A new selection model via clustering missingness indicators

Byungtae Seo

December 18, 2011

(Joint work with Drs. Hyekyung Jung and Joseph L. Schafer)

Missing data

- For each subject,

$$\begin{aligned} Y &= (y_1, y_2, \ldots, y_p) : \text{complete responses} \\ &= (Y_{obs}, Y_{mis}) \\ R &= (r_1, r_2, \ldots, r_p) : \text{missingness indicators} \\ &\quad (1{=}\text{observed}, 0{=}\text{missing}) \end{aligned}$$

- For statistical analysis, the joint distribution of Y and R should be modeled.
- If we factor $P(Y, R)$ into $P(R|Y)$ and $P(Y)$ (i.e. $P(Y, R) = P(R|Y)P(Y)$), we need the following statistical models:
  - $P(Y; \theta)$: Complete data model
  - $P(R|Y; \eta)$: missingness mechanism

Ignorable missing

- Missing Completely At Random (MCAR):

$$P(R|Y) = P(R)$$

$\rightarrow$ Missingness does not depend on any data

- Missing At Random (MAR):

$$P(R|Y) = P(R|Y_{obs})$$

$\rightarrow$ Missingness depends on the observed data but not unobserved data

- Probability distribution of the observed data under MAR

$$\begin{aligned} P(R, Y_{obs}) &= \int P(R, Y) dY_{mis} \\ &= \int P(R|Y;\eta) P(Y;\theta) dY_{mis} \\ &= \int P(R|Y_{obs};\eta) P(Y_{obs}, Y_{mis};\theta) dY_{mis} \\ &= P(R|Y_{obs};\eta) P(Y_{obs};\theta) \end{aligned}$$

- Since $P(R, Y_{obs})$ is factored into $P(R|Y_{obs};\eta)$ and $P(Y_{obs};\theta)$, we do not have to consider $P(R|Y_{obs},\eta)$ when θ is the parameter of interest. That is, the missing data mechanism is ignorable.

Nonignorable missing

- Not Missing At Random (NMAR):

$$P(R|Y) \neq P(R|Y_{obs})$$

$\rightarrow$ Missingness depends on both observed and unobserved data

- Under NMAR, missing mechanism is not ignorable.
- Require a certain model for $P(Y, R)$ in addition to the complete data model.

Two common types of NMAR models

- Pattern-Mixture Model (Little, 1993):

$$P(Y, R) = P(Y|R)P(R)$$

$$\rightarrow P(Y_{obs}, R) = \int P(Y_{obs}, Y_{mis}|R; \alpha)P(R; \beta)dY_{mis}$$

  - Can give the inference to the subpopulation of cases ($R = 0$ and $R = 1$).
  - The computation is easier than the selection model.
  - The inference for the parameter of interest, θ, is difficult.

- Selection Model (Diggle & Kenward, 1994):

$$P(Y, R) = P(R|Y)P(Y)$$

$$\rightarrow P(Y_{obs}, R) = \int P(R|Y_{obs}, Y_{mis}; \eta)P(Y_{obs}, Y_{mis}; \theta)dY_{mis}$$

  - A natural way of factoring $P(Y, R)$.
  - Require a parametric model for missingness mechanism $P(R|Y; \eta)$.
  - Based on likelihood based method or Bayesian approach, we can have direct interpretation and inference for θ.
  - In general, the likelihood is complicated.

