Bayesian Consumer Profiling

Arnaud De Bruyn
ESSEC Business School

Thomas Otter
Goethe University Frankfurt

ESSEC BUSINESS SCHOOL

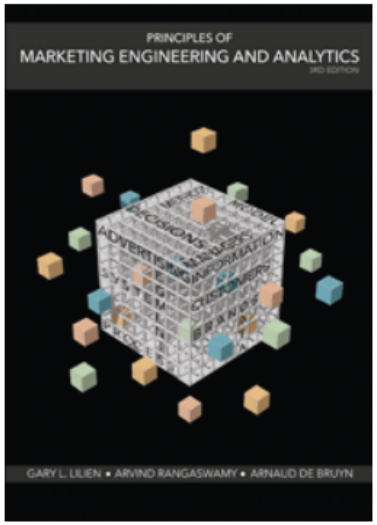
GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Illustration: Who visits my website?

DecisionPro

WHAT IS "MARKETING ENGINEERING"?

Several forces are transforming the nature, scope, and structure of the marketing profession. Marketers are seeing increasingly faster changes in the marketplace and are barraged with an ever increasing amount of information. While many view traditional marketing as art and some view it as science, the new marketing increasingly looks like engineering. "Marketing Engineering" is a term coined by Professor Gary Lilien, one of the founders of DecisionPro, Inc, to describe this new marketing philosophy.



DecisionPro, Inc. has developed a series of books and software products to further the field of analytical marketing. These books, combined with a comprehensive collection of leading-edge software models provides the know-how and tools to collect the right information and perform analysis to make better marketing plans, better product designs, and better decisions.

The purpose in writing these books is to help educate and train a new generation of marketing managers. Our goal is to train marketing engineers to translate concepts into context-specific operational decisions and actions using analytical, quantitative, and computer modeling techniques. We link theory to practice and practice to theory.

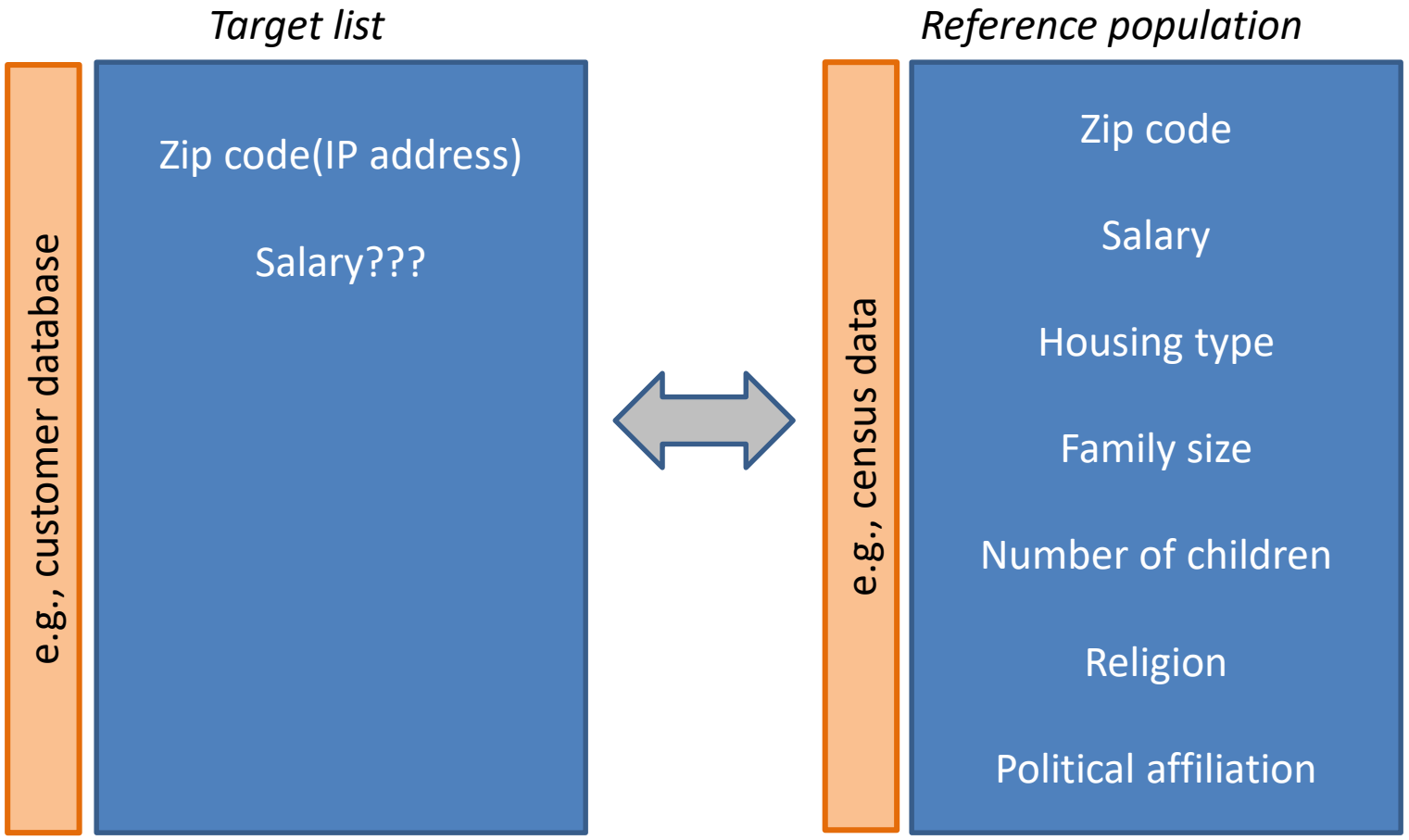
DOWNLOAD DEMO

Download a FREE demo of Marketing Engineering for Excel to learn how easy it is to introduce analytics into your marketing efforts!

