Some issues in item response modeling of multiple-recapture census problems. [a]

Educational Testing Service
Princeton NJ
July 10, 1998

Brian W. Junker
Dept of Statistics
Carnegie Mellon University
Pittsburgh PA 15213
(412) 268-8873
brian@stat.cmu.edu

[a] Thanks to Fienberg, Patz, Basu, Johnson. My research was supported in part by the by the National Institutes of Health, Grant CA54852, and the National Science Foundation, Grants DMS-94.04438 and DMS-97.05032

- Capture-recapture population estimation
  - Motivation: the U.S. Census
- "Dual-systems"
  - Standard capture-recapture
  - What if the interaction is not zero?
- Multiple-recapture census
  - Analogies with educational testing
- An educational testing model
  - Model-fitting and testing issues
  - An example and some results
- New directions — Work in progress
  - A Bayesian formulation
  - Population size estimation as model selection
  - Reversible jump MCMC
  - Searching the World Wide Web
  - What the heck is "no k-way interactions"?

Capture-recapture population estimation

Motivation: *U.S. Census*

- Mandated by U.S. Constitution.
- Taken every 10 years (April 1) since 1790.
- Exhaustive count is the goal; huge and expensive undertaking.
- Allocation based on local counts.
  - * Representatives in U.S. Congress
  - * Tax money to 39,000 state and local governments
  - * *A fixed-sum game!*
- Undercount vs. Overcount.

Darroch, Fienberg, Glonek and Junker (1993, *JASA*).

"Dual-systems" (Capture-Recapture)

- Estimates undercount at local levels.

- Sources (systems, lists, traps).

  * Census Enumeration (E).
  ("exhaustive")

  * Post-Eumeration Survey (P).
  (165,000 households)

- Within a small region:

		(P)	
	n_{ep}	1	0
	1	n_{11}	n_{10}
(E)	0	n_{01}	$n_{00} =?$

# in original census (E):	$n_{11} + n_{10}$	
# added by P.E.S. (P):	n_{01}	Undercount:
# missed by both:	n_{00}	$n_{01} + n_{00}$

Total Population: $n_{11} + n_{10} + n_{01} + n_{00}$

How do we estimate n_{00}?

Standard Capture-Recapture

	$n_{y_e y_p}$	(P) 1	0
(E)	1	n_{11}	n_{10}
	0	n_{01}	n_{00} =?

Assume *independent samples*, so odds ratio=1:

$$1 = \frac{n_{11} n_{00}}{n_{01} n_{10}}$$

Then

$$\hat{n}_{00} = \frac{n_{10} n_{01}}{n_{11}}$$

Equivalent to estimating the log-linear model with no interaction:

$$\log p_{y_e y_p} = \alpha + \beta_1 y_e + \beta_2 y_p$$

- y_e is 1 if you are in the Census (E); else 0.
- y_s is 1 if you are in the PES (P); else 0.
- If we model $\log n_{y_e y_p}$, then $\exp \hat{\alpha} = \hat{n}_{00}$.

What if the interaction is not zero?

- *Log-linear models:*

$$\begin{aligned}\log p_{y_1 y_2 y_3 \ldots} &= \alpha + \beta_1 y_1 + \beta_2 y_2 + \beta_3 y_3 + \ldots \\ &+ \beta_{12} y_1 y_2 + \beta_{23} y_2 y_3 + \ldots \\ &+ \ldots\end{aligned}$$

- *Heterogeneity* $\Rightarrow$ positive interactions.
- *Trap attraction* $\Rightarrow$ positive interactions.
- *Trap avoidance* $\Rightarrow$ negative interactions.
- *List effort dependence* $\Rightarrow$ either positive or negative dependence.
- Need at least *3 lists* to estimate interactions. . .

US Census:

Admin Lists (A)

1

		PES (P)	
		1	0
Census (E)	1	n_{111}	n_{101}
	0	n_{011}	n_{001}

0

		PES (P)	
		1	0
Census (E)	1	n_{110}	n_{100}
	0	n_{010}	n_{000} =?

