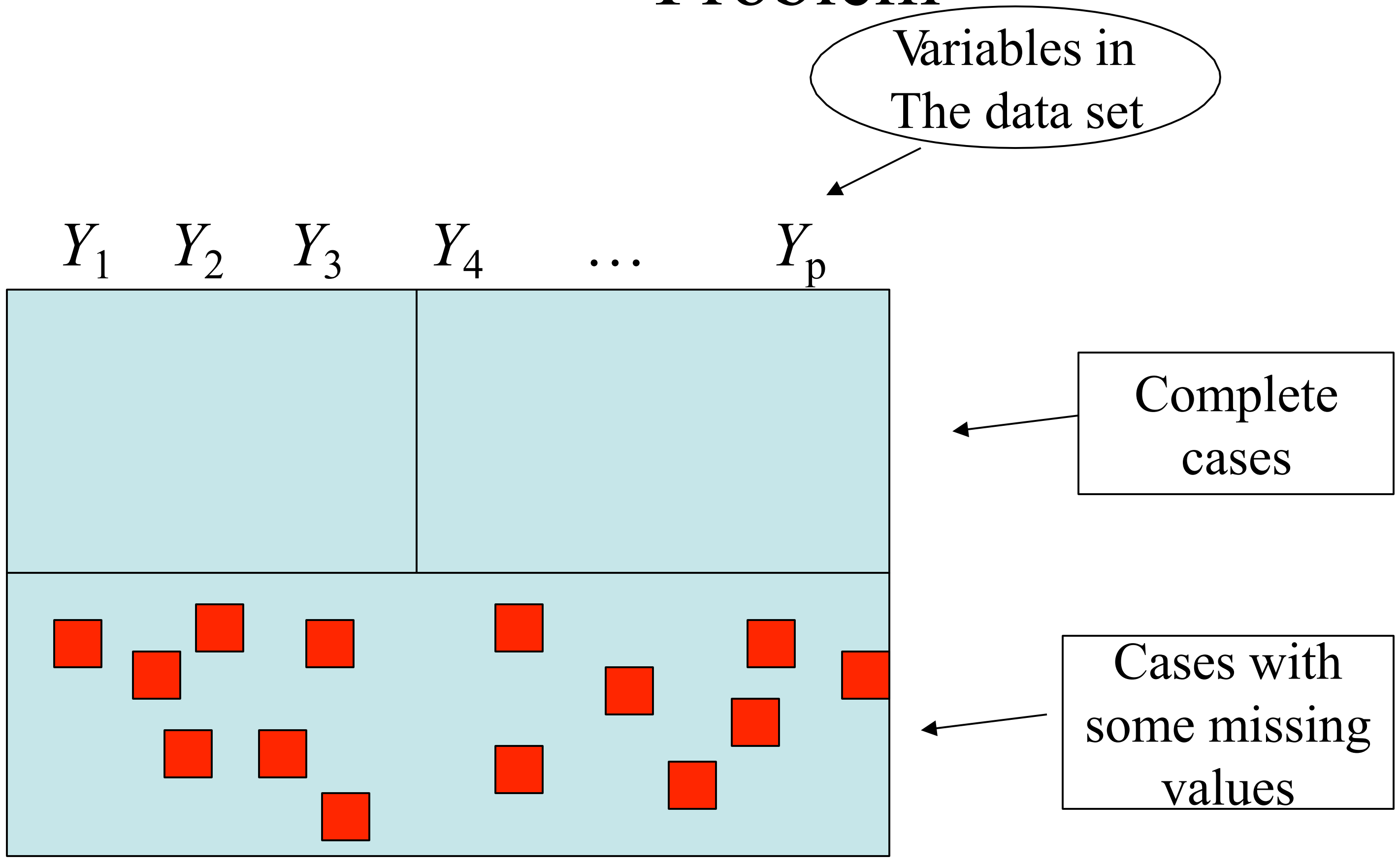
Statistical Analysis with Missing Data

Module 9
Multiple Imputation using Sequential Regression/Chained Equations

UNIVERSITY OF MICHIGAN

Problem



$D_{(0)}$ = Observed data:

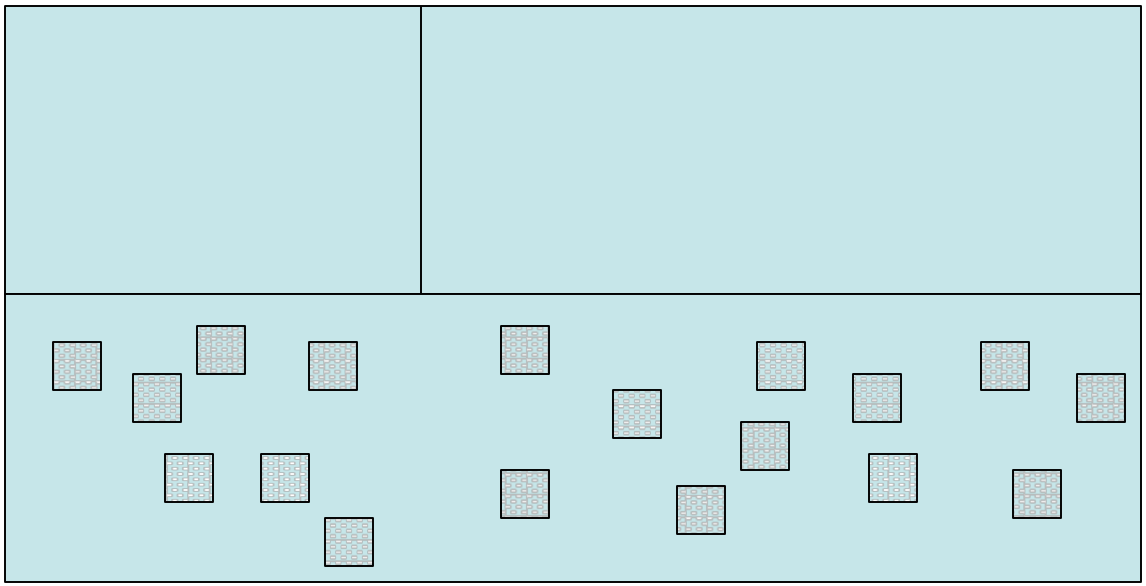
$D_{(1)}$ = Missing data:

Y: Discrete, continuous or semi-continuous as well as multivariate

Setting

- Multiple users analyzing different subsets of variables
- Multiple analytical techniques
- Different skill levels dealing with incomplete data
- Analysis to be performed with complete data is known
- Software to perform complete data analysis is available
- Assume missing at random.
  - That is conditional on the observed characteristics the residual differences between those with missing and those with no missing values are random

Imputation



Important issues:

Imputations are not real values

Uncertainties associated with imputes

$$\text{"}\textit{Ideal}\text{"}\ \textit{imputations}:$$

$$\textit{Draws from}\ \ \Pr(D_{(1)} \mid D_{(0)})$$

Practical Issues

- Hot deck imputation is limited
  - Variables have to be completely observed
  - Continuous variables have to be categorized
- Explicit Model is difficult
  - Large number of variables of different types
  - Restrictions
    - Question is valid only for certain subjects
    - Skip pattern
  - Bounds
    - Variables are bounded. *Example: Years smoked cannot exceed Age for current smokers and (Age-Years since Quit smoking) for former smokers. It can become more complex, if a question about teen age smoking was asked and age when started smoking was also asked*
    - Bracketed responses

Sequential Regression/Chained Equation/Flexible Conditional Specification Approach

Variables With Missing Values:

$$Y_1, Y_2, \cdots, Y_p$$

Variables With No Missing Values: U

Each step involves draws from the predictive distribution

Iteration 1:

$$Y_1 \mid U$$

$$Y_2 \mid Y_1^{(1)}, U$$

$$\vdots$$

$$Y_j \mid U, Y_1^{(1)}, \cdots, Y_{j-1}^{(1)}$$

$$\vdots$$

$$Y_p \mid U, Y_1^{(1)}, Y_2^{(1)}, \cdots, Y_{p-1}^{(1)}$$

Iteration t=2,3,…:

$$Y_1 \mid U, Y_2^{(t-1)}, \cdots, Y_p^{(t-1)}$$

$$Y_2 \mid U, Y_1^{(t)}, Y_3^{(t-1)}, \cdots, Y_p^{(t-1)}$$

$$\vdots$$

$$Y_j \mid U, Y_1^{(t)}, \cdots, Y_{j-1}^{(t)}, Y_{j+1}^{(t-1)}, \cdots, Y_p^{(t-1)}$$

$$\vdots$$

$$Y_p \mid U, Y_1^{(t)}, \cdots, Y_{p-1}^{(t)}$$

Example: bivariate continuous MAR data

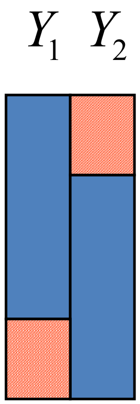
Continuous data with missing data on both variables

MAR mechanism

Recall that Gibbs' for bivariate normal model makes the restrictive assumption that regressions are linear:

$$\text{missing } y_{i2} : (y_{i2}^{(t+1)} \mid y_{i1}, \theta^{(t)}) \sim_{ind} N(\beta_{20\cdot 1}^{(t)} + \beta_{21\cdot 1}^{(t)} y_{i1}, \sigma_{22\cdot 1}^{(t)})$$

$$\text{missing } y_{i1} : (y_{i1}^{(t+1)} \mid y_{i2}, \theta^{(t)}) \sim_{ind} N(\beta_{10\cdot 2}^{(t)} + \beta_{12\cdot 2}^{(t)} y_{i2}, \sigma_{11\cdot 2}^{(t)})$$



With chained equations approach, we might replace linearity by modeling the two regressions with splines:

$$\text{missing } y_{i2} : (y_{i2}^{(t+1)} \mid y_{i1}, \theta^{(t)}) \sim_{ind} N(\text{spline}(y_{i1} \mid \theta^{(t)}), \sigma_{22\cdot 1}^{(t)})$$

$$\text{missing } y_{i1} : (y_{i1}^{(t+1)} \mid y_{i2}, \theta^{(t)}) \sim_{ind} N(\text{spline}(y_{i2} \mid \theta^{(t)}), \sigma_{11\cdot 2}^{(t)})$$

(Penalized splines can be modeled using random effects models. There is no joint distribution of Y_1 and Y_2 corresponding to these conditionals, but the added flexibility is still very useful in practice.

Flexible Features

- Ability to specify individual regression model
- Types of variables
  - Continuous (Normal)
  - Categorical (Logistic or generalized logistic)
  - Count (Poisson)
  - Mixed or semi-continuous (Logistic/Normal)
  - Ordinal (ordered probit)
- Parametric or semi-parametric regression models
- Restrictions
  - Regression model is fitted only to the relevant subset
- Bounds
  - Draws from a truncated distribution from the corresponding regression model
- Models each conditional distribution. There is no guarantee that a joint distribution exists with these conditional distributions
- How many iterations?
  - Empirical studies show that nothing much changes after 5 or 6 iterations

Software

- Sequential regression imputations
  - R and Stata (MICE, ICE, MI)
  - Standalone (SRCWARE)
  - SAS (IveWare), PROC MI
- MI-Analysis
  - PROC MIANALYZE
  - IveWare (can handle complex sample survey)
  - SRCWARE
  - MICOMBINE/MITOOLS (STATA)
  - SUDAAN

IveWare

- SAS interface
  - A collection of SAS, C and Fortran routines
  - Handles linear (Continuous), logistic (Binary), multinomial logistic (categorical), Poisson (Count) and two-stage linear/logistic (Mixed or semi-continuous)
    - Currently at work: Ordered probit and semiparametric regression models
    - Ability to recode while imputing
  - Stepwise selection possible at each step to save computation time (use with caution and only if it is absolutely necessary)
  - Add interaction terms
  - Specify bounds
  - Specify logical restrictions and skip patterns

IveWare

- Uses Normal approximation for the posterior distribution of the parameters
- Sampling Importance Resampling to handle non-normal posterior
- Non-informative prior
- Single chain or multiple chain (starting with different seeds)
- Iterations and Multiples control the length of the chain
- Version 0.3 soon to be released will work as stand alone and also with SAS, R, STATA and SPSS and any possible combination