SOFTWARE MODELS

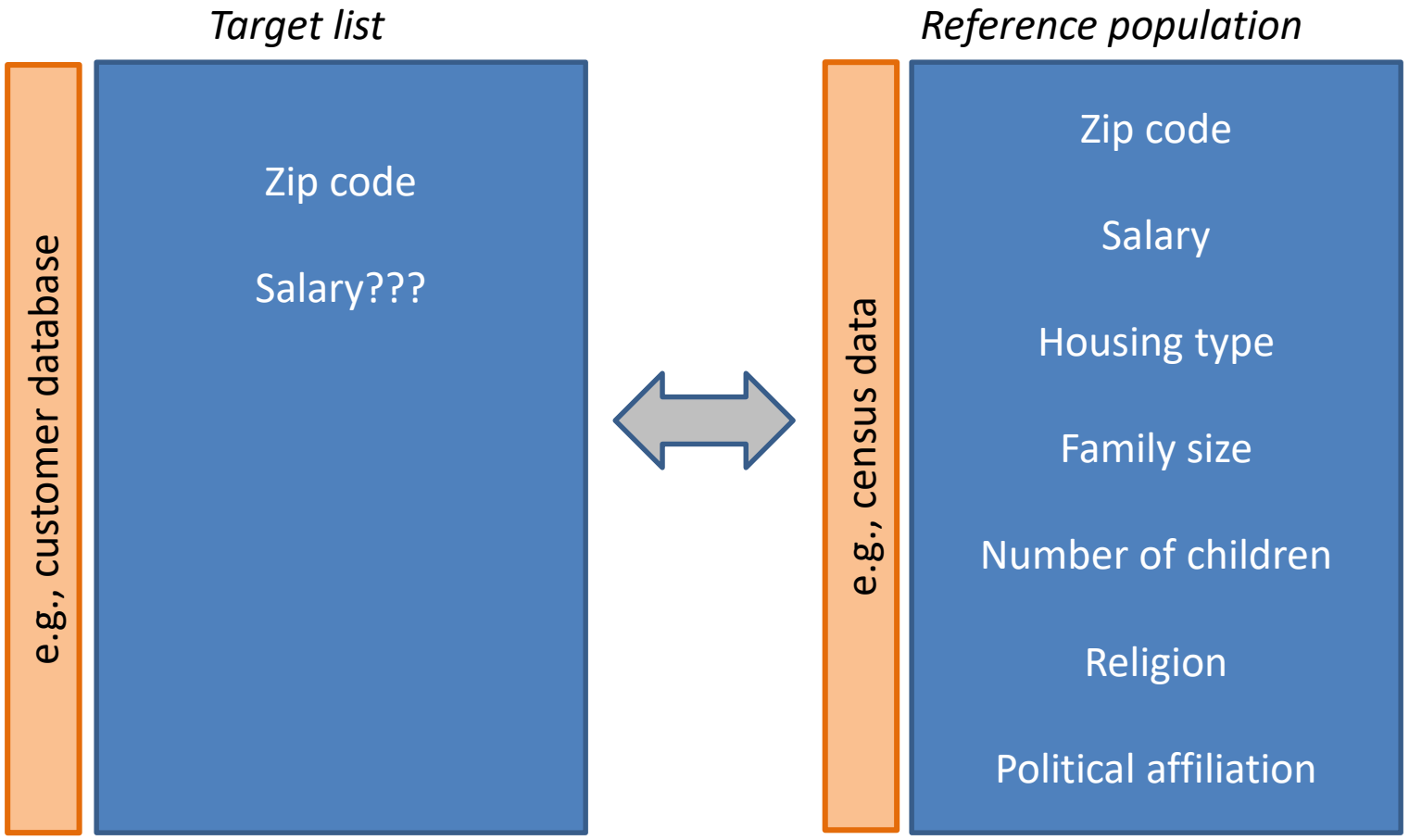
- BASS FORECASTING MODEL
- CONJOINT ANALYSIS MODEL
- CUSTOMER CHOICE MODEL
- CUSTOMER LIFETIME VALUE MODEL
- GE/MCKINSEY PORTFOLIO MATRIX MODEL

Problem setup



Traditional solution: The simple count method

- Use p(Zip code, salary) in the reference population to compute p(salary | Zip code)
- Integrate over Zip code distribution in the target list to infer the salary distribution in the target list



Simple count method

- Probability to fall in salary category *j* given zip code *X* in the reference population *Y*

$$p(j|X) = \frac{Y_{j,X}}{Y_{.X}}$$

Simple count method

- Proportion of individuals in target list in salary category *j*

$$p(j|L) = \frac{1}{I}\sum_{i=1}^{I}\frac{Y_{j,X_i}}{Y_{.,X_i}}$$

Who visits my website?