Multiple-Recapture Census

- Two or more lists or captures
  (Analysis similar to 2-list case)

- Animal population estimation.
  * Burnham and Overton (1978)
  * Seber (1982)
  * Chao (1987)

- Visual scanning problems.
  * Sanathanan (1972,73)

- Survey and epidemiological work.
  * Sekar and Deming (1949)
  * The Diabetes Project (1995)

- US Census.
  * Kadane and Ericksen (1985); Fienberg (1991, ...)
  * Wolter (1986, 91); Freedman (1991)
  * Mulrey and Spencer (1991)
  * Zaslavsky and Wolfgang (1993)

Analogies with educational measurement

Multiple Recapture Census	*Measurement*
• Presence on a list	• Positive item response
• Capture effectiveness	• Item parameters
• Varying catchability	• Latent variable
• *Inference:* n_{000}	• *Inference: items/persons*
• Independent captures within persons, or strata, plus monotone catchability, implies positive dependence among captures (Kadane, Meyer and Tukey, 1992)	• LI, M, $d = 1$ implies conditional association (Holland and Rosenbaum, 1986)

An educational testing model

- Each person has a *capture record* $y = (y_1, y_2, y_3)$, and a *catchability parameter* θ.
- Each list j [1=(E), 2=(P), 3=(A)] has an *effectiveness parameter* β_j
- For $P_j(\theta) = P[y_j = 1|\theta]$, assume

$$\log \frac{P_j(\theta)}{1 - P_j(\theta)} = \theta + \beta_j$$

- In sampling from the population under this model, probabilities in the 3-way table are

$$p(y_1, y_2, y_3) = \int_{-\infty}^{\infty} \prod_{j=1}^{3} P_j(\theta)^{y_j} (1 - P_j(\theta))^{1-y_j} dF(\theta)$$

- By Dutch Identity (Holland, 1990)

$$\log p(y_1, y_2, y_3) = \alpha + \sum_{j=1}^{3} \beta_j y_j + \gamma(y_1 + y_2 + y_3) \tag{DI}$$

where the $\gamma(k)$ "parametrize" the dependence among the lists.

Model fitting and testing issues

- *No 3-way interaction?*

  For the log-linear estimate $\hat{n} = \exp \hat{\alpha}$ to be valid, must assume $y_1 y_2 y_3$ not in model.

  Impose on model, get (for 3 lists):

$$\log p(y_1, y_2, y_3) = \alpha + \sum_{j=1}^{3} \beta_j y_j + \gamma \cdot (y_1 + y_2 + y_3)^2$$

- *Consistent with the mixture model?*

  Need $\gamma > 0$. Check $\hat{\gamma}$.

- *Rasch model tests?*

  $P[Y_i = 1 | Y_i + Y_j = 1, h(\text{other items}) = c]$ independent of c? (Rosenbaum, 1987)

(P vs A) $\coprod$ E?		In PES (P)	In A. Lists (A)
In Census (E)	Yes	n_{110}	n_{101}
	No	n_{010}	n_{001}

An example: Four "Post-strata" in a St. Louis (MO, USA) trial run [Own/Rent, 20's/30's].

Log-odds-ratio	O2	R2	O3	R3
(A vs E) $\coprod$ P	–2.20	–2.04	–2.48	–1.98
(P vs A) $\coprod$ E	2.34	2.00	2.32	2.89
(E vs P) $\coprod$ A	–0.14	0.05	0.15	–0.91

Chi-Squared	O2	R2	O3	R3
(A vs E) $\coprod$ P	66.50	37.35	58.30	38.36
(P vs A) $\coprod$ E	74.84	37.14	57.09	63.19
(E vs P) $\coprod$ A	0.15	0.01	0.11	3.44

- (E vs P) $\coprod$ A can be achieved by

$$\log \frac{P_j(\theta_1, \theta_2)}{1 - P_j(\theta_1, \theta_2)} = \begin{cases} \theta_1 + \beta_i, & i = 1, 2 \\ \theta_2 + \beta_3, & i = 3 \end{cases}$$

- Integrate out θ's, get:

$$\log p(y_1, y_2, y_3) = \alpha + y_1\beta_1 + y_2\beta_2 + y_3\beta_3 + \gamma(y_1 + y_2, y_3)$$

where $\gamma(2, 0) \geq 2\gamma(1, 0)$ and $\gamma(2, 1) + \gamma(0, 1) \geq 2\gamma(1, 1)$.

