Bayesian inference for sample surveys

Roderick Little

Trivellore Raghunathan

Module 5: Bayesian models for simple random samples

University of Michigan

Consulting Example

- In India (during the late 70’s), any person possessing a radio, transistor or television has to pay a license fee.
- In a densely populated area with mostly makeshift houses practically no one was paying these fees.
- It was determined that for enforcement to be fiscally meaningful, the proportion of households possessing one or more of these devices must exceed certain limit.

Consulting example (continued)

$N =$ Population Size

$$Y_i = \begin{cases} 1, & \text{if household } i \text{ has a device} \\ 0, & \text{otherwise} \end{cases}$$

$$Q = \sum_{i=1}^{N} Y_i / N \quad \text{Proportion of households with a device}$$

Question of Interest: $\Pr(Q \geq 0.3)$

- If the probability of Q exceeding 0.3 is very high then enforcement might be fiscally sensible
- Conduct a small scale survey to answer the question of interest
- Note that question only makes sense under Bayes paradigm

General Setup

- Model for $I = (I_1, I_2, \cdots, I_N)$: Sample Design
- Model for $Y = (Y_1, Y_2, \cdots, Y_N)$: Prior
- Frame/Design Variables: Z
- Joint distribution: $\Pr(Y, I \mid Z)$
- Observed Data: (Y_{inc}, I, Z)
- Missing or Unobserved Data: Y_{exc}
- Inference: $\Pr(Y_{exc} \mid Y_{inc}, I, Z)$

I	Z	Y
1		
1		Y_{inc}
1		
0		
0		
0		Y_{exc}
0		
0		

Simple Random Sample

- Consider *Z* is not available
- *Pr(Y,I)=Pr(Y)Pr(I)*
- Exchangeable Prior/Model for *Y*
  - For any two permutations of the labels or index used in *Y*

$$(i_1, i_2, \cdots, i_N) \text{ and } (j_1, j_2, \cdots, j_N)$$

$$\Pr(Y_{i_1}, Y_{i_2}, \cdots, Y_{i_N}) = \Pr(Y_{j_1}, Y_{j_2}, \cdots, Y_{j_N})$$

- That is, the labels have no “information” relevant for the inference
- de Finetti (1937), Hewitt & Savage (1955) and Diaconis & Freedman (1980)
- Exchangeable distribution can be expressed as

$$\Pr(Y_1, Y_2, \cdots, Y_N) = \int \prod_{i=1}^{N} \Pr(Y_i \mid \theta)\pi(\theta)d\theta$$

Consulting example

srs of size n, $Y_{\text{inc}} = \{Y_1,...,Y_n\}$, $Y_{\text{exc}} = \{Y_{n+1},...,Y_N\}$

$Y_i \mid \theta \sim iid\ \text{Bernoulli}(\theta)$

Model for observable

$\pi(\theta) = 1 \quad \theta \in (0,1)$

Prior distribution

$$x = \sum_{i=1}^{n} Y_i$$

$$f(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$Q = \sum_{i=1}^{N} Y_i / N = \left(x + \sum_{i=n+1}^{N} Y_i \right) / N$$

Estimand

Binomial Example

The posterior distribution is

$$p(\theta \mid x) = \frac{f(x \mid \theta)\pi(\theta)}{\int f(x \mid \theta)\pi(\theta)d\theta} \propto f(x \mid \theta)\pi(\theta)$$

$$p(\theta \mid x) = \frac{\binom{n}{x}\theta^{x}(1-\theta)^{n-x} \times 1}{\int \binom{n}{x}\theta^{x}(1-\theta)^{n-x}d\theta}$$

$$\theta \mid x \sim Beta(x+1, n-x+1)$$

$$Q = (x + \sum_{i=n+1}^{N} Y_i)/N$$

$$\left(\sum_{i=n+1}^{N} Y_i \mid \theta, x\right) \sim \mathrm{Bin}(N-n, \theta)$$

Infinite Population

$$For\ N \to \infty,\ \bar{Y}_N \to \theta$$

$$\Pr(\bar{Y}_N \geq 0.3 \mid x) \approx \Pr(\theta \geq 0.3 \mid x)$$

Compute using cumulative distribution function

of a beta distribution which is a standard function

in most software such as SAS, R

What is the maximum proportion of households in the population with devices that can be said with great certainty?