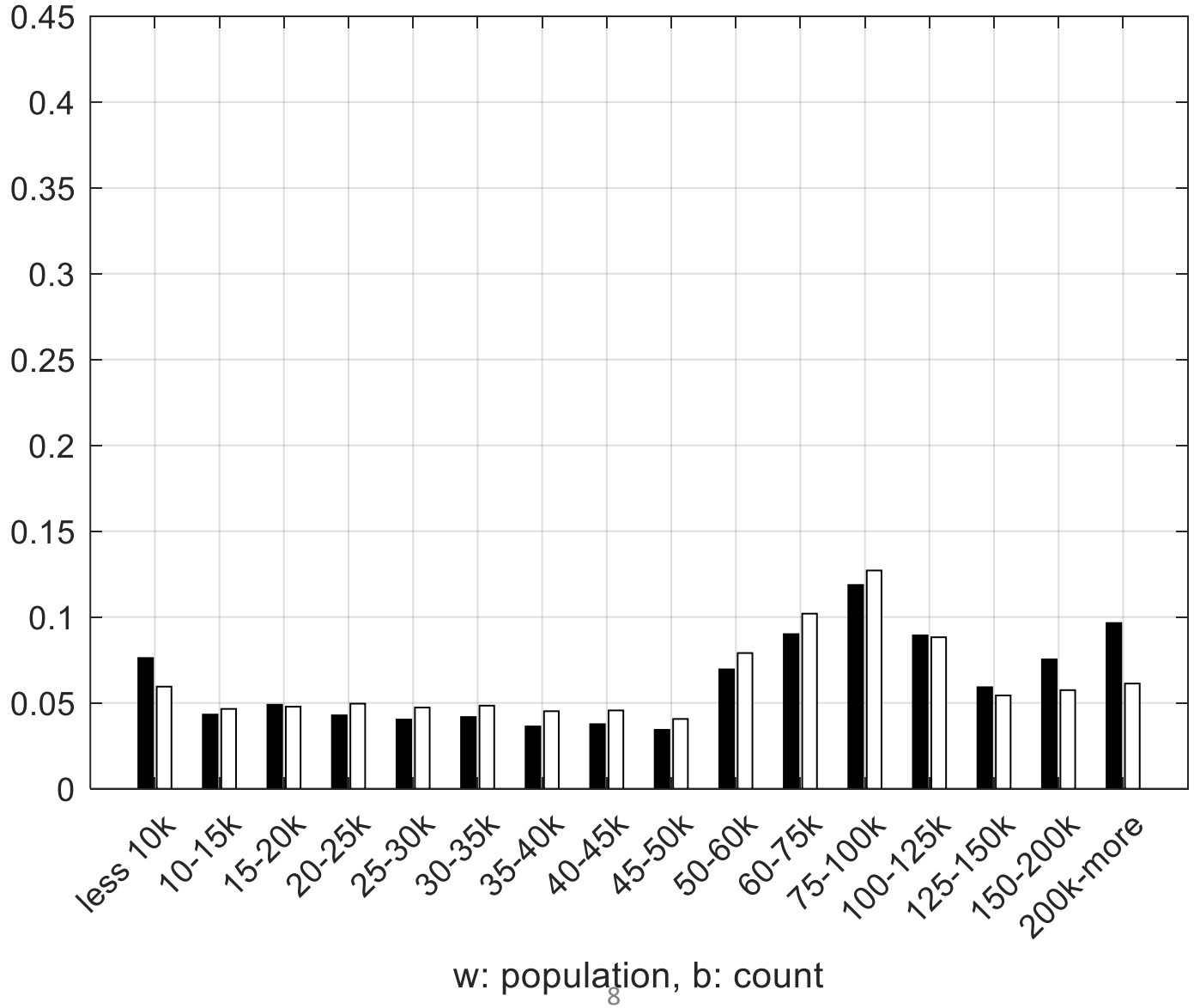


Illustration: French elections

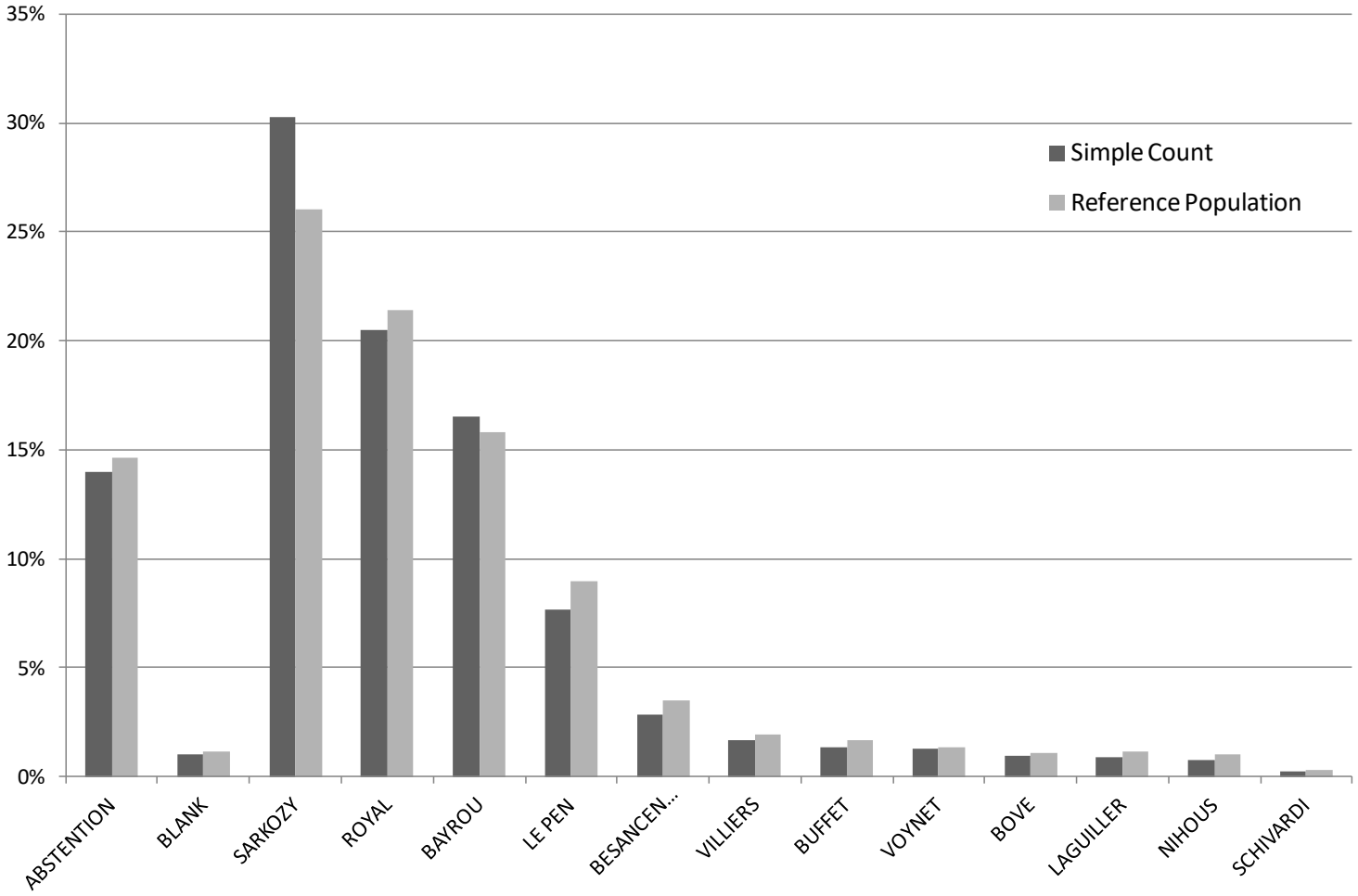
Data set

- 2007 presidential elections in France
  - 2-round elections
  - Nicolas Sarkozy won the second round
  - 12 candidates competed in the first round

- Target list
  - 18,981 individuals, private political group with shared preferences for candidates
  - Zip codes only
  - Can we determine the prevailing political preference?

- Reference list
  - Detailed results of the elections for each of the 36,239 voting districts (Interior Minister's Web site)

Results


35%
30%
25%
20%
15%
10%
5%
0%
Simple Count
Reference Population
ABSTENTION
BLANK
SARKOZY
ROYAL
BAYROU
LE PEN
BESANCEN...
VILLIERS
BUFFET
VOYNET
BOVE
LAGUILLER
NIHOUS
SCHIVARDI

A curious result?

