# NUMERICAL COMPARISON AMONG DIFFERENT EMPIRICAL PREDICTION INTERVALS

Masayo Yoshimori

Research Fellow of JSPS, Graduate School of Engineering Science, Osaka University

(The research was conducted under the supervision of Professor Partha Lahiri at the University of Maryland, College Park.)

September 4th, 2013

# Outline

1. Empirical Bayes estimator under the Fay-Herriot model
2. Confidence Interval
3. Simulation study
4. Conclusion

# The Fay Herriot Bayesian Model

**Ref:** Fay and Herriot (JASA, 1979)

For $i = 1, \cdots, m$,

*Level 1: (Sampling Distribution)*: $y_i|\theta_i \sim N(\theta_i, D_i)$;

*Level 2: (Prior Distribution)*: $\theta_i \sim N(\mathbf{x}_i'\boldsymbol{\beta}, A)$

where

- $m$ : number of small area;
- $y_i$ : direct survey estimate of $\theta_i$;
- $\theta_i$ : true mean for area $i$;
- $\mathbf{x}_i$ : $p \times 1$ vector of known auxiliary variables;
- $D_i$: *known* sampling variance of the direct estimate;
- The $p \times 1$ vector of regression coefficients $\boldsymbol{\beta}$ and model variance $A$ are unknown.

# Bayes Estimator of $\theta_i$

**The purpose is to predict a true mean for $i$ area, $\theta_i$**

When model variance $A$ is known, the following Bayes estimator of $\theta_i$ is obtained by minimizing $MSE(\hat{\theta}_i)$ among all linear unbiased predictors of $\theta_i$, where $MSE(\hat{\theta}_i) = E[(\hat{\theta}_i - \theta_i)^2]$ and $E$ is the expectation with respect to the Fay-Herriot model:

$$\hat{\theta}_i^B = (1 - B_i)y_i + B_i \mathbf{x}_i' \hat{\boldsymbol{\beta}},$$

where

- $B_i \equiv B_i(A) = \frac{D_i}{A+D_i}$
- $\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}(A) = (X'V^{-1}X)^{-1}X'V^{-1}y$ where $V \equiv V(A) = diag(A + D_1, \cdots, A + D_m)$.

# Empirical Bayes (EB) Estimator of $\theta_i$

Let model variance $\hat{A}$ be a consistent estimator of $A$, for large $m$.

An EB of $\theta_i$ is given by

$$\hat{\theta}_i^{EB} = (1 - \hat{B}_i)y_i + \hat{B}_i \mathbf{x}_i' \hat{\boldsymbol{\beta}}.$$

where

- $\hat{B}_i = \frac{D_i}{\hat{A}+D_i}$
- $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{A})$

*Ref:* Efron and Morris (JASA, 1975), Fay and Herriot (JASA, 1979)

# Confidence Interval for $\theta_i$

An interval, denoted by $I_i$, is called a $100(1-\alpha)\%$ interval for $\theta_i$ if

$$P(\theta_i \in I_i | \beta, A) = 1 - \alpha, \ \forall \beta \in R^p, A \in R^+,$$

where

- the probability $P$ is with respect to the joint distribution of $\{(y_i, \theta_i), i = 1, \cdots, m\}$ under the Fay-Herriot model;
- $R^+$ is the positive part of the real line.

# A General Form of Confidence Interval for $\theta_i$

Most of the intervals proposed in the literature can be written as:

$$(\hat{\theta}_i + q_1(\alpha)\hat{\tau}_i(\hat{\theta}_i), \hat{\theta}_i + q_2(\alpha)\hat{\tau}_i(\hat{\theta}_i))$$

where

- $\hat{\theta}_i$ is an estimator of $\theta_i$;
- $\hat{\tau}_i(\hat{\theta}_i)$ is an estimate of the measure of uncertainty of $\hat{\theta}_i$;
- $q_1(\alpha)$ and $q_2(\alpha)$ are chosen suitably in an effort to attain coverage probability close to the nominal level $1 - \alpha$.