$$\Pr(\theta \leq ? \mid x) = 0.9$$

Inverse CDF of Beta Distribution

Numerical Example

- $N=270$ households
- $n=20$ SRS sample
- $x=8$ out 20 had a device
- *Simulation*
  - *Write a R-code for* $\theta \sim Beta(9,13); X_{N-n} \sim Bin(250,\theta);$ *and compute* $\bar{Y}_N = (x + X_{N-n})/N$
  - *Generate Treat 250 households with missing values and use missing data package (for example IVEware or MICE in R or MI in Stata) which fits the model*

$$\Pr(\bar{Y}_N \geq 0.3 \mid x) = \Pr(Y_{N-n} \geq 0.3 \times N - x \mid x)$$

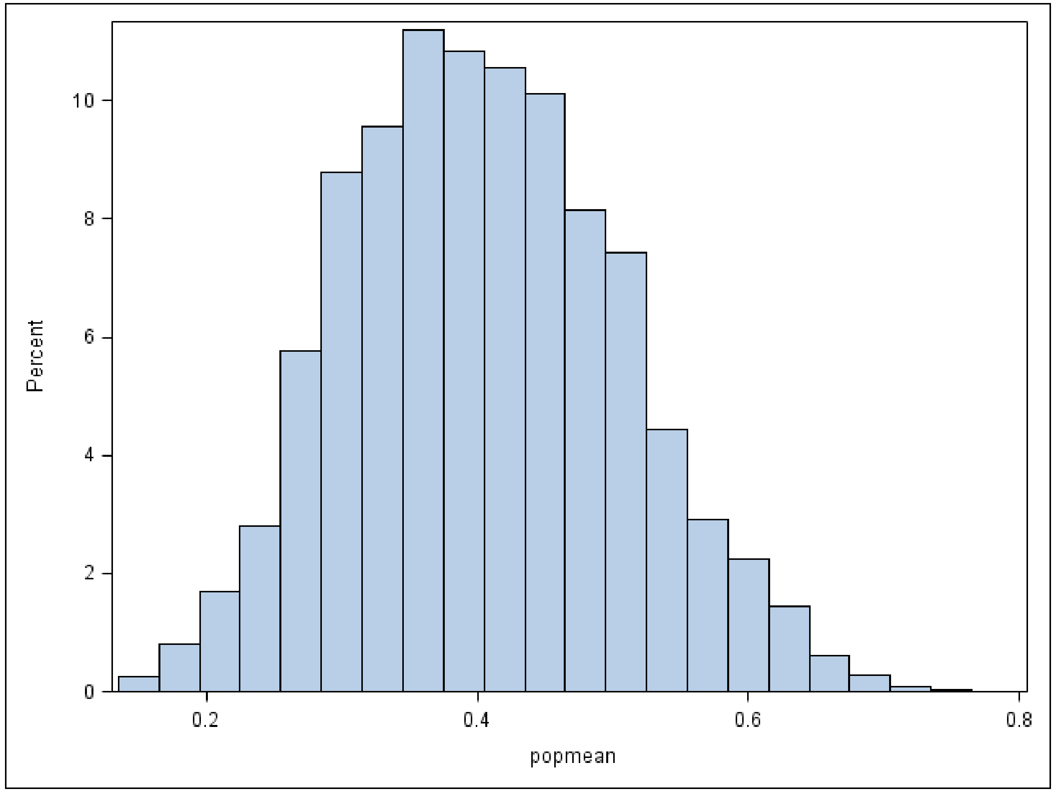
$$= \int_0^1 \Pr(Y_{N-n} \geq 0.3 \times N - x \mid \theta, x)\pi(\theta \mid x)d\theta$$

$$\Pr(Y = 1) = \exp(\beta)/(1 + \exp(\beta)), \pi(\beta) \propto 1$$

or

$$Y \sim Bern(\theta), \pi(\theta) \propto \theta^{-1}(1-\theta)^{-1}$$

Histogram of the 2,500 Draws



Proportion of Draws exceeding 0.3=84%
Posterior mean: 0.4051
Posterior standard deviation: 0.1007
Normal Approximation credible interval:
(0.2077, 0.6025)

Point Estimates

- Point estimate is often used as a single summary “best” value for the unknown Q
- Some choices are the mean, mode or the median of the posterior distribution of Q
- For symmetrical distributions an intuitive choice is the center of symmetry
- For asymmetrical distributions the choice is not clear. It depends upon the “loss” function.