- Administrative lists (A) are "different".

Results

Strata	O2	R2	O3	R3
$\hat{x}_{000}$: Saturated	290.06	670.47	496.83	825 .97
S.E. (delta method)	118.96	335.01	247.31	449.62
Fit / d.f.	—/0	—/0	—/0	—/0
$\hat{x}_{000}$: 2-dim Rasch	291.55	669.50	489.84	767.80
S.E. (delta method)	119.46	334.28	242.12	402.86
Fit / d.f.	0.15/1	0.01/1	0.11/1	3.45/1
$\hat{x}_{000}$: Rasch	46.69	102.06	42.33	81.91
S.E. (delta method)	14.40	35.76	13.41	28.72
Fit / d.f.	94.04/2	52.99/2	67.89/2	73.01/2
$\hat{x}_{000}$: Independent Lists	44.64	48.26	27.51	34.98
S.E. (delta method)	5.69	6.78	3.80	5.06
Fit / d.f.	94.06/3	58.34/3	70.15/3	80.06/3

Some preliminary conclusions

- Rasch heterogeneity is an improvement over independence, but may not be enough.
- Allowing interactions between lists and the total number of captures y_+ turns out to be useful.
- Sometimes these interactions can be interpreted as evidence of "latent multidimensionality"
- Sometimes they simply model dependence in the lists that the monotone unidimensional latent variable models can't see. (Cf. e.g. Diabetes Project, 1995a,b); Fienberg and Carriquiry, 1997).

Related work (!)

- Stegelman (1983, *Psy*): Multidimensional Rasch.
- Fischer (1983, *Psy*; ...): LLTM (stratum variables).
- More recently, Wilson, Adams and Wu (1997; ConQuest).
- Kelderman (1992, *Psy*; ...): Log-Linear Rasch. Also Lindsay, Clogg and Grego (1991, *JASA*); Agresti and Lang (1993 *Stat Med?*).
- Fienberg (1992, *Statistics Canada*): Survey of literature in Census, Animal and Epidemiology settings.
- Diabetes Project (1995a,b, *Am. J. Epi.*). Yip, Bruno, Tajima, Seber, Buckland, Cormack, Unwin, Fienberg, Junker, Chang, LaPorte, Libman, McCarty.
- Fienberg and Carriquiry (1997, *Ency. Biostat.*): Innovative use of Rasch model and its generalizations.
  See also Fienberg (1998). Biometric Society/ENAR short course.

New directions — Work in progress

- A Bayesian formulation
- Population size estimation as model selection
- Reversible jump MCMC
- Searching the World Wide Web
- What the heck is "no k-way interactions"?

A Bayesian formulation

- Let $y_{ij} = 1$ if person i on list j; 0 else.
  Let $p_{ij} = P[y_{ij} = 1]$ ($\approx$ ICC's)

- Let $\mathcal{Y}_{full}$ be the complete N-row 0/1 matrix;
  Let $\mathcal{Y}_{obs}$ be the n rows of $\mathcal{Y}_{full}$ with $y_+ > 0$;
  Let $\mathcal{Y}_{miss}$ be the rows with $y_+ = 0$.

Then

$$P[\mathcal{Y}_{full}|N, \underline{p}] = \prod_{i=1}^{N} \prod_{j=1}^{k} p_{ij}^{x_{ij}} (1 - p_{ij})^{1-x_{ij}}$$

$$= \prod_{i \in \mathcal{Y}_{obs}} \prod_{j=1}^{k} p_{ij}^{x_{ij}} (1 - p_{ij})^{1-x_{ij}} \cdot \prod_{i \in Y_{miss}} \prod_{j=1}^{k} (1 - p_{ij})$$

and

$$\begin{aligned} P[\mathcal{Y}_{obs}|N, \underline{p}] &= \sum_{\mathcal{Y}_{miss}} P[\mathcal{Y}_{obs}, \mathcal{Y}_{miss}|N, \underline{p}] \\ &= \binom{N}{n} \prod_{i=1}^{N} \prod_{j=1}^{k} p_{ij}^{x_{ij}} (1 - p_{ij})^{1-x_{ij}} \end{aligned}$$

George and Robert (1992, *Bmka*); Basu (1997a,b *NIU tech rept's*).

