

A Bayesian Approach to Data Disclosure: Optimal Intruder Behavior for Continuous Data

SE Fienberg, UE Makov and AP Sanil (1997)

Presented by
Lisa Denogean

Outline

1.0 Introduction - Brief Review of “Disclosure”

2.0 Basic Disclosure Model - Intruder’s Perspective

3.0 Bayesian Approach - Prior & Posterior Distributions

4.0 Case Study - Canadian Survey of Elite Attitudes

4.1 Data

4.2 Implementation of Bayesian Model

4.3 Computational Results

5.0 Further Work

1.0 Introduction - Brief Review of “Disclosure”

- Fellegi (1972) - *Disclosure requires both the recognition of an individual member of a population and gaining information about that individual*
 - ◇ Identification of a sample member from data release without prior knowledge of individual being a member of the sample
 - ◇ Identification of additional identifiable characteristics
- Dalenius (1997) - *If the release of certain statistical information make it possible to determine a particular value relating to a known individual more accurately than is possible without access to that data, then a disclosure has taken place*
 - ◇ Seek to limit disclosure

- Lambert (1993) - Bayesian approach to disclosure identification from intruder perspective
 - ◇ Utilizes a model for the intruder
 - ◇ Does not optimize intruder behavior
 - ◇ Does not explicitly allow for mechanism that might prevent intruder from making a match
- Fuller (1993) - Measurement error as part of a masking process intended to foil an intruder's attempts at identification

2.0 Basic Disclosure Model

- Model describing the data with unknown parameters with exchangeable priors
- Probabilistic mechanism which introduces bias into the responses
- Probabilistic mechanism which generates errors in the database
- All variables are continuous and unique
- *(A1)* The intruder possesses verified (exact) information, \mathbf{x} , on several individuals
- An agency releases (possibly corrupted or modified) data \mathbf{y}
- The intruder seeks to obtain additional information by identifying the released information of individuals with records in \mathbf{x}
- Intruder information \mathbf{x} is accurate, allowing for upper bound on posterior probability of correct identification

The data available to the intruder consists of exact data on an individual at the center of the investigation, $\mathbf{x}_0 = \{x_{01}, x_{02}, \dots, x_{0k}\}$, a k attribute vector of observations on that individual (*key variables*)

The agency records data on a population of N individuals, $\mathbf{y}^{(N)} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$, where $\mathbf{y}_i = \{y_{i0}, y_{i1}, \dots, y_{iq}\}$, $q > k$, where y_{i0} is a u -vector of identifying attributes not released to the public (e.g. SSN)

There is a nonzero probability that $\mathbf{y}^{(N)}$ does not contain any information on \mathbf{x}_0

The agency releases the censored records of n individuals, $\mathbf{z}^{(n)} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$, $\mathbf{z}_i = \{z_{i1}, \dots, z_{iq}\}$, arranged so that z_{ij} corresponds to x_{0j} for $j = 1, \dots, k < q$

Define indicator function $\mathcal{J} = 1, 2, \dots, n + 1$, such that $\mathcal{J} = j$ if \mathbf{x}_0 is associated with the individual whose released record is given by \mathbf{z}_j , and $\mathcal{J} = n + 1$ if \mathbf{x}_0 is associated with an individual whose record has not been released

The aim of the intruder is to find the value of $\mathcal{J} = r$, yielding disclosure of attributes $y_{r(k+1)}, \dots, y_{rq}$

Note that $\mathcal{J} = r \leq n$ does not imply that $z_{rj} = x_{0j}$ due to errors in the agency data

The problem faced by the intruder is to attempt to match record \mathbf{x}_0 with one of those released

Assumption A2

- Distribution of attributes among individuals denoted $f(\mathbf{z}^{(n)}|\mu)$, where parameters $\mu = \{\mu_1, \dots, \mu_q\}$
- Independence: $f(\mathbf{z}^{(n)}|\mu) = \prod_i f(\mathbf{z}_i|\mu)$
- $\mu \sim t(\mu)$, for some density t

Assumption A3

- $x_{0j} = \theta_{ij} + \xi_j, \mathcal{J} = i; i = 1, \dots, n + 1; j = 1, \dots, k$
- Bias removing parameter denoted θ_{ij}
($\theta_{(n+1)j}$ is used for all un-released records)
- $\xi_j \sim N(0, \sigma_j^2)$, with $\sigma^2 \sim \nu(\sigma_j^2)$
(ξ_j varies across attributes, independent of θ_{ij})