List selection mechanisms

- *Simple random sampling*: $p(i \in L) = p(i \in L|X_i, S_i)$

- *Selection based on observed X:*

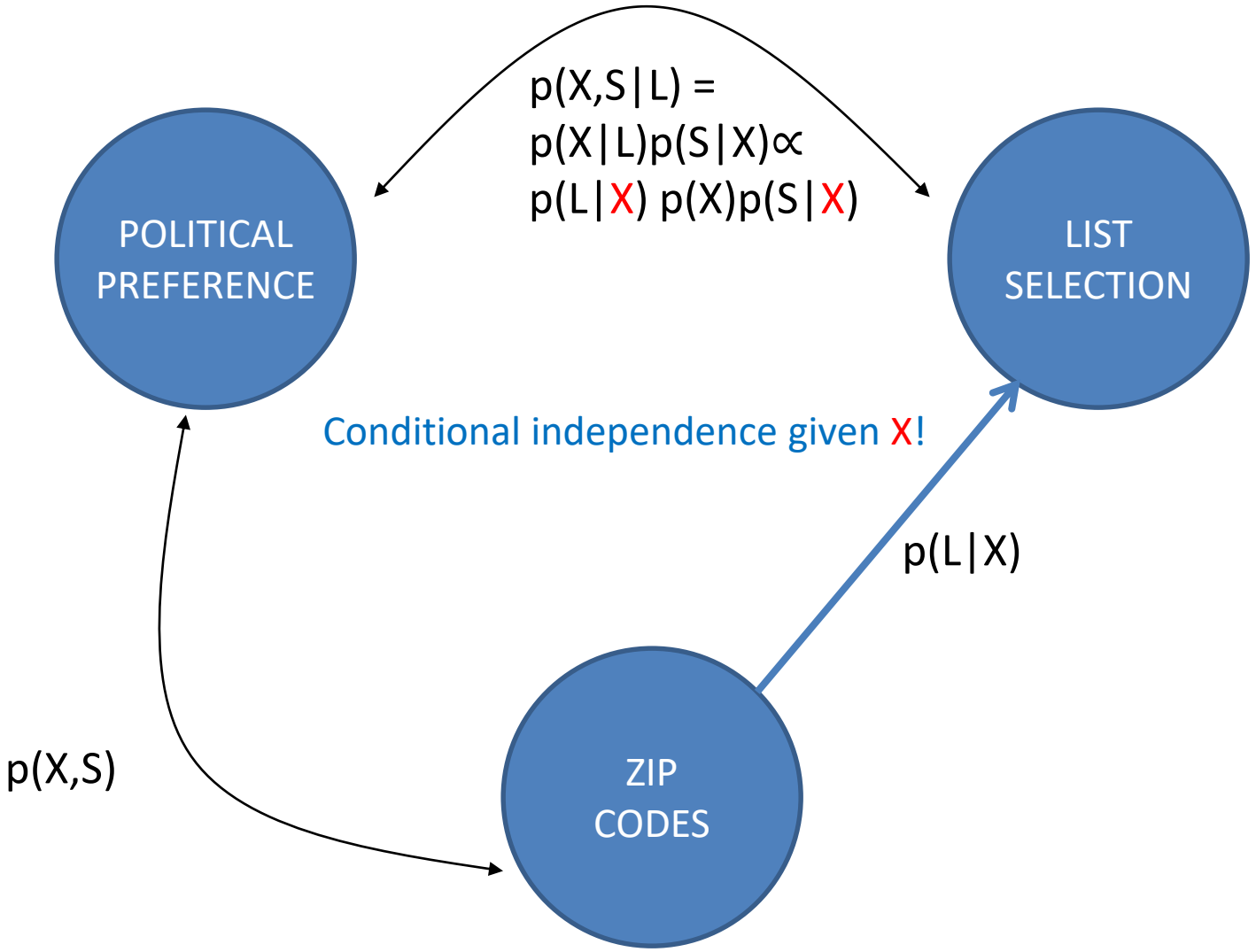
  $p(i \in L) \neq p(i \in L|X_i, S_i)$ and
  $p(i \in L|X_i) = p(i \in L|X_i, S_i) \rightarrow$
  $p(X, S|L) = p(X|L) \times p(S|X)$

- *Selection based on unobserved S:*

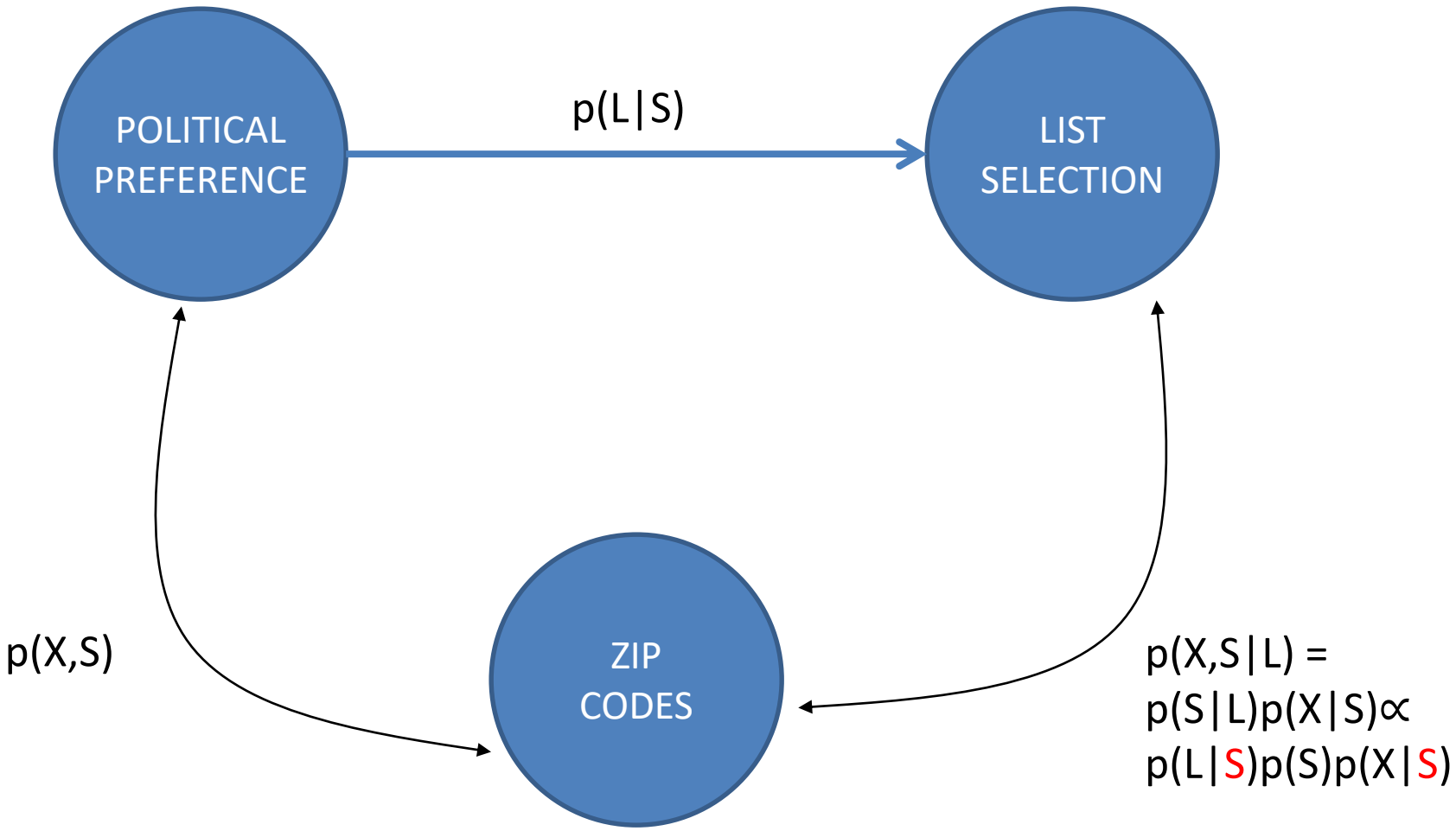
  $p(i \in L) \neq p(i \in L|X_i, S_i)$ and
  $p(i \in L|S_i) = p(i \in L|X_i, S_i) \rightarrow$
  $p(X, S|L) = p(S|L) \times p(X|S)$

- *Selection based on both X and S*

Selection based on X – The simple count method



Selection based on unobserved S



Conditional independence given S!