IveWare

- Issues
  - Convergence
  - Several completed data statistics seem to converge to the same value regardless of seeds
  - Zhu and Raghunathan (JASA 2015) establish conditions for convergence
  - Good fitting models are needed to get results with desirable repeated sampling properties

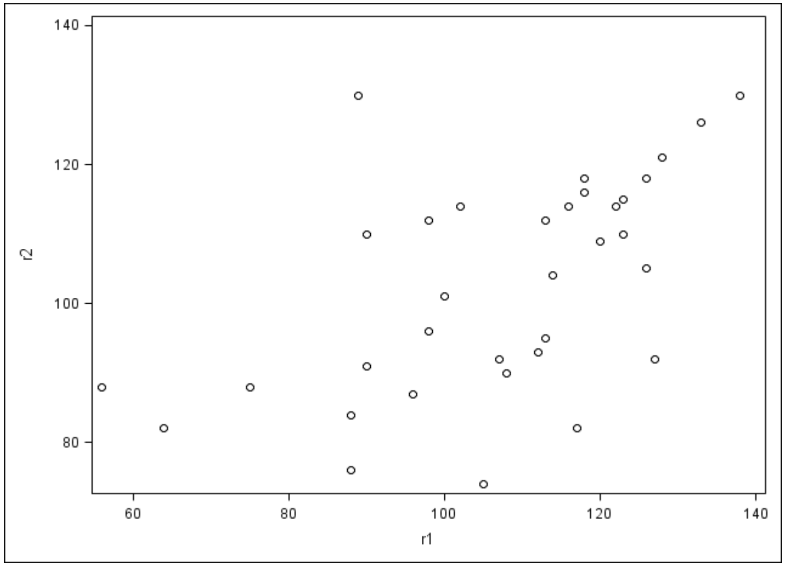
St. Louis Risk Study

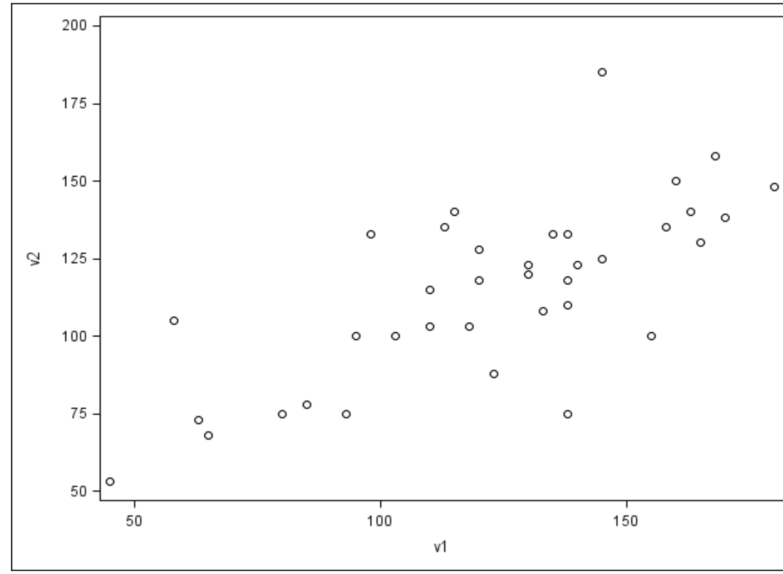
(Little and Rubin, 2002)

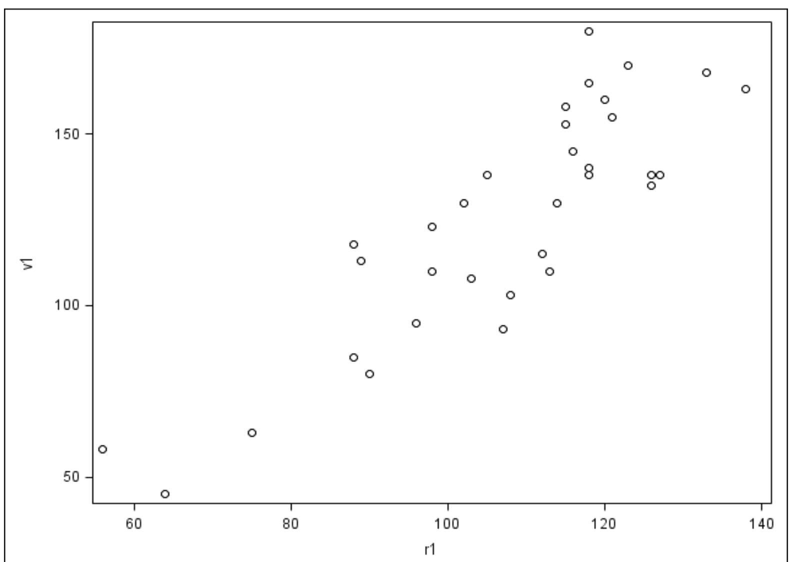
- A study was conducted to evaluate the effects parental psychological disorders on various aspects of the development of the children. Data from 69 families with two children were collected. Families were classified into risk group of the parent (G) with
  - G=1 normal or control group
  - G=2 Moderate risk group with one parent having some psychiatric illness
  - G=3 High risk group with one or more parent having schizophrenia or affective mental disorder