Interval Estimation

- Better summary is an interval estimate
- Fix the coverage rate $1-\alpha$ in advance and determine the *highest posterior density* region C to include most likely values of Q totaling $1-\alpha$ posterior probability
- Fix the value Q_o in advance, determine C by the collection of values of Q more likely than Q_o and calculate the coverage $1-\alpha$ as the posterior probability of this C

Interval Estimates

C is such that

(1) $p(Q \mid Y_{\text{inc}}) > p(Q' \mid Y_{\text{inc}})$

$Q \in C, Q' \notin C$

(2) $\Pr(Q \in C \mid Y_{\text{inc}}) = 1 - \alpha$

"Most likely" is usually defined by highest posterior density

- Highest Posterior Density Region
- For symmetric unimodal posterior distributions, $(1-\alpha)$ HPD interval is $(q_{\alpha/2}, q_{1-\alpha/2})$ where $\Pr(Q \leq q_{\alpha/2}) = \alpha/2$
- In the Binomial example, the beta density of θ used to determine the interval estimate of Q

Normal simple random sample

$$Y_i \sim \text{iid } N(\mu, \sigma^2); i = 1, 2, ..., N$$

$$\pi(\mu, \sigma^2) \propto \sigma^{-2}$$

simple random sample results in $Y_{\text{inc}} = (y_1, ..., y_n)$

$$Q = \overline{Y} = \frac{n\overline{y} + (N-n)\overline{Y}_{\text{exc}}}{N}$$

$$= f \times \overline{y} + (1 - f) \times \overline{Y}_{\text{exc}}$$

Derive posterior distribution of Q

Normal simple random sample

$$Y_i \sim \text{iid } N(\mu, \sigma^2); i = 1, 2, ..., N$$

$$\pi(\mu, \sigma^2) \propto \sigma^{-2}$$

simple random sample results in $Y_{\text{inc}} = (y_1, ..., y_n)$

$$Q = \overline{Y} = \frac{n\overline{y} + (N-n)\overline{Y}_{\text{exc}}}{N}$$

$$= f \times \overline{y} + (1-f) \times \overline{Y}_{\text{exc}}$$

Derive posterior distribution of Q

Normal Example

Posterior distribution of (μ, σ^2)

$$p(\mu, \sigma^2 \mid Y_{\text{inc}}) \propto (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i \in \text{inc}} \frac{(y_i - \mu)^2}{\sigma^2} \right)$$

$$\propto (\sigma^2)^{-n/2-1} \exp - \frac{1}{2} \left(\sum_{i \in \text{inc}} (y_i - \bar{y})^2 / \sigma^2 - n(\mu - \bar{y})^2 / \sigma^2 \right)$$

The above expressions imply that

(1) $\sigma^2 \mid Y_{\text{inc}} \sim \sum_{i \in \text{inc}} (y_i - \bar{y})^2 / \chi^2_{n-1}$

(2) $\mu \mid Y_{\text{inc}}, \sigma^2 \sim N(\bar{y}, \sigma^2 / n)$

Posterior Distribution of Q

$$\bar{Y}_{\text{exc}} \mid \mu, \sigma^2 \sim N(\mu, \frac{\sigma^2}{N-n})$$

$$\bar{Y}_{\text{exc}} \mid \sigma^2, Y_{\text{inc}} \sim N\left(\bar{y}, \frac{\sigma^2}{N-n} + \frac{\sigma^2}{n} = \frac{\sigma^2}{(1-f)n}\right)$$

$$Q = f \times \bar{y} + (1-f) \times \bar{Y}_{\text{exc}}$$

$$Q \mid \sigma^2, Y_{\text{inc}} \sim N\left(\bar{y}, \frac{(1-f)\sigma^2}{n}\right)$$

$$\bar{Y}_{\text{exc}} \mid Y_{\text{inc}} \sim t_{n-1}\left(\bar{y}, \frac{s^2}{(1-f)n}\right)$$

$$Q \mid Y_{\text{inc}} \sim t_{n-1}\left(\bar{y}, \frac{(1-f)s^2}{n}\right)$$

HPD Interval for Q

Note the posterior t distribution of Q is symmetric and unimodal -- values in the center of the distribution are more likely than those in the tails.

Thus a (1-α)100% HPD interval is:

$$\bar{y} \pm t_{n-1,1-\alpha/2} \sqrt{\frac{(1-f)s^2}{n}}$$

Like frequentist confidence interval, but recovers the t correction

Some other Estimands

- Suppose Q=Median or some other percentile
- One is better off inferring about all non-sampled values
- As we will see later, simulating values of Y_{exc} add enormous flexibility for drawing inferences about any finite population quantity
- Modern Bayesian methods heavily rely on simulating values from the posterior distribution of the model parameters and predictive-posterior distribution of the nonsampled values
- Computationally, if the population size, N, is too large then choose any arbitrary value K large relative to n, the sample size
  - National sample of size 2000
  - US population size 306 million
  - For numerical approximation, we can choose $K=2000/f$, for some small f=0.01 or 0.001.