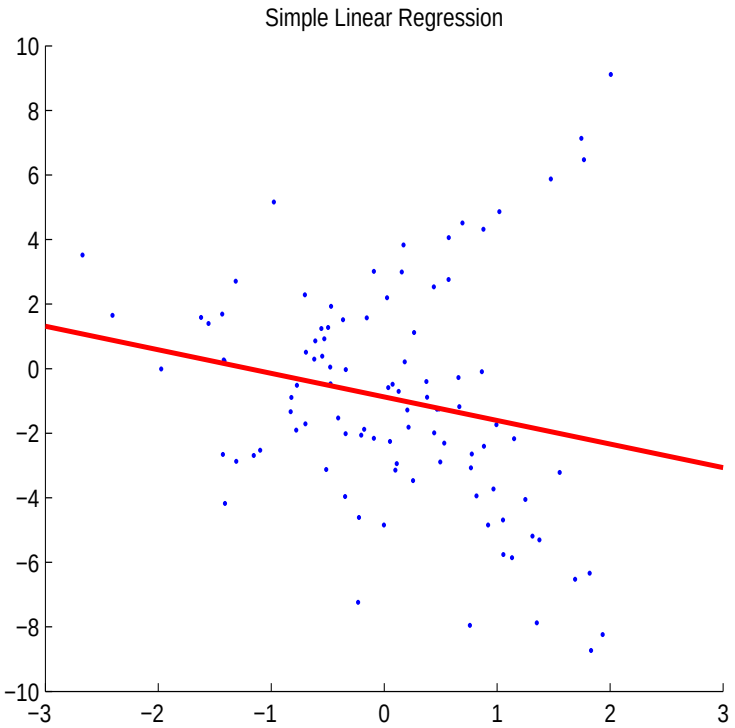
- Some observations on the traditional selection model
  - In selection model, missingness mechanism $P(R|Y)$ is not flexible enough to explain the true mechanism.
  - The estimation in the selection model is often instable.
  - In many cases, a large number of missing-data patterns may be reduced to just a few prototypes.
- From these observations, we propose the following latent class regression model for missingness mechanism:

$$P(R|Y) = \sum_{l=1}^{C} P(L = l|Y)P(R|Y, L = l)$$

Example: Latent class regression model

- Simple linear regression: $y = \beta_0 + \beta_1 x + \epsilon$ where $\epsilon \sim N(0, \sigma^2)$.

$$\Longrightarrow f(y; \beta_0, \beta_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \beta_0 - \beta_1 x)^2}{2\sigma^2}\right)$$

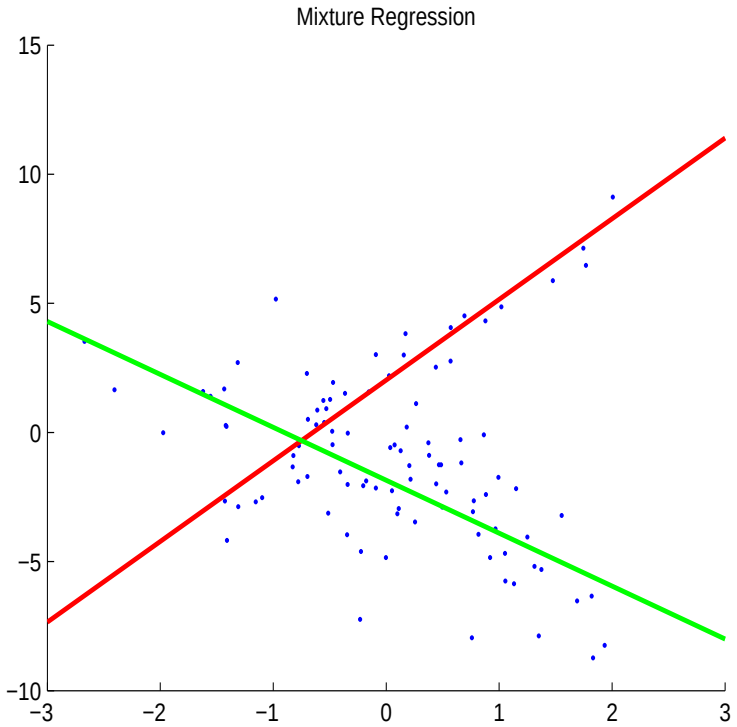


- Latent class regression

$$\frac{p}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(y-\beta_{01}-\beta_{11}x)^2}{2\sigma_1^2}\right)+\frac{1-p}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(y-\beta_{02}-\beta_{12}x)^2}{2\sigma_2^2}\right)$$

- Latent class regression

$$\frac{p}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(y-\beta_{01}-\beta_{11}x)^2}{2\sigma_1^2}\right)+\frac{1-p}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(y-\beta_{02}-\beta_{12}x)^2}{2\sigma_2^2}\right)$$