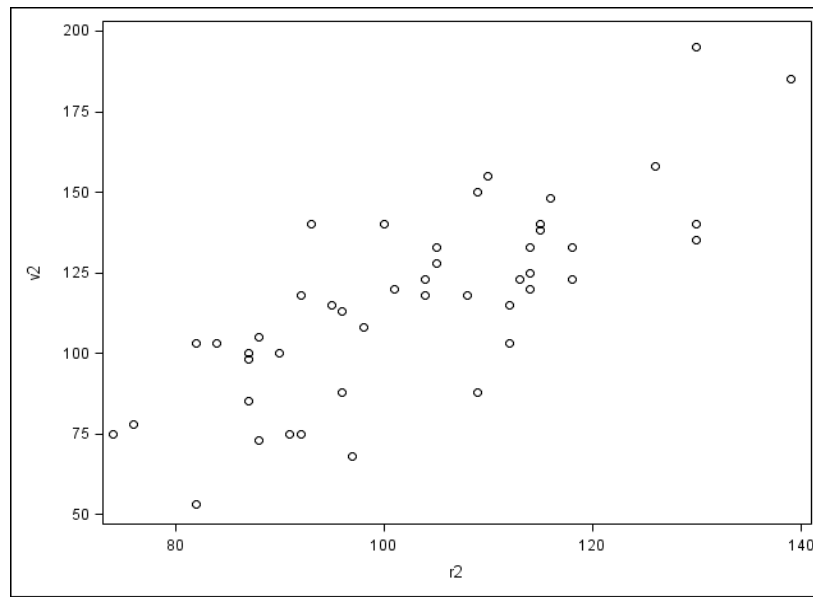
St. Louis Risk Study

- Variables measured on Child 1
  - D1= Number of symptoms (1=Low, 2=High)
  - V1= Standardized verbal comprehension score
  - R1=Standardized Reading score
- Variables Measured on Child 2
  - D2, V2, R2
- G is always observed and other variables are missing with variety of different combinations


r2
r1


v2
v1


v1
r1


v2
r2

St. Louis Risk Study

- Sequential regression approach used to impute the missing values R1, R2, V1, V2 using normal linear regression model and D1, D2 using the logistic regression model
- Analysis
  - Regress R on G and D , treating the Family ID as “cluster” or “Repeated” factor
  - Regress V on G and D, treating the Family ID as “cluster” or “Repeated” factor
  - Regress D on G, treating Family ID as “cluster” or “Repeated” factor

Results

Multiple Imputation Analysis

Parameter	Reading	Verbal	Symptoms
Intercept	114.23 (5.49)	152.67 (15.50)	-0.32 (0.41)
Group 2 vs 1	-9.85 (3.93)	-25.14 (13.56)	1.05 (0.74)
Group 3 vs 1	-9.69 (5.09)	-19.41 (11.03)	0.47 (0.51)
Symptoms	-1.02 (3.35)	-10.86 (10.97)	

Applications of MI

- Survey of Consumer Finances, 1992
  - 5 multiply imputed data sets
- National Health and Nutritional Examination Survey
  - 5 multiply imputed data sets for a selected set of variables in NHANES-III. Uses general location model.
- National Health Interview Survey 1997-Present
  - Multiple imputation of missing family income and personal earnings.
- Numerous applications in a variety of fields. Becoming a very common approach.

Software for Multiple Imputation Analysis

For Creating Imputations

- SAS
  - PROC MI
  - IVEware
- Standalone
  - SRCware
- STATA
  - MI IMPUTE
- R
  - MICE
- SOLAS
- SPSS (Version 22)

For Analysis of multiply imputed data

- SAS
  - PROC MIANALYZE
  - IVEware
- Standalone
  - SRCware
- STATA
  - MI ESTIMATE
- SUDAAN
- R
- SPSS (Version 22)

Conclusion

- Sequential Regression/Chained Equation is a flexible approach for handling missing data with varying type of variables and complex structure
- Standard regression diagnostics can be used to fine tune the model to fit the observed data well
- Models can be parametric, semi-parametric or non-parametric
- Many software available to implement the method
- It is easy to program using a macro environment

Example: Multivariate Normal Gibbs

- P-Step: Bayes for complete data – inverse Wishart for $\Sigma^{(t)}$, normal for $\mu^{(t)} \mid \Sigma^{(t)}$, using current imputed data
- I-Step: given current parameter draws $\theta^{(t)} = (\mu^{(t)}, \Sigma^{(t)})$ enter loop over cases; for case *i*:

  (a) Sweep for regression coefficients, res cov matrix of missing variables on observed variables, as functions of $\theta^{(t)}$;

  (b) fill in missing values y_{ij} with draws $\hat{y}_{ij}^{(t)}$ computed using regression equation with coefficients from (a);

  (c) add case *i* to vector of running means, sum of squares and cross products (sscp) matrix;

  ~~(d) add residual covariance matrix of missing variables given observed variables in case *i* to sscp matrix;~~

- recompute cd sufficient statistics from mean, sscp matrix

Multivariate Normal Gibbs

• Monitor Gibbs' chain for convergence (recommended: multiple chains to check that they merge)

• Draws after burn-in period can be used to simulate posterior distribution, or for multiple imputation

•Draws of functions of parameters are functions evaluated at draws: in particular we can simulate posterior distributions for regression parameters

Example: Discriminant Analysis and Logistic Regression, no missing data

Consider random sample on $(Y, X): \left((y_i, x_i), i = 1, ..., n\right)$

y_i is binary, taking values 0 and 1

$x_i = (x_{i1},...x_{iK})$ is $(1 \times K)$ vector of K continuous variables

Logistic regression model for y_i given x_i:

$$(y_i \mid x_i, \beta_0, \beta_x) \sim_{\text{ind}} \text{BERNOULLI}(\pi_i),\ \text{logit}(\pi_i) = \beta_0 + \beta_x^T x_i$$

Normal discriminant analysis model for x_i given y_i:

$$(x_i \mid y_i = j, \mu_0, \mu_1, \Omega) \sim_{\text{ind}} N_K(\mu_j, \Omega)\ (j = 0 \text{ or } 1)$$

These two models are related, as follows:

Discriminant Analysis and Logistic Regression

Combine normal discriminant analysis model for X given Y:

$(x_i \mid y_i = j, \mu_0, \mu_1, \Omega) \sim_{\text{ind}} N_K(\mu_j, \Omega)$ ($j = 0$ or 1)

with Bernoulli model for Y :

$(y_i \mid \pi) \sim_{\text{ind}}$ BERNOULLI(π)

The resulting conditional distribution of y_i given x_i is also logistic regression:

$$\frac{\Pr(y_i = 1 \mid x_i, \theta)}{\Pr(y_i = 0 \mid x_i, \theta)} = \frac{\Pr(y_i = 1 \mid \theta)}{\Pr(y_i = 0 \mid \theta)} \times \frac{\Pr(x_i \mid y_i = 1, \theta)}{\Pr(x_i \mid y_i = 0, \theta)}, \text{ where } \theta = (\pi, \mu_0, \mu_1, \Omega)$$