Comparison of Two Populations

- Population 1

$$
\begin{aligned}
&Population\ size = N_1\\
&Sample\ size = n_1\\
&Y_{1i} \sim ind\ N(\mu_1, \sigma_1^2)\\
&\pi(\mu_1, \sigma_1^2) \propto \sigma_1^{-2}
\end{aligned}
$$

$$
\begin{aligned}
&Sample\ Statistics:\ (\bar{y}_1, s_1^2)\\
&Posterior\ distributions:\\
&(n_1 - 1)s_1^2 / \sigma_1^2 \sim \chi^2_{n_1 - 1}\\
&\mu_1 \sim N(\bar{y}_1, \sigma_1^2 / n_1)\\
&Y_{1i} \sim N(\mu_1, \sigma_1^2), i \in \text{exc}
\end{aligned}
$$

- Population 2

$$
\begin{aligned}
&Population\ size = N_2\\
&Sample\ size = n_2\\
&Y_{2i} \sim ind\ N(\mu_2, \sigma_2^2)\\
&\pi(\mu_2, \sigma_2^2) \propto \sigma_2^{-2}
\end{aligned}
$$

$$
\begin{aligned}
&Sample\ Statistics:\ (\bar{y}_2, s_2^2)\\
&Posterior\ distributions:\\
&(n_2 - 1)s_2^2 / \sigma_2^2 \sim \chi^2_{n_2 - 1}\\
&\mu_2 \sim N(\bar{y}_2, \sigma_2^2 / n_2)\\
&Y_{2i} \sim N(\mu_2, \sigma_2^2), i \in \text{exc}
\end{aligned}
$$

Estimands

- Examples
  - $\bar{Y}_1 - \bar{Y}_2$ (Finite sample version of Behrens-Fisher Problem)
  - Difference $\Pr(Y_1 > c) - \Pr(Y_2 > c)$
  - Difference in the population medians
  - Ratio of the means or medians
  - Ratio of Variances
- It is possible to analytically compute the posterior distribution of some these quantities
- It is a whole lot easier to simulate values of non-sampled $Y_1^{'s}$ in Population 1 and $Y_2^{'s}$ in Population 2

Bayesian Nonparametric Inference

- Population: $Y_1, Y_2, Y_3, ..., Y_N$
- All possible distinct values: $d_1, d_2, \ldots, d_K$
- Model: $\Pr(Y_i = d_k) = \theta_k$
- Prior: $\pi(\theta_1, \theta_2, ..., \theta_k) \propto \prod_k \theta_k^{-1} \text{ if } \sum_k \theta_k = 1$
- Mean and Variance:

$$E(Y_i \mid \theta) = \mu = \sum_k d_k \theta_k$$

$$Var(Y_i \mid \theta) = \sigma^2 = \sum_k d_k^2 \theta_k - \mu^2$$

Bayesian Nonparametric Inference (continued)

- SRS of size n with n_k equal to number of d_k in the sample
- Objective is to draw inference about the population mean: $Q = f \times \bar{y} + (1 - f) \times \bar{Y}_{\text{exc}}$
- As before we need the posterior distribution of μ and σ^2

Nonparametric Inference (continued)

- Posterior distribution of θ is Dirichlet:

$$\pi(\theta \mid Y_{\text{inc}}) \propto \prod_k \theta_k^{n_k - 1} \text{ if } \sum_k \theta_k = 1 \text{ and } \sum_k n_k = n$$

- Posterior mean, variance and covariance of θ

$$E(\theta_k \mid Y_{\text{inc}}) = \frac{n_k}{n}, Var(\theta_k \mid Y_{\text{inc}}) = \frac{n_k(n - n_k)}{n^2(n+1)}$$

$$Cov(\theta_k, \theta_l \mid Y_{\text{inc}}) = -\frac{n_k n_l}{n^2(n+1)}$$

Inference for Q

$$E(\mu \mid Y_{\text{inc}}) = \sum_k d_k \frac{n_k}{n} = \bar{y}$$

$$Var(\mu \mid Y_{\text{inc}}) = \frac{s^2}{n}\frac{n-1}{n+1}; s^2 = \frac{1}{n-1}\sum_{i \in \text{inc}} (y_i - \bar{y})^2$$

$$E(\sigma^2 \mid Y_{\text{inc}}) = s^2 \frac{n-1}{n+1}$$

Hence posterior mean and variance of Q are:

$$E(Q \mid Y_{\text{inc}}) = f \times \bar{y} + (1-f)E(\mu \mid Y_{\text{inc}}) = \bar{y}$$