Assumption A4

- The θ_{ij} are exchangeable with respect to i
- $\theta_{ij} \sim g(\theta_{ij}|\varphi_j)$, where $\varphi_j \sim h(\varphi_j)$

3.0 Bayesian Approach

Intruder's Goal: Obtain the posterior distribution of \mathcal{J} , $P(\mathcal{J}|\mathbf{x}_0; \mathbf{z}^{(n)})$

Intruder's "Optimal Behavior": Decide that the confidential records associated with \mathbf{x}_0 are in \mathbf{z}_m , where m is the value of \mathcal{J} for which $P(\mathcal{J}|\mathbf{x}_0; \mathbf{z}^{(n)})$ is maximized

$$P(\mathcal{J}|\mathbf{x}_0; \mathbf{z}^{(n)}) \propto f(\mathbf{x}_0|\mathcal{J} = i; \mathbf{z}^{(n)})f(\mathcal{J} = i|\mathbf{z}^{(n)}), i = 1, \dots, n + 1 \quad (1)$$

$$f(\mathcal{J} = i|\mathbf{z}^{(n)}) = \begin{cases} 1/N & \text{for } i = 1, \dots, n \\ (N - n) & \text{for } i = n + 1 \end{cases} \quad (2)$$

$$\begin{aligned}
f(\mathbf{x}_0 | \mathcal{J} = i; \mathbf{z}^{(n)}) &= \prod_{j=1}^k \int_{\theta_{ij}} \int_{\sigma_j^2} f(x_{0j} | \theta_{ij}; z_{ij}; \sigma_j^2) \cdot f(\theta_{ij} | \mathcal{J} = i; \mathbf{z}^{(n)}) \\
&\quad \cdot f(\sigma_j^2 | \mathcal{J} = i; \mathbf{z}^{(n)}) d\theta_{ij} d\sigma_j^2 \quad (3)
\end{aligned}$$

$$\begin{aligned}
f(\mathbf{x}_0 | \mathcal{J} = n + 1; \mathbf{z}^{(n)}) &= \\
\prod_{j=1}^k \int_{\theta_{(n+1)j}} \int_{\sigma_j^2} \int_{z_{(n+1)j}} &f(x_{0j} | \theta_{(n+1)j}; z_{(n+1)j}; \sigma_j^2) \cdot f(\theta_{(n+1)j} | \mathcal{J} = n + 1; \mathbf{z}^{(n)}) \\
&\quad \cdot f(z_{(n+1)j} | \mathbf{z}^{(n)}) f(\sigma_j^2 | \mathcal{J} = n + 1; \mathbf{z}^{(n)}) d\theta_{ij} dz_{(n+1)j} d\sigma_j^2 \quad (4)
\end{aligned}$$

(Integrating out the unknown realization of $z_{(n+1)j}$)

where $f(z_{(n+1)j} | \mathbf{z}^{(n)}) = \int f(z_{(n+1)j} | \mu) f(\mu | \mathbf{z}^{(n)}) d\mu$
and $f(\mu | \mathbf{z}^{(n)}) \propto f(\mathbf{z}^{(n)} | \mu) t(\mu)$

4.0 Case Study

Variables Measured on 5 point scale from 1981 phone interviews

Civil-liberties

C1 – Free speech is just not worth it

C2 – We have gone too far in pushing equal rights in this country

C3 – It is better to live in an orderly society than to allow people so much freedom

C5 – Free speech ought to be allowed for all political groups

Attitudes towards Jews

A15 – Most Jews don't care what happens to people who are not Jews

A18 – Jews are more willing than others to use shady practices to get ahead

Canada–U.S. relationship

CUS1 – Ensure independent Canada

CUS5 – Canada should have free trade with the U.S.A.

CUS6 – Canada's way of life is influenced strongly by U.S.A.

CUS7 – Canada benefits from U.S. investments

In addition, we have data on two approximately continuous variables:

Personal information

Income – Total family income before taxes (with top-coding at 80,000 USD)

Age – Based on year of birth

4.1 Data

662 observations on variables:

$$\text{Civil} = C1 + C2 + C3 + (8 - C5) + N(0, \frac{1}{2})$$

$$\text{Attitude} = A15 + A18 + N(0, \frac{1}{2})$$

$$\text{Can/US} = (5 - CUS1) + CUS5 + (5 - CUS6) + CUS7 + N(0, \frac{1}{2})$$

$$\text{Age} = \text{Age} + N(0, 4)$$

$$\text{Income} = \begin{cases} \text{Income} + U[0, \$10,000] & \text{Income} < \$80,000 \\ \$60,000 + 25,000 * t(8)_{.38} & \text{Income} \geq \$80,000 \end{cases}$$