# Direct Confidence Interval

The choice $\hat{\theta}_i = y_i$ leads to the direct interval $I_i^D$ given by

$$I_i^D : y_i \pm z_{\alpha/2}\sqrt{D_i},$$

where $z_{\alpha/2}$ is the upper $100(1-\alpha/2)\%$ point of $N(0,1)$.

**Remarks**:

- The coverage probability is $1-\alpha$;
- When $D_i$ is large, the length is too large to make any reasonable conclusion.

# Synthetic Confidence Interval

**Ref:** Hall and Maiti (JRSS, 2006)

$$(x_i'\hat{\beta} + q_1(\alpha)\sqrt{\hat{A}},\ x_i'\hat{\beta} + q_2(\alpha)\sqrt{\hat{A}})$$

where

- $\hat{A}$ are consistent estimators of $A$. For example, residual maximam likelihood estimator (REML).
- $L_i^*[q_2(\alpha)] - L_i^*[q_1(\alpha)] = 1 - \alpha$ where $L_i^*$ is a parametric bootstrap approximation of the distribution $L_i$ of $\frac{\theta_i - x_i'\hat{\beta}}{\sqrt{\hat{A}}}$.

**Remarks**:

- The method is synthetic (Rao 2005).
- This approach could be useful in situations especially when $y_i$ is missing for the $i$th area.

# Bayesian Credible Interval

Assume $\beta$ and $A$ are known.

$$I_i^B(A) : \hat{\theta}_i^B(A) \pm z_{\alpha/2}\sigma_i(A),$$

where

- $\hat{\theta}_i^{\mathrm{B}} \equiv \hat{\theta}_i^B(A) = (1 - B_i)y_i + B_i x_i' \beta,$
- $B_i \equiv B_i(A) = \frac{D_i}{D_i + A},$
- $\sigma_i(A) = \sqrt{\frac{AD_i}{A+D_i}}$

**Remarks**:

- $\theta_i | y_i; \beta, A \sim N[\hat{\theta}_i^B(A), g_{1i} = \sigma_i^2(A)].$
- The Bayesian credible interval cuts down the length of the direct confidence interval by $100 \times (1 - \sqrt{1 - B_i})\%$
- The maximum benefit from the Bayesian methodology is achieved when $B_i$ is near 1.

# Empirical Bayes Confidence Interval

**Ref:** Cox (1975)

$$I_i^{Cox}(\hat{A}) : \hat{\theta}_i^{EB}(\hat{A}) \pm z_{\alpha/2}\sigma(\hat{A}),$$

where

- $x_i^T\beta = \mu$ is estimated by the sample mean $\bar{y} = m^{-1}\sum_{i=1}^m y_i$ and
- $A$ by the ANOVA estimator:
  $\hat{A}_{ANOVA} = \max\left\{(m-1)^{-1}\sum_{i=1}^m (y_i - \bar{y})^2 - D, 0\right\}$.

**Remarks**:

- The length of the Cox interval is smaller than that of the direct interval.
- The distribution of $\frac{\theta_i - \hat{\theta}_i^{\mathrm{EB}}}{\sigma(\hat{A})}$ is not a standard Normal. Thus, it is not appropriate to use the Normal quantile $z_{\alpha/2}$ as the cut-off points.
- The Cox empirical Bayes confidence interval introduces a coverage error of the order $O(m^{-1})$, not accurate enough in most small area applications.
- length of the interval is zero when $\hat{A}_{ANOVA} = 0$