Bayesian profiling for selection based on unobserved S

Likelihood of observed X

By integrating out unobserved S

$$p(X = k|L) = \sum_{j=1}^{J} p(X = k, S = j|L) = \sum_{j=1}^{J} p(S = j|L) \times p(X = k|S = j)$$

Use

$$p(S = j|L) = \frac{p(L|S = j) \times p(S = j)}{p(L)} = \frac{p(L|S = j) \times p(S = j)}{\sum_{j=1}^{J} p(L|S = j) \times p(S = j)}$$

to rewrite as

$$p(X = k|L) = \sum_{j=1}^{J} \frac{p(L|S = j) \times p(S = j)}{\sum_{j=1}^{J} p(L|S = j) \times p(S = j)} \times p(X = k|S = j)$$

Likelihood of observed X

$$\ell\left(\#(X=1),\ldots,\#(X=K)\middle|w_1,\ldots w_J,L\right)=$$

$$\prod_{k=1}^{K}\left(\sum_{j=1}^{J}\frac{w_j\times p(S=j)}{\sum_{j=1}^{J}w_j\times p(S=j)}\times p(X=k|S=j)\right)^{\#(X=k)}$$

Prior structure

Unordered categories: lognormal prior with mode at 1 for each *w*

Ordered categories: beta-shape prior on the (ordered) weights with lognormal hyperpriors on *a* and *b in beta(a,b)*

Likelihood information / (econometric) identification

X variables with more categories and a conditional distribution $p(X|S)$ in the reference population that differs more from the corresponding marginal distribution $p(X)$ increase the statistical information about the weights

Who was supported by list members?

	National Averages	Simple Count Method	Bayesian Profiling	
			Estimate	*Post S.D.*
ABSTENTION	0.162	0.140	0.000	*< 0.001*
BLANK	0.012	0.010	0.001	*< 0.001*
SARKOZY	**0.257**	**0.303**	**0.999**	*< 0.001*
ROYAL	0.214	0.205	0.000	*< 0.001*
BAYROU	0.153	0.165	0.000	*< 0.001*
LE PEN	0.086	0.076	0.000	*< 0.001*
BESANCENOT	0.034	0.028	0.000	*< 0.001*
VILLIERS	0.018	0.017	0.000	*< 0.001*
BUFFET	0.016	0.013	0.000	*< 0.001*
VOYNET	0.013	0.013	0.000	*< 0.001*
BOVE	0.011	0.010	0.000	*< 0.001*
LAGUILLER	0.011	0.009	0.000	*< 0.001*
NIHOUS	0.009	0.008	0.000	*< 0.001*
SCHIVARDI	0.003	0.002	0.000	*< 0.001*

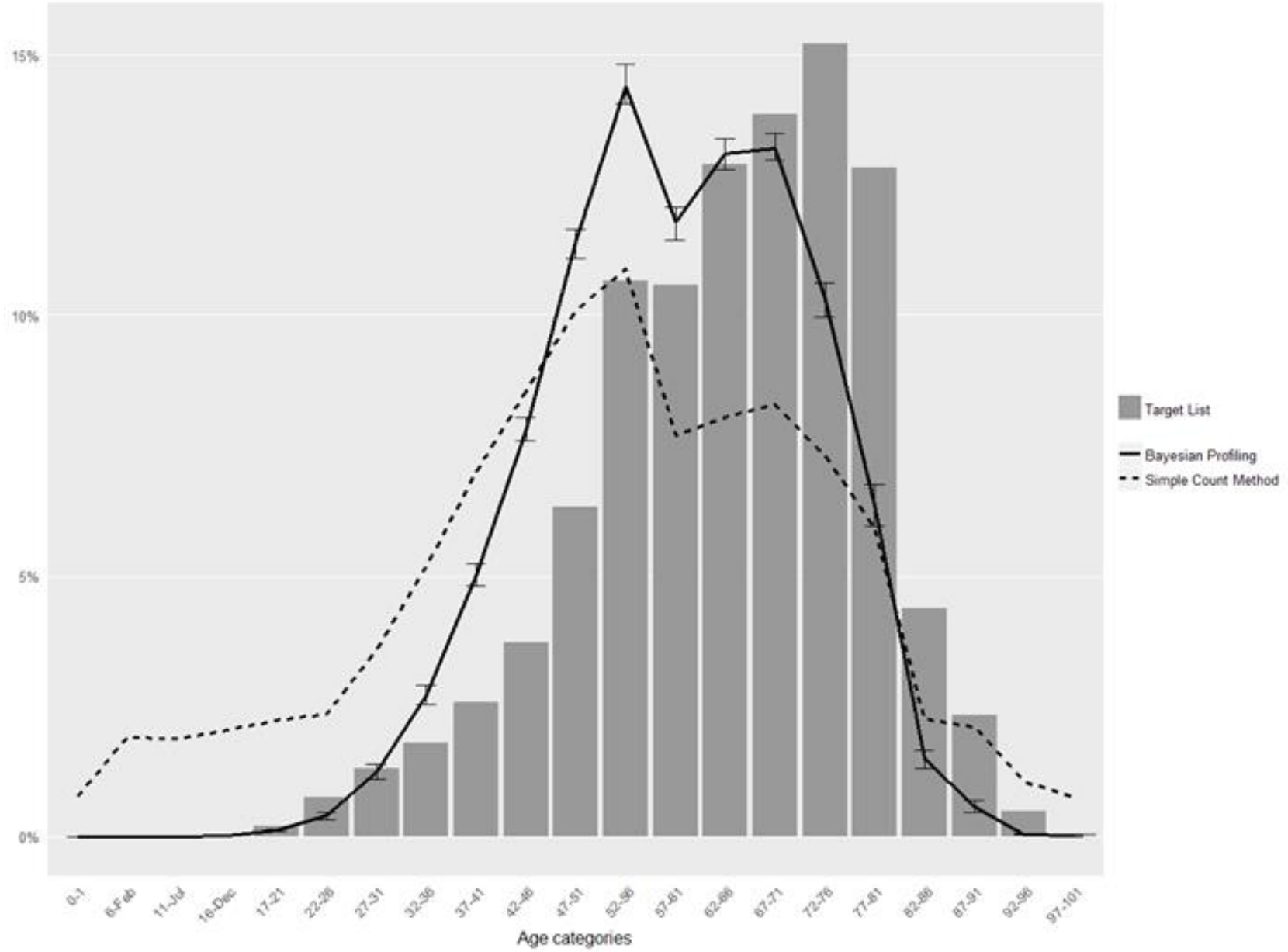
Table 11 - Abstentions, blank votes, and valid votes per candidate in the first round of the 2007 French presidential elections: national averages (leftmost column) versus target list estimates using the simple count and the Bayesian profiling methods. Although actual voting behavior of the target list is unknown, list members are expected to be extremely loyal to the candidate Sarkozy, a phenomenon that is predicted with striking accuracy by the Bayesian method.

Illustration #2: age profiling using first names

Age profiling using first names

- Empirical illustration
  - 14,046 individuals
  - List originates from the private banking sector
  - Highly educated, wealthy
- Data
  - First name
- Want to know
  - Age
- Methods
  - Simple count method
  - Bayesian profiling approach

Does it work?


15%
10%
5%
0%
0-1
6-Feb
11-Jul
16-Dec
17-21
22-26
27-31
32-36
37-41
42-46
47-51
52-56
57-61
62-66
67-71
72-76
77-81
82-86
87-91
92-96
97-101
Age categories
Target List
Bayesian Profiling
Simple Count Method

Does it work?