Agency releases all variables except ‘Attitudes’ (toward Jews)

‘Attitudes’ is unavailable to the intruder and at the center of the intruder’s investigation

Released data: $\mathbf{z} = \{z_{ij}, i = 1, \dots, 662; j = 1, \dots, 4\}$ ($n = N$)

Intruder’s data: $x_{0j} = z_{ij} * \theta_{ij} + \xi_j$, where $\theta_{ij} \sim N(1, \varphi_j), \xi \sim N(0, \sigma_j^2)$

	φ_j	σ_j^2
Civil	0.1732	25
Can/U.S.	0.1732	25
Age	0.1732	9
Income (in 10,000’s USD)	0.1732	4

Table 4.1. First 10 records of \mathbf{x} and \mathbf{z}

Age		Civil		Can/U.S.		Income (USD)	
\mathbf{x}	\mathbf{z}	\mathbf{x}	\mathbf{z}	\mathbf{x}	\mathbf{z}	\mathbf{x}	\mathbf{z}
44.607011	31.00364	24.803494	23.26688	0.2782396	7.230798	80351.260	86680.77
44.356572	58.36153	17.330712	17.76006	6.8772395	4.480846	95886.344	64127.42
35.260936	49.43488	10.930148	14.58399	12.3295743	7.632419	120247.969	88728.53
47.740238	40.87560	20.582654	16.54536	10.9438634	9.448964	106980.348	80348.58
32.257831	30.38650	21.430536	14.09269	16.5630154	10.828120	74050.109	76234.72
16.964057	21.51478	21.842566	14.62777	12.4165201	14.206017	105327.918	81986.36
43.319185	52.79831	10.552020	16.20542	2.3126535	5.881757	54703.225	73593.64
36.162886	42.55710	22.629968	21.26227	2.0760983	5.632542	39358.922	63209.81
31.119159	32.50106	15.738561	20.83966	8.4469488	14.843309	4466.606	42866.14
56.607847	82.03417	19.465088	18.70948	4.3898451	7.309711	102111.234	119271.81

4.2 Implementation of Bayesian Model

Uniform prior: $f(\mathcal{J} = j | \mathbf{z}^{(662)}) = \frac{1}{662}$

Previous assumptions $\Rightarrow x_{ij} | \theta_{ij}, z_{ij}, \sigma_j^2 \sim N(\theta_{ij} z_{ij}, \sigma_j^2)$
and $x_{ij} | \theta_{ij}, z_{ij}, \varphi_j \sim N(z_{ij}; \sigma_j^2 + z_{ij}^2 \varphi_j)$

Eqn (3) becomes (8):

$$f(\mathbf{x}_0 | \mathcal{J} = i; \mathbf{z}^{(n)}) = \prod_{j=1}^4 \int \int f(x_{0j} | \mathbf{z}^{(n)}; \sigma_j^2; \varphi_j; \mathcal{J} = i) \cdot f(\varphi_j) \cdot f(\sigma_j^2) d\varphi_j d\sigma_j^2$$

Identification Rule: Choose the value of i that maximizes (8)
(Generate observations from $f(\varphi_j)$ and $f(\sigma_j^2)$)

Priors on φ_j and σ_j^2 : $\nu(\sigma_j^2), h(\varphi_j) \sim \text{Gamma}(\alpha, \beta)$

Assume: $\sqrt{\sigma_j^2} = \frac{1}{6} \text{Range}\{\text{Variable}_j\}$,

and $\sqrt{\varphi_j} = \frac{1}{6} \text{Range}\{\theta_j\} = \frac{1}{6} \text{Range}\{[0.75, 1.25]\}$

Table 4.2. Values of the parameters of the Gamma priors distributions used for φ and σ^2

	Attribute	α	β
φ	Age	400	5,000
	Civil	400	5,000
	Can/U.S.	400	5,000
	Income (in 10,000's USD)	400	5,000
σ^2	Age	400	3.5
	Civil	400	34
	Can/U.S.	400	63
	Income (in 10,000's USD)	400	110

4.3 Computational Results

4 simulation scenarios:

- The released data contains no bias or noise ($\varphi_j = 0, \sigma_j^2 = 0$)
- The released data contains only noise ($\varphi_j = 0$)
- The released data contains only bias ($\sigma_j^2 = 0$)
- The released data contains both bias and noise (Table 4.2)