# Other EB Confidence Intervals

1. Replace $\sigma(\hat{A})$ by a measure of uncertainty that captures uncertainty due to estimation of the hyperparameters $\beta$ and $A$ (e.g., $\sqrt{g_{1i} + g_{2i} + 2g_{3i}}$) (Ref: Morris (JASA, 1983) Prasad and Rao (JASA, 1990))
2. Replace $z_{\alpha/2}$ by $z_{\alpha/2} c_i(\hat{A})$ to reduce the coverage error to $O(m^{-1.5})$ (Datta et al., Scand. Stat. 2002; Basu et al. 2003; Sasase and Kubokawa, JRSS., 2005; Yoshimori, Comm. Stat., 2013)
3. Parametric bootstrap (Laird and Louis, JASA 1987; Carlin and Louis 1996; Chatterjee et al., AS 2008)

# Parametric Bootstrap Confidence Interval

**Ref:** Chatterjee, Lahiri and Li (AS, 2008)

- Use the distribution of $\frac{\theta_i^* - \hat{\theta}_i^{\mathrm{EB}*}}{\sigma_i(\hat{A}^*)}$ to approximate the distribution of $\frac{\theta_i - \hat{\theta}_i^{\mathrm{EB}}}{\sigma_i(\hat{A})}$.
- Compute $\hat{\beta}$ and $\hat{A}$;
- Draw bootstrap sample from the following bootstrap model:
  (i) $y_i^* | \theta_i^* \stackrel{ind}{\sim} N(\theta_i^*, D_i)$
  (ii)$\theta_i^* \stackrel{ind}{\sim} N(x_i' \hat{\beta}, \hat{A})$
- Compute $\hat{\beta}^*$ and $\hat{A}^*$ from $y^*$. Then we have $\hat{\theta}_i^{\mathrm{EB}*} = (1 - \hat{B}^*) y_i^* + \hat{B}^* x_i' \hat{\beta}^*$, and $\sigma_i^2(\hat{A}^*) = \frac{A^* D_i}{A^* + D_i}$;
- Compute $(\theta_i^* - \hat{\theta}_i^{\mathrm{EB}*}) / \sigma_i(\hat{A}^*)$.

**Remarks**:

- When REML estimates gets zero, we need to truncated by some small values.

# Parametric Bootstrap Confidence Interval

**Parametric Bootstrap Confidence Interval**

$$\mathrm{CI}_i^{\mathrm{PB}} = [\hat{\theta}_i^{\mathrm{EB}} + q_1(\alpha)\sigma_i(\hat{A}),\ \hat{\theta}_i^{\mathrm{EB}} + q_2(\alpha)\sigma_i(\hat{A})],$$

where $L_i^*[q_2(\alpha)] - L_i^*[q_1(\alpha)] = 1 - \alpha$, and $L_i^*$ is a parametric bootstrap approx. of the distribution of $\frac{\theta_i - \hat{\theta}_i^{\mathrm{EB}}}{\sigma_i(\hat{A})}$.

**Theorem**

Under reg. cond. $\Pr(\theta_i \in \mathrm{CI}_i^{\mathrm{PB}}) = 1 - \alpha + O(m^{-1.5})$,

# A Research Question

Which of the confidence intervals one should use when REML is used to estimate $A$?

**Restricted Maximum Likelihood estimator (REML estimator)**

$$\hat{A}_{RE} = \max\{\arg\max_{0<A<\infty} |X'V^{-1}(A)X|^{-1/2}|V|^{-1/2}\exp\{-\frac{1}{2}y'Py\} \times K,\ 0\}$$

where $K$ is a generic constant free from $A$ and
$P \equiv P(A) = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$.

# Simulation set-up: The Fay-Herriot Model with Unequal Sampling Variances

$m = 15,\ 45,$
$x_i'\beta = 0,\ A = 1$
There are two patterns of sampling variance $D_i$;

- Pattern (a)$\{0.7, 0.5, 0.4, 0.3\}$,
- Pattern (b)$\{20, 6, 5, 4, 2\}$.

(When REML estimate gets zero, we truncated it as 0.01.)