$$Var(Q \mid Y_{\text{inc}}) = (1-f)\frac{s^2}{n}\frac{n-1}{n+1}$$

Posterior Predictive Distribution

$$Sample: y_1, y_2, \ldots, y_n$$
$$Non-sample: y_{n+1}, y_{n+2}, \ldots, y_N$$
$$Predictive\ distribution:$$
$$Pr(y_{n+1,}y_{n+2}, \ldots, y_N \mid y_1, y_2, \ldots, y_n) = \Pr(y_{n+1} \mid y_1, y_2, \ldots, y_n) \times$$
$$\Pr(y_{n+2} \mid y_{n+1}, y_1, y_2, \ldots, y_n) \times \Pr(y_{n+3} \mid y_{n+2}, y_{n+1}, y_1, y_2, \ldots, y_n) \times$$
$$\ldots \times \Pr(y_N \mid y_{N-1}, \ldots, y_{n+1}, y_{n+1}, y_1, y_2, \ldots, y_n)$$

The Polya Urn Model can be used to obtain draws from the posterior predictive distribution (Ghosh and Meeden (1997), Feller (1967))

Simple random Sample with Auxiliary Variables

Ratio and Regression Estimates

- Population: $(y_i, x_i;\ i=1,2,\ldots N)$
- Sample: $(y_i,\ i \in \text{inc},\ x_i,\ i=1,2,\ldots,N)$.

Objective: Infer about the population mean

$$Q = \sum_{i=1}^{N} y_i$$

Excluded Y's are missing values

y_1	x_1
y_2	x_2
.	.
.	.
.	.
y_n	x_n
	x_{n+1}
	x_{n+2}
	.
	.
	.
	x_N

Model Specification

$$(Y_i \mid x_i, \beta, \sigma^2) \sim \text{ind } N(\beta x_i, \sigma^2 x_i^{2g})$$

$$i = 1, 2, ..., N$$

g known

Prior distribution: $\pi(\beta, \sigma^2) \propto \sigma^{-2}$

g=1/2: Classical Ratio estimator. Posterior variance equals randomization variance for large samples

g=0: Regression through origin. The posterior variance is nearly the same as the randomization variance.

g=1: HT model. Posterior variance equals randomization variance for large samples.

Note that, no asymptotic arguments have been used in deriving Bayesian inferences. Makes small sample corrections and uses t-distributions.

Some Remarks

- For large samples, estimate and its variance under nonparametric model assumptions are very nearly the same as those under the normal model assumptions
- For large N, the population size, the finite population quantity is very nearly same as the model parameter ($Q \approx \mu$).
- For large samples,

$$\frac{Q - E(Q \mid Y_{\text{inc}})}{\sqrt{Var(Q \mid Y_{\text{inc}})}} \sim N(0,1)$$

Remarks (Continued)

- Bayesian Interpretation: Summary of the excluded portion of the population has approximate normal distribution conditional on the observed data. *That is Y_{inc} is fixed and Q is random.*
- Frequentist Interpretation: Under repeated sampling, the distribution of estimates of Q. *That is Q is fixed and Y_{inc} is random.*
- For large samples, the frequentist and Bayes will nearly give the same numerical answers but interpretations would differ.

Remarks

- In much practical analysis the prior information is diffuse, and the likelihood dominates the prior information.
- Jeffreys (1961) developed “noninformative priors” based on the notion of very little prior information relative to the information provided by the data.
- Jeffreys derived the noninformative prior requiring invariance under parameter transformation.
- In general,

$$\pi(\theta) \propto |J(\theta)|^{1/2}$$

where

$$J(\theta) = -E\left(\frac{\partial^2 \log f(y \mid \theta)}{\partial\theta\partial\theta^t}\right)$$

Examples of noninformative priors

Normal: $\pi(\mu,\sigma^2) \propto \sigma^{-2}$

Binomial: $\pi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$

Poisson: $\pi(\lambda) \propto \lambda^{-1/2}$

Normal regression with slopes β: $\pi(\beta,\sigma^2) \propto \sigma^{-2}$

In simple cases these noninformative priors result in numerically same answers as standard frequentist procedures

Summary

- Considered Bayesian predictive inference for population quantities
- Focused here on the population mean, but other posterior distribution of more complex finite population quantities Q can be derived
- Key is to compute the posterior distribution of Q conditional on the data and model
  - Summarize the posterior distribution using posterior mean, variance, HPD interval etc
- Modern Bayesian analysis uses simulation technique to study the posterior distribution
- Models need to incorporate complex design features like unequal selection, stratification and clustering