normal prior and Laplace Prior
  Normal Prior:
  $$p(\beta|\sigma^2) = \prod_{j=1}^{p} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\beta_j^2/2\sigma^2}$$
  Conditional Laplace Prior:
  $$p(\beta|\sigma^2) = \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma^2}}$$
- Expect to gain robustness by switching from L2 constraint to L1 constraint.

Laplace Prior

- The absolute value in Laplace distribution raises problems in optimization.
- The problem is solved by reform Laplace distribution into a scale mixture of normal with an exponential mixing density: (Andrews and Mallows 1974)
  $\frac{\alpha}{2}e^{-\alpha|z|} = \int_0^\infty \frac{1}{\sqrt{2\pi s}}e^{-z^2/(2s)}\frac{\alpha^2}{2}e^{-\alpha^2 s/2}ds$
- And Laplace prior turns into a two-level hierarchical model:
  $(\beta_1^T, ..., \beta_H^T)^T|\beta_h^*, D_\tau, \sigma^2 \sim MVN(\beta_h^*, \sigma^2 D_{\tau h})$
  $\sigma^2, \tau_1^2, ...\tau_{Hp}^2 \sim 1/\sigma^2 \prod_{j=1}^{Hp} \frac{\lambda^2}{2}e^{-\lambda^2\tau_j^2/2}$

Weight Smoothing with Laplace Prior

- The overall hierarchical model for weight smoothing model with Laplace prior is presented as following:

$$
\begin{aligned}
y_{hi}|x_{hi}, \beta_h, \sigma^2 &\sim N(x_{hi}^T \beta_h, \sigma^2) \\
(\beta_1^T, ..., \beta_H^T)^T | \beta_h^*, D_\tau, \sigma^2 &\sim MVN(\beta_h^*, \sigma^2 D_{\tau h}) \\
\beta_h^* | \sigma_0^2 &\sim MVN(0, \sigma_0^2 I_p) \\
D_{\tau h} &= diag(\tau_{h1}^2, ..., \tau_{hp}^2) \\
\sigma^2, \tau_1^2, ...\tau_{Hp}^2 &\sim 1/\sigma^2 \prod_{j=1}^{Hp} \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2 / 2} \\
\lambda^2 &\sim Gamma(\gamma = 1, \delta = 1.78)
\end{aligned}
$$

- The close forms for all full conditional distributions exist, and the model could be simulated through Gibbs steps.

Simulation: Population Setting

$$y_i|x_i, \beta, \sigma^2 \sim N(\beta_0 + \sum_{h=1}^{20} \beta_h (x_i - h)_+, \sigma^2)$$

$$x_i \sim UNI(0, 10), i = 1, ..., N = 20,000.$$

$$1. \beta_a = (0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)'$$

$$2. \beta_b = (0, 0, 0, 0, 0, 0, .5, .5, .5, .5, 1, 1, 1, 1, 2, 2, 2, 2, 4, 4, 4)'$$

$$3. \beta_c = (0, 22, -4, -4, -2, -2, -2, -2, -1, -1, -1, -1,$$

$$-.5, -.5, -.5, -.5, 0, 0, 0, 0, 0)'$$

$$\sigma^2 = 10^l, l = 1, 3, 5$$

$$P(I_i|H_i) = \pi_i \propto (1 + H_i/15)H_i.$$

$$H_i = [2X_i]/2$$

Goal, Sampling and Simulation Details

- Goal: To estimate population slope B
- Sample Size: $n = 1000$
- Simulation Count: 200
- Data-based prior for β
- 50,000 iterations with 10,000 burn-in
- Compare weight smoothing with Laplace prior(HWS) with unweighted estimate(UWT), fully weighted estimate(FWT), weight smoothing with exchangeable random effect(XRS):

$$y_{hi}|x_{hi}, \beta_h, \sigma^2 \sim N(x_{hi}^T \beta_h, \sigma^2)$$

$$(\beta_1^T, ..., \beta_H^T)^T | \beta^*, \Sigma \sim MVN(\beta^*, \Sigma)$$

$$p(\sigma, \beta^*, \Sigma) \propto \sigma^{-2} \mid \Sigma \mid^{-(p+1/2)} \exp(-1/2tr\{2\Sigma^{-1}\})$$

Simulation Result

Table 1: RMSE relative to fully weighted estimator (nominal 95% CI coverage in parenthesis)

	β_a Variance log_{10}			β_b Variance log_{10}			β_c Variance log_{10}		
	1	3	5	1	3	5	1	3	5
UWT	0.73 (95)	0.69 (95)	0.72 (96)	10.20 (0)	2.44 (2)	0.73 (95)	6.23 (0)	2.18 (1)	0.76 (90
WT	1 (94)	1 (93)	1 (96)	1 (100)	1 (92)	1 (96)	1 (100)	1 (97)	1 (95)
XRS	1.49 (99)	0.72 (96)	0.72 (96)	1.01 (95)	2.21 (94)	1.20 (94)	1.87 (6)	2.05 (1)	0.76 (91
HWS	1.05 (96)	0.98 (95)	0.77 (99)	0.45 (97)	0.95 (96)	0.77 (99)	0.34 (85)	0.94 (96)	0.77 (99

Application: Dioxin study from NHANES

- We present the performance of weight smoothing model with Laplace prior on Dioxin data from 2003-2004 NHANES study.
- The target is to estimate the linear effect of Age, Gender on log TCDD in blood.
- Altogether 1250 individuals sampled from 25 Strata, 2 MVU each.
- Reading below measurement threshold is corrected with Multiple Imputation, resulting in 5 replicates.

Application: Dioxin study from NHANES

Table 2: Relative RMSE for Dioxin study

Model	Age only	Gender only	Age and Gender	
			Age	Gender
UWT	0.840	1.960	0.846	1.464
FWT	1	1	1	1
HWS	0.312	0.953	0.315	0.919

Model	Age and Gender Interaction		
	Age	Gender	Interaction
UWT	1.412	0.488	0.448
FWT	1	1	1
HWS	0.770	0.393	0.364

Conclusion and Discussion

- By applying Laplace prior, the weight smoothing model is able to obtain robust estimator with less complicated structure, leading to a faster algorithm.
- The Bayesian finite population inference provide more than just a shrinkage estimator between fully weighted estimate and unweighted estimate. In some situation, it could provide estimate with overall smaller RMSE than both.
- Extensions to GLM (logistic regression) have been done.
  - Less savings on RMSE (10-15%)
  - Coverage similar to fully-weighted estimator (both substantially undercover when weight/slope correlation is weak).
- The gaining in RMSE sometimes comes with a cost of moderate drop in 95% coverage. It is worth exploring the model's mechanism in reducing the RMSE and the limit of the scenarios under which it still maintains reasonable converage.