Latent Class Selection Model

- Latent Class Selection Model (LCSM)

$$P(Y, R) = P(R|Y)P(Y)$$

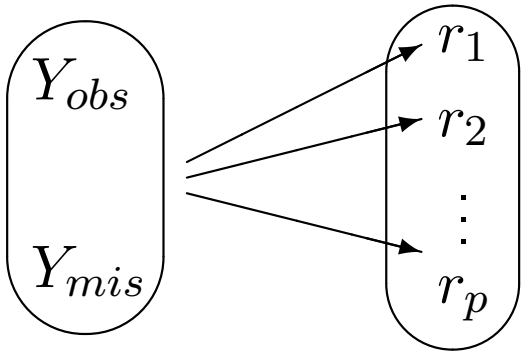
with

$$P(R|Y) = \sum_{l=1}^{C} P(R|L = l)P(L = l|Y)$$

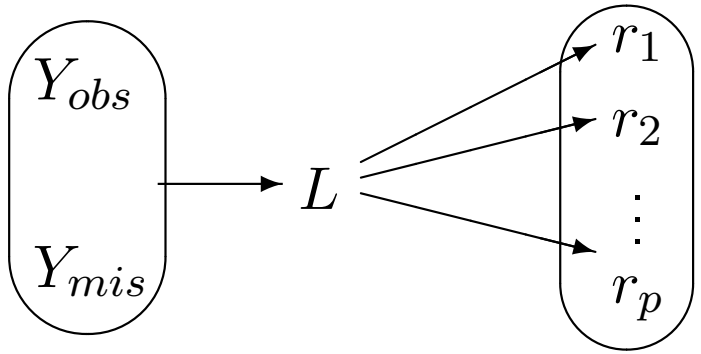
- LCSM assumes $P(R|Y, L) = P(R|L)$. That is, R depends on Y only through L.
- L describes the classes of respondents, summarizing missingness patterns.

- Diagrams

<Selection Model>



<LCSM>



- Potential advantage of LCSM
  - Can capture the information in missingness indicators in parsimonious way.
  - Can avoid instability and extreme sensitivity of conventional selection models

Example

- Complete data model: $Y \sim N_3(\mu, \Sigma)$.
- Missingness mechanism:
  - Missingness mechanism in each given class

$$P(r_1, r_2, r_3 | L = l) = \prod_{j=1}^{3} \rho_{j|l}^{r_j} (1 - \rho_{j|l})^{1-r_j}$$

    where $\rho_{j|l}$ is the probability that y_j is observed in l-the class
  - Class probability given Y

$$\pi_l(Y) = P(L = l|Y) = \frac{\exp(\beta_{0l} + \beta_{1l} y_1 + \beta_{2l} y_2 + \beta_{3l} y_3)}{1 + \sum_{l=1}^{C-1} \exp(\beta_{0l} + \beta_{1l} y_1 + \beta_{2l} y_2 + \beta_{3l} y_3)}$$

    where $\beta_{0C} = \beta_{1C} = \beta_{2C} = \beta_{3C} = 0$
- Probability distribution of the observed data under LCSM

$$P(Y_{obs}, R) = \int \left[\sum \pi_l(Y) \prod_{j=1}^{3} \rho_{j|l}^{r_j} (1 - \rho_{j|l}^{r_j})^{1-r_j} \right] N_3(Y; \mu, \Sigma) dY_{mis}$$

Simulation study 1: nonignorable missing data

- Estimate μ's under nonignorable missing-data mechanism
- Draw $n = 2000$ observations for 100 replicates following the simulation scheme:
  - Generate $(Y_1, Y_2, Y_3) \sim MVN(\boldsymbol{\mu}, \Sigma)$ with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & \gamma & \gamma \\ \gamma & 1 & \gamma \\ \gamma & \gamma & 1 \end{pmatrix}$$