	Pearson's R	χ^2	RMSE	Log-predictive-likelihood	Critical errors
Reference Population	-.092	22,971	5.97%	44,505	21.8%
Simple Count Method	.750	6,698	3.60%	37,407	6.6%
Bayesian Profiling	.880	4,573	2.61%	35,332	>0.1%

Targeting in the list

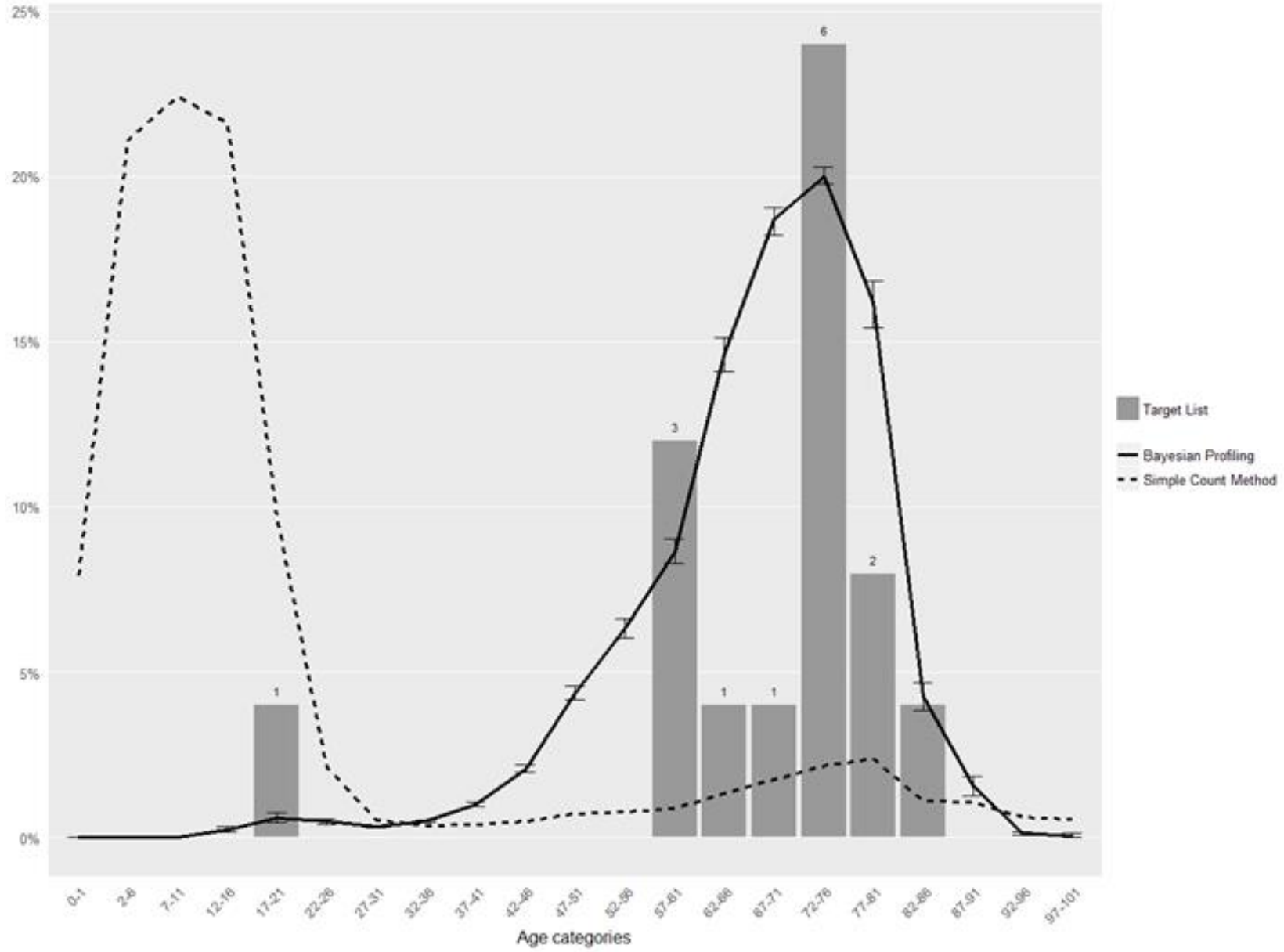
$$p\left(S = i \middle| w_1, \dots, w_J, X_i = k, L\right)$$

$$= \frac{w_i \times p(S = i) \times p(X = k|S = i)}{\sum_{j=1}^{J} w_j \times p(S = j) \times p(X = k|S = j)}$$

$$= \frac{w_i \times p(S = i|X = k)}{\sum_{j=1}^{J} w_j \times p(S = j|X = k)}$$

→ **Equivalent to simple count for all** w**=constant**

‘Targeting Adrien’