Taking logs:

$$\begin{aligned}
&\operatorname{logit} \Pr(y_i = 1 \mid x_i, \theta) \\
&\quad = \log(\pi / (1-\pi)) - 0.5(x_i - \mu_1)^T \Omega^{-1} (x_i - \mu_1) + 0.5(x_i - \mu_0)^T \Omega^{-1} (x_i - \mu_0) \\
&\quad = \log(\pi / (1-\pi)) + 0.5(\mu_0{}^T \Omega^{-1} \mu_0 - \mu_1{}^T \Omega^{-1} \mu_1) + (\mu_1 - \mu_0)^T \Omega^{-1} x_i \\
&\quad = \beta_0 + \beta_x x_i, \text{ where}
\end{aligned}$$

$$\beta_0 = \log(\pi / (1-\pi)) + 0.5(\mu_0{}^T \Omega^{-1} \mu_0 - \mu_1{}^T \Omega^{-1} \mu_1), \beta_x = (\mu_1 - \mu_0)^T \Omega^{-1}$$

ML estimation

For logistic regression: scoring or Newton-Raphson

Asymptotic SE's from information matrix, or bootstrap

For normal discriminant analysis:

(special case of multivariate regression)

$\hat{\mu}_j =$ sample mean in group j; $\hat{\Omega} =$ pooled cov matrix;

$\hat{\pi} =$ sample proportion of ones

Small-sample (Bayes or frequentist) inference based on

multivariate t distribution

An alternative to logistic regression

- An alternative to ML (Bayes) for parameters β of the logistic regression model is ML (Bayes) for discriminant analysis parameters $(\pi, \mu_0, \mu_1, \Omega)$, and hence ML (Bayes) for β expressed as functions of these parameters, e.g.

$$\hat{\beta}_0 = \log(\hat{\pi} / (1-\hat{\pi})) + 0.5(\hat{\mu}_0{}^T \hat{\Omega}^{-1} \hat{\mu}_0 - \hat{\mu}_1{}^T \hat{\Omega}^{-1} \hat{\mu}_1), \hat{\beta}_x = (\hat{\mu}_1 - \hat{\mu}_0)^T \hat{\Omega}^{-1}$$

- Assumes normality for covariates, but
- simpler computationally than logistic regression – for example, ML is non-iterative, and
- more statistically efficient, small sample inference is easy
- This discriminant analysis approach is now out of fashion – people don't like normality assumptions, and computational efficiency less of an issue than in the past

Logistic regression with missing covariates

- With missing data in the X's, we need to assume a distribution for the X's that are missing: three options:
- (a) Bayes or ML for discriminant analysis model. E-step for EM or I-step for Gibbs' is easy because X given Y is normal
- (b) Multiply impute missing X's assuming discriminant analysis model. Analyze filled-in data using discriminant analysis model, and MI combining rules. (Similar to a)
- (c) Multiply impute missing X's assuming discriminant analysis model. Analyze filled-in data using logistic regression, and MI combining rules.
- Note that (c) is potentially more robust, because normality of X's is only invoked for the imputed values

Logistic regression with missing covariates

- What about a chained equation MI?
- If the missing values are for variables are continuous and assumed normal, chained equations approach assumes the discriminant analysis model, so is the same as option (c)
- However, if some of the missing values are for variables that are not normal, chained equation MI does not have to assume normality for imputing these variables, and hence is potentially more robust

General Location Model

- The general location model generalizes the discriminant analysis model to arbitrary mixtures of continuous and categorical variables
  - Y set of categorical variables, X set of continuous variables
  - General location model factorizes joint distribution as

$$[Y, X]=[Y]\text{x}[X|Y]$$

  - Discriminant analysis is special case where Y is a single binary variable

General Location Model

- The general location model generalizes the discriminant analysis model to arbitrary mixtures of continuous and categorical variables
  - Also allows restrictions on the model parameters: loglinear models for the categorical variables, MANOVA models for continuous variables; so it's a pretty interesting and flexible model
  - Multiple imputation under this joint model is relatively simple: for a general pattern, we can use the Gibbs' sampler: algorithm involves elements of Gibbs' for incomplete normal models and Gibbs for loglinear models with incomplete categorical data
  - Lay out the main points here – see book for details.

General Location Model

- Model for mixtures of continuous and categorical variables

$Y =$ set of V categorical variables, variable j has I_j levels,

Total number of cells $C = \prod_{j=1}^{V} I_j$

X = set of K continuous variables

For unit i:

$x_i = (1 \times K)$ vector of continuous variables

$w_i = (1 \times C)$ vector with $w_i = E_c$ if unit falls in cell c of Y

$E_c =$ $(1 \times C)$ vector with 1 in cth entry, 0's elsewhere.

General Location Model

$$\Pr\left(w_i = E_c \mid \theta\right) = \pi_c, \qquad c = 1,\dots,C;\ \Sigma_c \pi_c = 1.$$

$$\left(x_i \,\middle|\, w_i = E_c, \theta\right) \sim_{\text{ind}} N_K\left(\mu_c, \Omega\right),\ k-\text{variate normal distribution}$$

with $\mu_c = \left(\mu_{c1},\dots,\mu_{cK}\right)$, cov matrix Ω

General Location Model

Special Cases:

(a) Y_1 binary with values 0 and 1, then

regression of Y_1 on other variables is logistic:

$$(y_{i1} \mid y_{i2}, ..., y_{iV}, x_i) \sim \text{BERNOULLI}(\theta_i), \theta_i = \frac{\exp(L_i)}{1 + \exp(L_i)}$$

With $V = 1$ and Y_1 is binary, this is two-group normal discriminant

analysis model, with categorical predictors also allowed

(b) $V = 1$ and Y_1 is categorical and completely missing,

model yields a form of parametric cluster analysis.