  - Generate L_i such that

$$L_i = \begin{cases} 1 & \text{with probability } \frac{\exp(\beta_0+\beta_1 y_1+\beta_2 y_2+\beta_3 y_3)}{1+\exp(\beta_0+\beta_1 y_1+\beta_2 y_2+\beta_3 y_3)} \\ 2 & \text{with probability } \frac{1}{1+\exp(\beta_0+\beta_1 y_1+\beta_2 y_2+\beta_3 y_3)} \end{cases}$$

    with $(\beta_0, \beta_1, \beta_2, \beta_3)=(0, 1, 1, 1)$
  - Generate missingness indicators according to response probabilities within each class:

$$\begin{aligned} (\rho_{1|1}, \rho_{2|1}, \rho_{3|1}) &= (0.1, 0.1, 0.999) \\ (\rho_{1|2}, \rho_{2|2}, \rho_{3|2}) &= (0.9, 0.9, 0.999) \end{aligned}$$

Table: Simulation results for a nonignorable missing data, bias=bias$\times 100$, se=standard error$\times 100$

		μ_1		μ_2		μ_3	
γ	Method	bias	se	bias	se	bias	se
0.2	CC	-41.07	3.35	-40.77	3.19	-40.28	3.15
	MAR	-29.73	3.12	-29.74	3.32	0.37	2.39
	LCSM2	1.30	6.32	0.47	6.63	0.35	2.39
	LCSM3	-1.28	10.88	-1.32	8.81	0.36	2.39
0.5	CC	-51.84	3.56	-51.95	2.91	-51.60	2.80
	MAR	-24.62	3.29	-24.80	3.07	-0.02	2.20
	LCSM2	0.78	5.83	0.45	5.70	-0.04	2.21
	LCSM3	-0.23	7.61	-1.55	8.92	-0.03	2.20
0.8	CC	-60.96	2.51	-60.87	2.51	-60.76	2.62
	MAR	-11.45	2.75	-11.57	2.81	0.53	2.38
	LCSM2	1.47	3.55	0.63	4.52	0.52	2.38
	LCSM3	1.31	5.31	0.24	5.95	0.52	2.38

Simulation study 2: ignorable missing data

- Estimate μ's under ignorable missing-data mechanism
- Draw $n = 2000$ observations for 100 replicates following the simulation scheme:
  - Generate $(Y_1, Y_2, Y_3) \sim MVN(\boldsymbol{\mu}, \Sigma)$ with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & \gamma & \gamma \\ \gamma & 1 & \gamma \\ \gamma & \gamma & 1 \end{pmatrix}$$

  - Generate missingness indicators based on

$$P(R_1 = 0) = \frac{\exp(\beta_0 + \beta_1 y_3)}{1 + \exp(\beta_0 + \beta_1 y_3)}$$

$$P(R_2 = 0) = \frac{\exp(\beta_0 + \beta_2 y_3)}{1 + \exp(\beta_0 + \beta_2 y_3)}$$

with $\beta_0 = \beta_1 = \beta_2 = -1$

Table: Simulation results for a nonignorable missing data, bias=bias$\times 100$, se=standard error$\times 100$

		μ_1		μ_2		μ_3	
γ	Method	bias	se	bias	se	bias	se
0.2	CC	8.65	2.89	7.99	3.01	44.32	2.74
	MAR	-0.51	2.57	-0.49	2.58	-0.44	2.22
	LCSM2	-0.52	3.11	-0.63	3.17	-0.44	2.22
0.5	CC	22.16	3.30	21.95	3.20	43.93	2.89
	MAR	-0.05	2.49	-0.16	2.69	-0.34	2.16
	LCSM2	-0.21	2.94	-0.04	3.14	-0.34	2.16
0.8	CC	35.27	3.14	35.50	3.54	44.17	3.23
	MAR	-0.18	2.57	0.03	2.71	-0.20	2.44
	LCSM2	-0.23	2.81	0.14	2.81	-0.20	2.44

Summary

- LCSM can be considered as an adaptive version of the classical selection model.
- LCSM can capture the underlying group structure of the missingness mechanism.
- The number of classes can be chosen using BIC or PPCD.
- When the chosen number of classes is too large compared to the true value, MCMC procedure is unstable. A more fast and stable algorithm for the estimation should be further studied.
- More comprehensive numerical studies are required with various models.