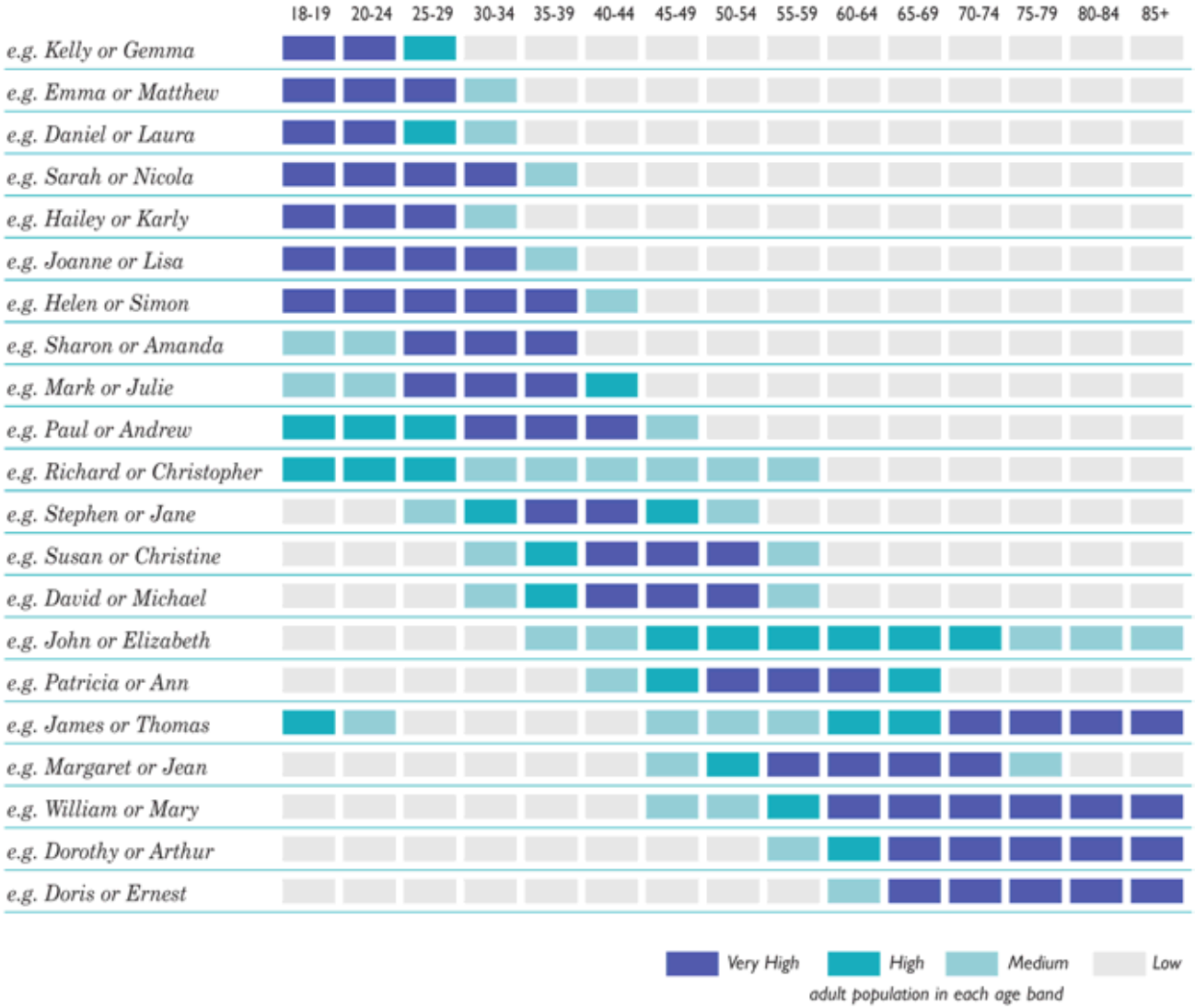


Can we do better?

- Recall what we know about the data:
  - 14,046 individuals
  - List originates from the private banking sector
  - (Highly educated, wealthy)

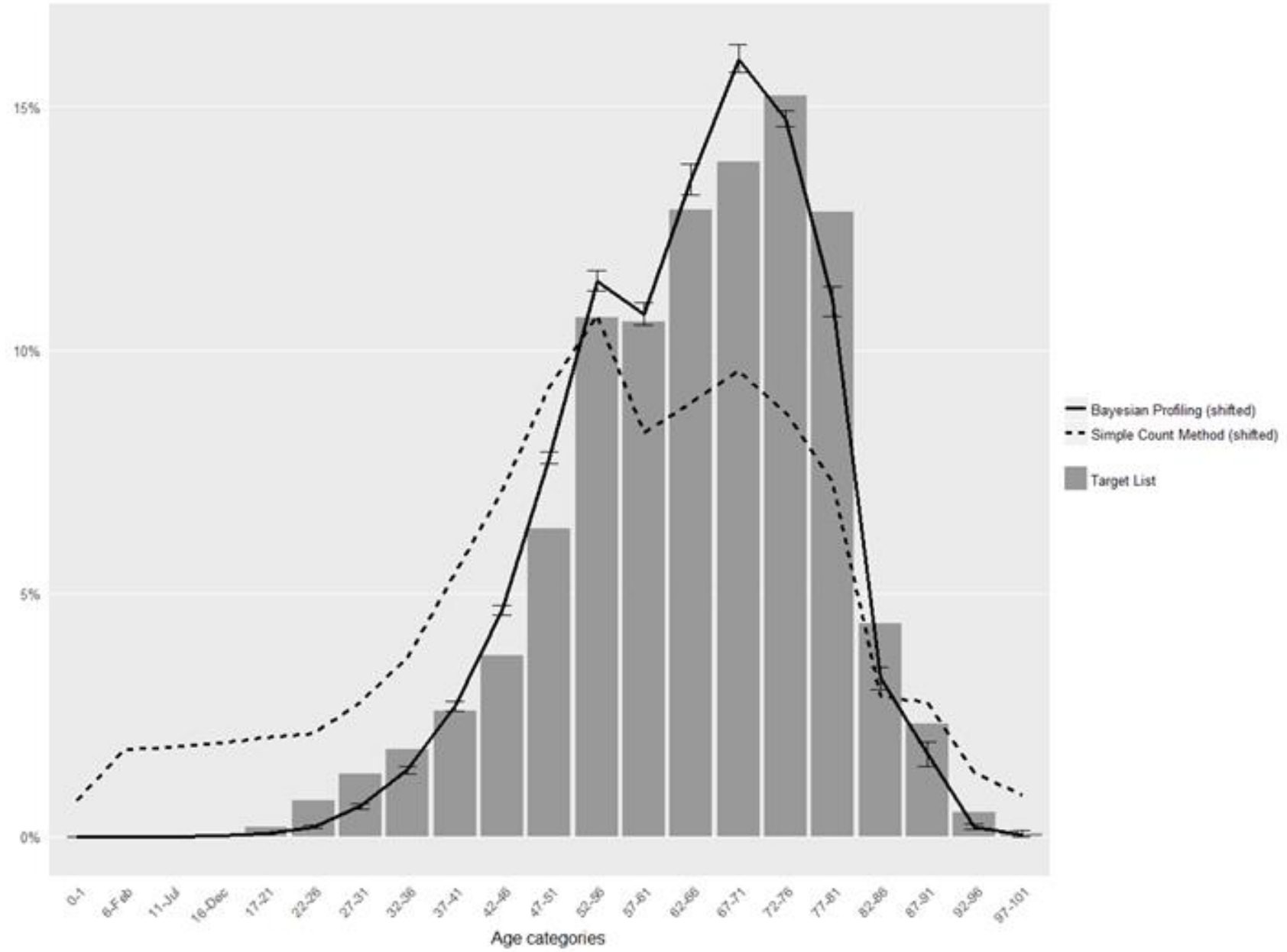
- Private banking → ideally condition on education, wealth, and age as unobserved selection criteria
- However, joint distribution with first names not available to us
- Alternative approach: Condition the first name – age reference table on the naming behaviors of educated wealthy families → like conditioning on above average wealth and education as *observed* selection criterion
- Naming behavior is ahead of the general population by about 4-5 years.

→ Shift age by one category to the left in the ***reference table***



Most likely age of typical first names in the United Kingdom.

‘Conditioning on known wealth’



Does it work better?

	Pearson's R	χ^2	RMSE	Log-predictive-likelihood	Critical errors
Reference Population	-.092	22,971	5.97%	44,505	21.8%
Simple Count Method	.750	6,698	3.60%	37,407	6.6%
Bayesian Profiling	.880	4,573	2.61%	35,332	>0.1%

With conditional reference table:

	Pearson's R	χ^2	RMSE	Log-predictive-likelihood	Critical errors
Simple Count Method	.889	4,234	2.81%	36,262	6.30%
Bayesian Profiling	.989	679	0.83%	33,901	0.01%

Who believes in the model you fit?