**CLL**:the parametric bootstrap confidence interval (Chatterjee et al, 2008);
**HM**:Synthetic Confidence interval (Hall and Maiti, 2006);
**Cox**:Cox empirical confidence interval (Cox, 1975);
**PR**:the method which is used second order unbiased estimator of MSE (Prasad and Rao, 1990);
**Y**:the method, which $z_{\alpha/2}$ is replaced by $z_{\alpha/2}c_i(\hat{A})$ for some $c_i$, (Under the Fay-Herriot model, Yoshimori, 2003).

# Simulation Results 1

**m=15, Pattern (a)**$\{0.7, 0.5, 0.4, 0.3\}$**, Pattern (b)**$\{20, 6, 5, 4, 2\}$**.**

Table: Average coverage and length for difference confidence intervals (average taken over the three areas within each group); nominal level=0.95

| Group | CLL | | HM | | Cox | | PR | | Y | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Pattern (a) | | | | | |
| 1 | 97.5 | (3.4) | 97.9 | (5.1) | 90.3 | (2.4) | 93.8 | (2.6) | 96.5 | (3.7) |
| 2 | 97.4 | (3.3) | 98.0 | (5.1) | 90.6 | (2.3) | 94.0 | (2.5) | 96.2 | (3.5) |
| 3 | 97.2 | (3.0) | 97.9 | (4.9) | 90.7 | (2.1) | 94.3 | (2.4) | 96.2 | (3.4) |
| 4 | 97.2 | (2.8) | 97.8 | (4.8) | 91.0 | (2.0) | 94.5 | (2.2) | 96.1 | (3.2) |
| 5 | 97.0 | (2.4) | 97.5 | (4.6) | 91.7 | (1.8) | 95.1 | (2.0) | 96.1 | (2.9) |
| | | | | | Pattern (b) | | | | | |
| 1 | 84.8 | (23.7) | 84.8 | (25.0) | 61.9 | (3.2) | 88.9 | (4.8) | 100.0 | (3421.6) |
| 2 | 85.3 | (20.2) | 85.3 | (23.4) | 61.9 | (2.9) | 95.1 | (5.1) | 99.9 | (3419.2) |
| 3 | 85.8 | (19.4) | 85.8 | (22.9) | 62.0 | (2.8) | 96.1 | (5.1) | 99.9 | (3418.5) |
| 4 | 86.0 | (18.2) | 86.0 | (22.2) | 62.0 | (2.7) | 97.4 | (5.2) | 99.8 | (3417.6) |
| 5 | 87.6 | (13.9) | 87.6 | (19.1) | 62.7 | (2.4) | 99.1 | (5.4) | 99.5 | (3413.3) |

# Simulation Results 2

**m=45, Pattern (a)**$\{0.7, 0.5, 0.4, 0.3\}$**, Pattern (b)**$\{20, 6, 5, 4, 2\}$**.**

Table: Average coverage and length for difference confidence intervals (average taken over the three areas within each group); nominal level=0.95

| Group | CLL | | HM | | Cox | | PR | | Y | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Pattern (a) | | | | | |
| 1 | 95.0 | (2.6) | 95.3 | (4.0) | 93.6 | (2.5) | 94.5 | (2.6) | 94.8 | (2.6) |
| 2 | 95.1 | (2.5) | 95.2 | (4.0) | 93.8 | (2.4) | 94.6 | (2.4) | 94.9 | (2.5) |
| 3 | 95.1 | (2.3) | 95.2 | (4.0) | 94.0 | (2.2) | 94.8 | (2.3) | 95.1 | (2.3) |
| 4 | 95.1 | (2.2) | 95.3 | (4.0) | 94.2 | (2.1) | 94.8 | (2.1) | 95.0 | (2.1) |
| 5 | 95.0 | (1.9) | 95.2 | (3.9) | 94.2 | (1.9) | 94.8 | (1.9) | 95.0 | (1.9) |
| | | | | | Pattern (b) | | | | | |
| 1 | 88.7 | (13.0) | 88.6 | (13.4) | 75.1 | (3.4) | 85.9 | (4.0) | 99.9 | (585.9) |
| 2 | 88.7 | (12.0) | 88.7 | (13.1) | 75.3 | (3.1) | 90.4 | (4.0) | 99.8 | (585.1) |
| 3 | 89.0 | (11.7) | 89.0 | (13.0) | 75.5 | (3.1) | 91.6 | (4.0) | 99.8 | (584.9) |
| 4 | 89.0 | (11.3) | 89.0 | (12.8) | 75.4 | (3.0) | 92.6 | (4.0) | 99.7 | (584.7) |
| 5 | 89.5 | (9.6) | 89.5 | (12.0) | 75.6 | (2.7) | 96.3 | (3.9) | 99.6 | (583.4) |