(c) X_1 continuous, regression of X_1 on other variables is

normal linear regression (continuous and categorical predictors)

ML with complete data

$$\ell(\Gamma,\Omega,\Pi) = \sum_{i=1}^{n} \log f\left(x_i \,\middle|\, w_i, \Gamma, \Omega\right) + \sum_{i=1}^{n} \log f\left(w_i \,\middle|\, \Pi\right)$$

$$= h(\Omega) - \tfrac{1}{2}\,\mathrm{tr}\left(\Omega^{-1}\sum_{i=1}^{n} x_i^T x_i\right) + \mathrm{tr}\,\Omega^{-1}\Gamma\left(\sum_{i=1}^{n} w_i^T x_i\right)$$

$$+ \sum_{c=1}^{C}\left[\left(\sum_{i=1}^{n} w_{ic}\right)\left(\log \pi_c - \tfrac{1}{2}\mu_c \Omega^{-1} \mu_c^T\right)\right],$$

ML estimates:

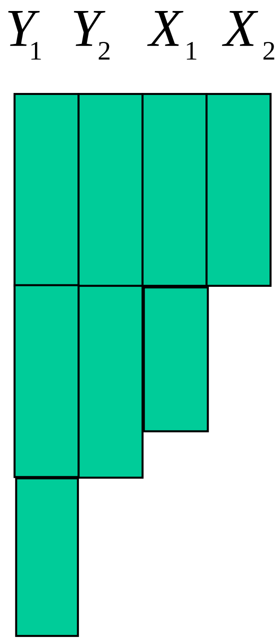
$\hat{\Pi} = n^{-1}\sum w_i$, (cell proportions)

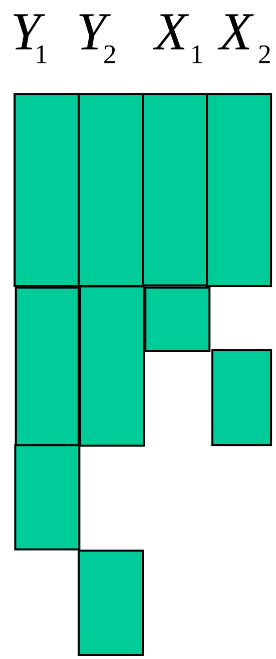
$\hat{\Gamma} = \left(\sum x_i^T w_i\right)\left(\sum w_i^T w_i\right)^{-1}$, (cell means)

$\hat{\Omega} = n^{-1}\sum\left(x_i - w_i\hat{\Gamma}\right)^T\left(x_i - w_i\hat{\Gamma}\right)$, (pooled within-cell cov matrix)

Factored Likelihoods

- For these patterns the factorization $[Y, X]=[Y]\text{x}[X|Y]$ in the general location model applies for multiple imputation





Y fully observed when any *X* is partially observed

- $Y_1 > Y_2 > X_1 > X_2$ $\qquad Y > X$

Fill in *Y*'s based on model for *Y*

Fill in *X*'s given *Y*'s based on model for *X* given *Y*

EM algorithm

- Combines aspects of EM for categorical and normal data
- Sweep again plays useful role for normal parts

Complete data is in exponential family with sufficient stats

$$\sum x_i^T x_i, \sum w_i^T x_i, \sum w_i$$

E-step: $x_{(0),i}$ = set of observed x_i, S_i = set of possible cells for unit i

$$T_{1i}^{(t)} = E(x_i^T x_i \,|\, x_{(0),i}, S_i, \theta^{(t)}),$$

$$T_{2i}^{(t)} = E(w_i^T x_i \,|\, x_{(0),i}, S_i, \theta^{(t)}),$$

$$T_{3i}^{(t)} = E(w_i \,|\, x_{(0),i}, S_i, \theta^{(t)}).$$

E-Step

$$\omega_{ic} = \Pr\left(w_i = E_c \,\middle|\, x_{(0),i}, S_i, \theta^{(t)}\right) = \frac{\exp(\delta_{ic})}{\sum_{d \in S_i} \exp(\delta_{id})}$$

$$\delta_{ic} = x_{(0)i}\Omega_{(0),i}^{-1}\mu_{(0),i}^{T} - \tfrac{1}{2}\mu_{(0)i}\Omega_{(0),i}^{-1}\mu_{(0),i}^{T} + \ell n(\pi_c)$$

$$E\left(w_{ic}x_{ij} \,\middle|\, x_{(0),i}, S_i, \theta^{(t)}\right) = \begin{cases} \omega_{ic}\hat{x}_{ij}^{(c)} & \text{if } x_{ij} \text{ is missing,} \\ \omega_{ic}x_{ij} & \text{if } x_{ij} \text{ is observed.} \end{cases}$$

where $\hat{x}_{ij}^{(c)} = E\left(x_{ij} \,\middle|\, x_{(0),i}, w_i = E_c, \theta^{(t)}\right)$,

E-Step

$$E\left(x_{ij}x_{ik}\,\middle|\,x_{(0),i}\,,S_i\,,\theta^{(t)}\right)$$

$$=\sum_{c\in S_i}\omega_{ic}E\left(x_{ij}x_{ik}\,\middle|\,x_{(0),i}\,,w_i=E_c\,,\theta^{(t)}\right)$$

$$=\begin{cases} x_{ij}x_{ik}, & x_{ij}, x_{ik} \text{ both observed;} \\ x_{ik}\sum_{c\in S_i}\omega_{ic}\hat{x}_{ij}^{(c)}, & x_{ij} \text{ missing, } x_{ik} \text{ observed;} \\ x_{ij}\sum_{c\in S_i}\omega_{ic}\hat{x}_{ik}^{(c)}, & x_{ik} \text{ missing, } x_{ij} \text{ observed;} \\ \sigma_{jk\cdot(0),i}+\sum_{c\in S_i}\omega_{ic}\hat{x}_{ij}^{(c)}\hat{x}_{ik}^{(c)}, & x_{ij}, x_{ik} \text{ both missing.} \end{cases}$$

Computations sweep an augmented cov matrix to make $x_{(0)i}$ independent -- see book for details.