- Can we empirically distinguish b/w models, e.g. selection based on X versus selection based on S, or selection based on $S^{(1)}$ versus $S^{(2)}$ or a combination?
- A comparison to "random selection" seems straightforward (this is the nested model with all weights constrained to equality)
- Use Bayesian model comparison techniques building on the idea of marginal likelihoods, i.e., the density of the data given a model, e.g., p(*X* | selection on *S*) instead of p(*X* | *w*-hat, selection on *S*). More formally:

$$p(\boldsymbol{X}|model\,) =$$
$$= \int \ell\big(\#(X=1),\dots,\#(X=K)\big|w_1,\dots,w_J,L\big)p(w_1,\dots,w_J)d\{w_1,\dots,w_J\}$$

Nice? – But what is the likelihood corresponding to the simple count method?

$$\ell(\#(X=1),\ldots,\#(X=K)\big|w_1^X,\ldots,w_K^X,L)$$
$$=\prod_{k=1}^{K}\left(\frac{w_k^X\times p(X=k)}{\sum_{j=1}^{K}w_j^X\times p(X=j)}\right)^{\#(X=k)}$$

Compare to Bayesian profiling:

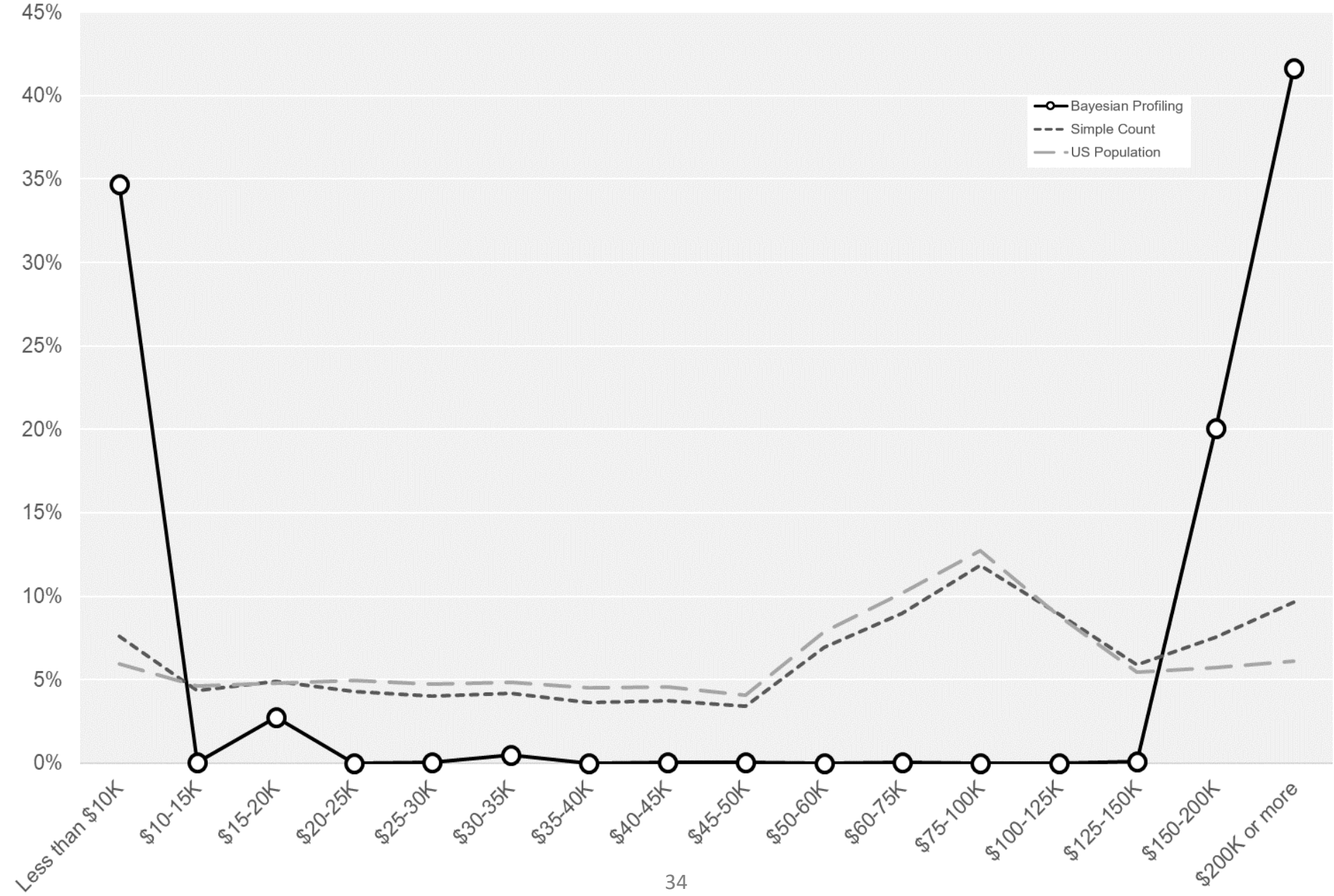
$$\ell\left(\#(X=1),\ldots,\#(X=K)\big|w_1,\ldots w_J,L\right)=$$
$$\prod_{k=1}^{K}\left(\sum_{j=1}^{J}\frac{w_j\times p(S=j)}{\sum_{j=1}^{J}w_j\times p(S=j)}\times p(X=k|S=j)\right)^{\#(X=k)}$$

Illustrative simulation results

100 X-levels [.01, .02, .03, …, .99, 1], 5 S-levels [.9, 0, -.9, .1, .9]

Model List		Random selection	Selection based on X	Selection based on S
List selected **on** ***S***	LML	-2,189,069	-2,189,067	-2,188,756
	Log posterior model probs	-313	-310	0
List selected **on** ***X***	LML	-2,179,500	-2,139,125	-2,140,486
	Log posterior model probs	-4,0375.5	0	-1,361

Illustration: Who visits my website?


45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
Bayesian Profiling
Simple Count
US Population
Less than $10K
$10-15K
$15-20K
$20-25K
$25-30K
$30-35K
$35-40K
$40-45K
$45-50K
$50-60K
$60-75K
$75-100K
$100-125K
$125-150K
$150-200K
$200K or more

Evidence supporting selection based on income

- LML selection on income -87,041
- LML random selection -89,038
- LML selection on ZIP code -230,275

(the maximal fit is -62,923, however using more than 32,000 parameters)

Relation to Little and Rubin's framework

- L&R distinguish b/w MCAR, MAR and non-ignorable patterns of missingness (NMAR)
- Their framework covers the case where some observations are incomplete
- In our case, the variable of interest is not observed at all in the target list, i.e., all observations are missing the same variable
- Combinations of MCAR, MAR and even NMAR (in the list) with our framework are possible (e.g., some age realizations could be observed in the target list of names and ages)

Summary

- Simple count method, i.e., the industry standard biased when observed X not a direct cause to selection
- Bayesian profiling: a technique to harness the information in observed ‘indicator’ variables
- Indicator variables depart from the reference because of selection based on a different, unobserved variable
- Various generalizations of the method in the paper

- In general, it helps to distinguish between causes and consequences ☺