# Conclusion

We compared the performances of several confidence intervals using the REML estimator of $A$.

**Our simulation results**

- All intervals perform well except for the Cox empirical Bayes confidence interval in pattern (a).
- The method based on the Taylor serious approximation can have large length for pattern (b).
- Overall, CLL and HM have similar coverage but CLL has usually shorter length than the HM method; both methods seems to have an under-coverage problem for pattern (b) even when we increase m from 15 to 45.
- REML method is not suitable for small area inference even when using a parametric bootstrap method.

**As future study**

We must improve the empirical prediction interval in order to find a better estimator than that of the REML for the unknown variance parameter $A$.

# Reference


- Basu, R., Ghosh, J. K. and Mukerjee, R. (2003). Empirical Bayes prediction intervals in a normal regression model: higher order asymptotic. *Statist. Prob. Letters*, **63** 197-203
- Carlin, B. P. and Louis, T. A. (1996). *Bayes and empirical Bayes methods for data analysis.* Chapman and Hall, London.
- Chatterjee, A, Lahiri, P and Li, H. (2008). Parametric bootstrap approximation to the distribution of EBLUP, and related prediction intervals in linear mixed models, *The Annals of Statistics* **36** 1221-1245.
- Cox, D. R. (1975). Prediction intervals and empirical Bayes confidence intervals. in:J.Gani(Ed.), *Perspectives in Probability and Statistics, Papers in Honor of M.S. Bartlett, Academic Press*. 47-55.
- Datta G. S., Ghosh, M., Smith, D. and Lahiri, P. (2002). On an asymptotic theory of conditional and unconditional coverage probabilities of empirical Bayes confidence intervals. *Scand. J. Statist.* **29** 139-152.
- Efron, B. and Morris, C. N. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.* **70** 311-319.
- Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data, *Journal of the American Statistical Association* **74** 269-277
- Laird, N. M. and Louis, T. A. (1987). Empirical Bayes Confidence Intervals Based on Bootstrap Samples. *J. Amer. Statist. Assoc.* **82** 739-750.
- Li, H and Lahiri, P. (2010). An adjusted maximum likelihood method for solving small area estimation problems, *Journal of Multivariate Analysis*, **101** 882-892
- Morris, C. N. (1983b). Parametric empirical Bayes confidence intervals. In *Proc. Conf. Sci. Infer. Data Anal. Robustness, Ed. G. E. P. Box, T. Leonard and C. F. J. Wu.* 25-50. New York: Academic Press.
- Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small area estimators. *J. Amer. Statist. Assoc.* **85** 163-171.
- Rao, J. N. K. (2005). Inferential issues in small area estimation: some new developments. *Statistics in transition*, **7**, 513–326.
- Sasase, Y and Kubokawa, T. (2005). Asymptotic Correction of Empirical Bayes Confidence Intervals and its Application to Small Area Estimation (in Japanese), *Journal of Japan statistical society*, **35**, 27-54
- **Yoshimori, M. (2013) Numerical comparison between different empirical prediction intervals under the Fay-Herriot Model,** *Communications in Statistics - Simulation and Computation*, **in press**