EM Algorithm

M-step:

$$\Pi^{(t+1)} = n^{-1}\sum_{i=1}^{n} T_{3i}^{(t)},$$

$$\Gamma^{(t+1)} = D^{-1}\left(\sum_{i=1}^{n} T_{2i}^{(t)}\right),$$

$$\Omega^{(t+1)} = n^{-1}\left[\sum_{i=1}^{n} T_{1i}^{(t)} - \left(\sum_{i=1}^{n} T_{2i}^{(t)}\right)^{T} D^{-1}\left(\sum_{i=1}^{n} T_{2i}^{(t)}\right)\right],$$

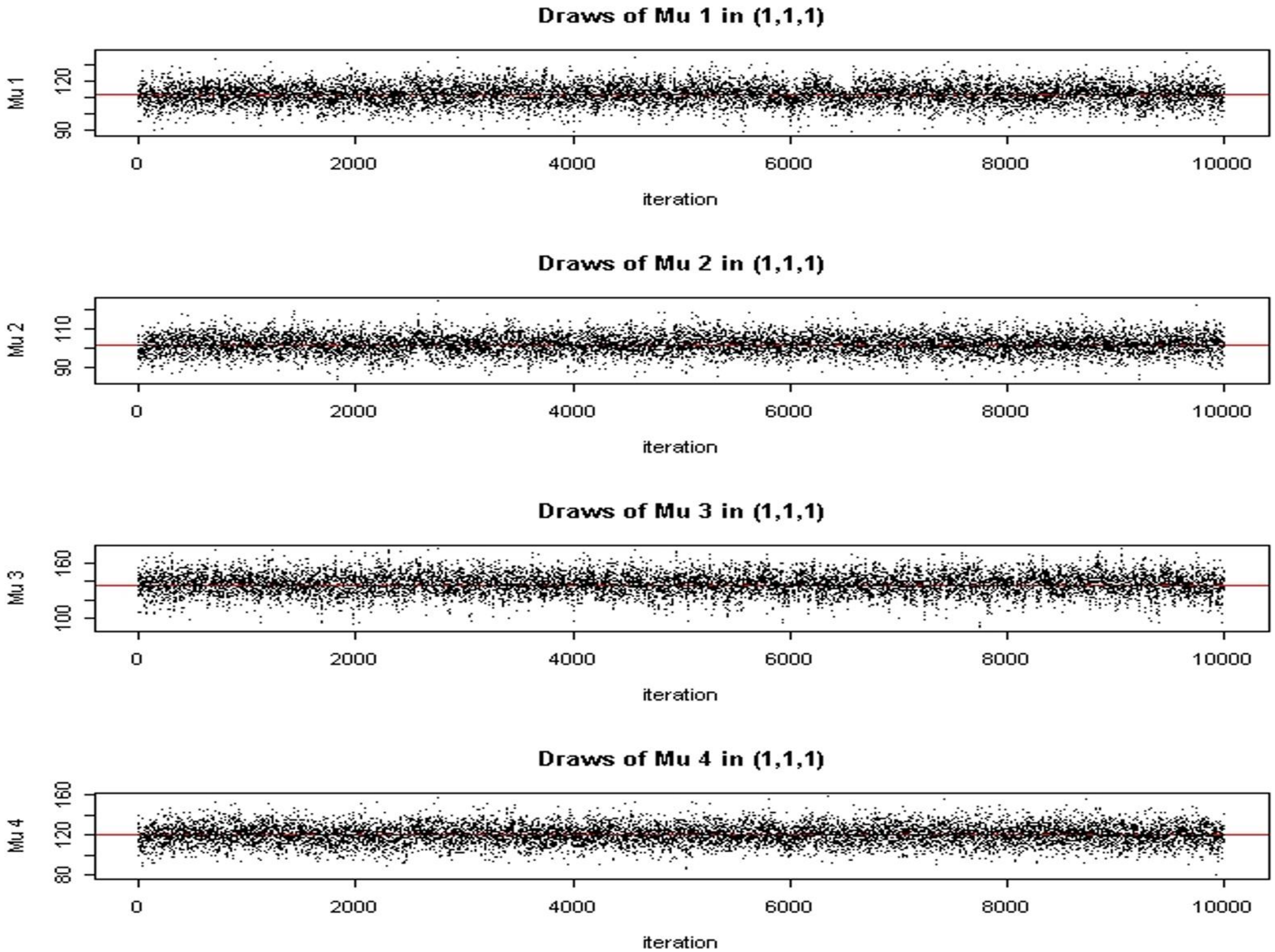
Bayes simulation

- Again I and P steps parallel the E and M steps for ML, involving Dirichlet and Inverse-Wishart / Normal draws
- See book Section 14.2.4 for details

St. Louis Risk Research Data

- See book examples 14.1, 14.2
- G = risk group of the parent (3 categories)
- D_1, D_2 = number of symptoms for first and second child (high or low)
- R_1, V_1, R_2, V_2 = reading and verbal comprehension for first and second child
- G, D_1, D_2 = for a 3-way contingency table with 3x2x2 = 12 cells
- R_1, V_1, R_2, V_2 are continuous.