Now it seems straightforward... For a prior $\pi(N, \underline{p})$,

$$\begin{aligned} P(N, \underline{p}|\mathcal{Y}_{obs}) &\propto P(\mathcal{Y}_{obs}|N, \underline{p})\pi(N, \underline{p}) \\ P(N|\mathcal{Y}_{obs}) &\propto \int P(\mathcal{Y}_{obs}|N, \underline{p})\pi(N, \underline{p})d\underline{p} \end{aligned}$$

But *the number of nuisance parameters is* $dim(\underline{p}) = Nk$*, which depends on the estimand* N*!*

Even for additive models like $\log p_{ij} = \theta_i - \beta_j$ or $\text{logit}(p) = \theta_i - \beta_j$, we still have $dim(\underline{p}) = N + k$.

Related problem:

- We can't condition on all of $\underline{p}$ when estimating N, because the dimension of the $\underline{p}$-matrix fixes N!

- In the "$\theta - \beta$" models above, it's OK to condition on $\underline{\beta}$ when estimating N (or drawing from a "complete conditional" in MCMC) *but we better not condition on* $\underline{\theta}$*!*

Basu proposes two MCMC estimation solutions:

1. Fix an absolute maximum M, maintain $\theta_1, \ldots, \theta_M$; N of the θ's simulated from complete conditionals given $\mathcal{Y}_{obs}$, and $M - N$ from their priors;

2. Observe that the complete conditional for $(N, \underline{\theta})$ has the form

$$P[N, \underline{\theta} | rest] \quad = \quad P[\underline{\theta} | N, rest] \cdot P[N | rest \ \ without \ \ \underline{\theta}]$$

First method is Basu's straw-man. Expect it not to converge well, because when $M - N$ is large, we are constantly re-introducing high-dimensional noise from the θ-prior into the complete conditionals of the other parameters.

Second method is clever, and really works!

In Basu's second method:

$$P[N, \underline{\theta}|rest] \quad = \quad P[\underline{\theta}|N, rest] \cdot P[N|rest\ \ without\ \ \underline{\theta}]$$

- $P[\underline{\theta}|N, rest]$ is just the complete conditional for $\underline{\theta}$ in IRT; has the nice forms we've seen before
- $P[N|rest\ \ without\ \ \underline{\theta}]$ has a nice marginal form,

$$P[N|rest\ \ without\ \ \underline{\theta}] \ \propto \ \pi(N) \binom{N}{n} \overline{p}_{\underline{0}}{}^{N-n}$$

  where

$$\overline{p}_{\underline{0}} = \int \prod_{j=1}^{k} (1 - P(\theta, \beta_j)) dF(\theta|\lambda)$$

  and $\pi(N)$ is the marginal prior on N.

- This solution is easily implementable as long as θ has a low-dimensional discrete distribution.

Population size estimation as model selection

A new way to look at the problem:

$$P[\mathcal{Y}_{obs}|N,\underline{p}] = \binom{N}{n} \cdot \prod_{i=1}^{N}\prod_{j=1}^{k} p_{ij}^{x_{ij}}(1-p_{ij})^{1-x_{ij}}$$

$$= \binom{N}{n} \cdot \prod_{i=1}^{n}\prod_{j=1}^{k} p_{ij}^{x_{ij}}(1-p_{ij})^{1-x_{ij}} \cdot \prod_{i=n+1}^{N}\prod_{j=1}^{k}(1-p_{ij})$$

defines a different <u>*model*</u> for $\mathcal{Y}_{obs}$, for each N.

- When $p_{ij} = P(\theta_i, \beta_j)$, the number and dimension of β's stays the same from model to model, but as N increases and decreases, the number of θ's changes.
- The last factor above functions like a partial prior on the β's with prior sample size $N - n$.