Each individual was taken in turn as the object of the intruder's efforts

The following tables give the number of times the correct record was ranked 1st through 10th and the average probability of a correct match

Table 4.3. Results for data without noise or bias

Rank	Number observed	Avg. prob. match (SD)	Top 10 cum. (SD)
1	319	0.017 (0.021)	0.113 (0.067)
2	144	0.009 (0.003)	0.078 (0.025)
3	106	0.007 (0.003)	0.066 (0.024)
4	41	0.006 (0.001)	0.06 (0.011)
5	34	0.006 (0.001)	0.058 (0.011)
6	8	0.005 (0.001)	0.051 (0.009)
7	3	0.005 (0.0001)	0.054 (0.001)
8	3	0.005 (0.001)	0.055 (0.008)
9	3	0.006 (0.0001)	0.056 (0.003)
10	1	0.007 –	0.074 –

Table 4.4. Results for data with only noise

Rank	Number observed	Avg. prob. match (SD)	Top 10 cum. (SD)
1	42	0.045 (0.057)	0.204 (0.14)
2	36	0.023 (0.021)	0.162 (0.087)
3	29	0.015 (0.012)	0.177 (0.092)
4	26	0.015 (0.01)	0.146 (0.089)
5	18	0.015 (0.007)	0.16 (0.082)
6	22	0.014 (0.008)	0.145 (0.094)
7	6	0.014 (0.01)	0.193 (0.177)
8	13	0.01 (0.003)	0.119 (0.043)
9	12	0.009 (0.004)	0.111 (0.06)
10	10	0.012 (0.007)	0.176 (0.156)
>10	448	0.005 (0.003)	0.109 (0.058)

Table 4.5. Results for data with only bias

Rank	Number observed	Avg. prob. match (SD)	Top 10 cum. (SD)
1	295	0.018 (0.026)	0.116 (0.071)
2	141	0.009 (0.004)	0.081 (0.029)
3	75	0.008 (0.003)	0.073 (0.022)
4	46	0.006 (0.001)	0.06 (0.012)
5	33	0.006 (0.001)	0.062 (0.012)
6	21	0.007 (0.002)	0.067 (0.02)
7	13	0.005 (0.001)	0.054 (0.008)
8	11	0.005 (0.001)	0.056 (0.01)
9	6	0.006 (0.002)	0.061 (0.019)
10	2	0.006 (0.002)	0.062 (0.016)
>10	19	0.005 (0.0001)	0.054 (0.005)

Table 4.6. Results for data with both bias and noise

Rank	Number observed	Avg. prob. match (SD)	Top 10 cum. (SD)
1	43	0.044 (0.068)	0.193 (0.149)
2	34	0.026 (0.022)	0.179 (0.088)
3	34	0.02 (0.014)	0.168 (0.097)
4	24	0.014 (0.008)	0.133 (0.073)
5	16	0.02 (0.012)	0.2 (0.115)
6	15	0.013 (0.006)	0.139 (0.073)
7	16	0.014 (0.007)	0.167 (0.119)
8	13	0.014 (0.006)	0.184 (0.116)
9	12	0.011 (0.005)	0.139 (0.091)
10	4	0.009 (0.003)	0.104 (0.042)
>10	451	0.005 (0.003)	0.108 (0.057)

Summary of Results

- **No bias or noise:** 319/662 cases the intruder's posterior probability of a match was highest for the correct record, the average value of the posterior probability associated with the highest ranked record is modestly small (0.02 to 0.10), no cases of low rank
- **Noise, no bias:** 42/662 cases correct rank, 448/662 cases the correct record was not in the top ten of records indicated from posterior probability
- **Bias, no noise:** 295/662 cases correct rank, 19/662 cases not in top ten
- **Bias & noise:** 43/662 cases correct rank, 451/662 cases not in top ten

⇒ Even with moderate bias and noise, the intruder has the correct match in the top five around 40% of the time. But, the intruder did not achieve true matches with high probability, and this should degrade as n grows.

5.0 Further Work

- Specify a loss function for the intruder, e.g. utilize a threshold for the posterior probability of a match
- Intruder could match more than one record at a time against released data, increasing probability of success
- Agency may release only a portion of sampled data, decreasing intruder's success
- Intruder could have only approximate information on individuals
- Data could include categorical variables
- Explore agency response for optimal intruder behavior, e.g. use of matrix masking and ability of intruder to undo agency's masking