  E.g. the more persons with $(0, 0, \ldots, 0)$ capture record $\Rightarrow$ the catch effort was smaller (the items were harder).

Reversible jump MCMC Green (1995, *Bmka*).

- Countable collection of models $\mathcal{M}_N$, with parameters $\tau^{(N)} \in \Re^{d_N}$, $d_N \nearrow$ with N:

$$p(y, \tau^{(N)}, \mathcal{M}_N) \quad = \quad p(y|\mathcal{M}_N, \tau^{(N)})\, p(\tau^{(N)}|\mathcal{M}_N)\, p(\mathcal{M}_N)$$

- Complete conditionals

$$\begin{aligned} p(\tau^{(N)}|rest) &\propto p(y|\mathcal{M}_N, \tau^{(N)})\, p(\tau^{(N)}|\mathcal{M}_N) \\ p(N|rest) &\propto p(y|\mathcal{M}_N, \tau^{(N)})\, p(\tau^{(N)}|\mathcal{M}_N)\, p(\mathcal{M}_N) \\ &\equiv p(N; y, \tau^{(N)}) \end{aligned}$$

- For Metropolis step, need
  - Proposal $N^* \sim q_k(N|N^{(k-1)})$
  - Acceptance ratio α^*:

$$\min\left\{\frac{p(N^*; y, \tau^{(N^*)})\, q_k(N^{(k-1)}|N^*)}{p(N^{(k-1)}; y, \tau^{(N^{(k-1)})})\, q_k(N^*|N^{(k-1)})}, 1\right\}$$

- Clearly need map $\phi : (\tau^{(N)}, \textit{aux var's}) \to (\tau^{(N+1)})$

- "Detailed balance" holds when $\phi(\cdot)$ is a bijection.

Currently...

Grad Student Matt Johnson (with Fienberg, me):

- Adapting Basu's 2^{nd} method to Rasch-like (additive logit) models; Basu did only log-additive case

- Developing reversible-jump code for the same problem

Goal: estimation methodology that allows:

- A more flexible model specification in the latent space, than multidimensional Rasch

- More flexible "ICC's", e.g. the whole 2PL, etc. litany

- Local dependence structures ("testlests": list$\times$list interactions that don't go away, conditional on θ); list$\times\theta$ interactions

- Broader model fit analysis (e.g. PPC's)

- Etc. Etc.

Searching the World Wide Web (Lawrence and Giles, 1998 *Science*)

How many web pages are there?

Look at 6 search engines, aggregated across 575 queries on scientific topics.

Basic pairwise cap-recap analysis of the search engines (from Lawrence and Giles, 1998):

Engines	$\hat{N}$ indexable pages ($\times 10^6$)	95% CI
Lycos and Infoseek	90	$\pm$ 6
Infoseek and Excite	220	$\pm$ 16
Excite and Northern Light	230	$\pm$ 15
Northern Light and AltaVista	230	$\pm$ 13
AltaVista and Hotbot	320	$\pm$ 34

Estimates of indexable web, from pairs of search engines, from the pair with the smallest coverage relative to the total observed pages n *to pair with the largest coverage.*

- We hope to try our methods soon with the raw data underlying the *Science* article
  - With six lists we can push our model development strategy
  - Should we be aggregating across the queries or stratifying on them?
  - Etc.
- Lawrence and Giles intend to repeat the experiement every six months
  - Analyze the data; monitor growth and decline of the WWW
  - Insights into designing the experiment...
    * Are these the right queries?
    * Do we need to do so many?
    * Are these the right search engines?
    * Other sources?

What the heck is no k-way interaction?

- In the log-linear Rasch approach (first part of the talk) it was necessary to impose the restriction *"no k-way interaction"* to project the fitted model onto the missing cell, to get an estimate of $N - n$.

- In the Bayesian formulation, this is not explicitly imposed; yet sensible estimates can be obtained (e.g. Basu's work; preliminary reversible-jump results).

- Thus the prior constraints on the parameters are imposing enough structure on the marginal distribution of the data, to "project to the missing cell".

- What is the relationship between *"no k-way interaction"* and the prior constraints in the Bayesian formulation?