Bayes output



Bayes output

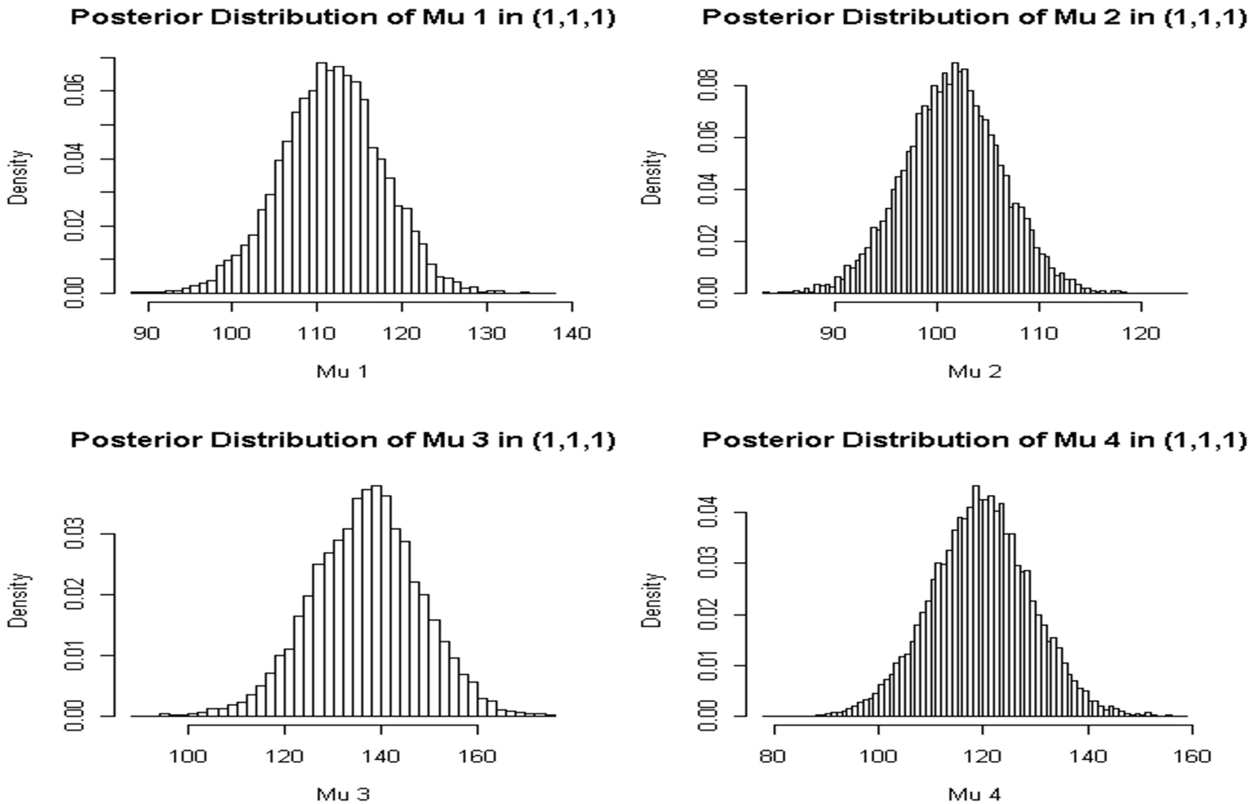
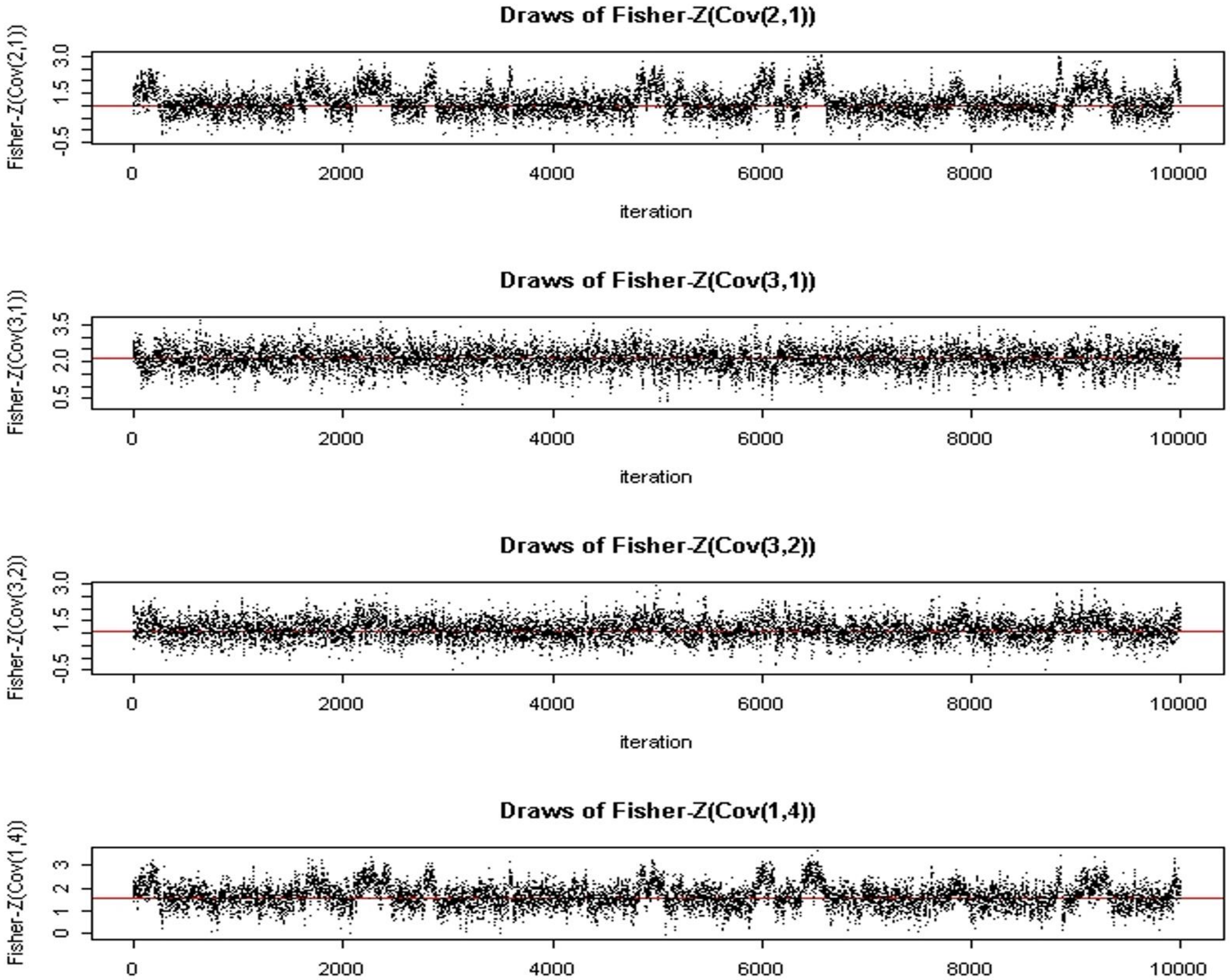


Figure 14.1A. Example 14.2, Above: Sequences of Draws from the Posterior Distributions of the Means (1 = R1, 2 = R2, 3 = V1, 4 = V2) in Cell (1,1,1). Below: Histograms of the Posterior Distributions for each Mean.

Bayes Output



Bayes Output

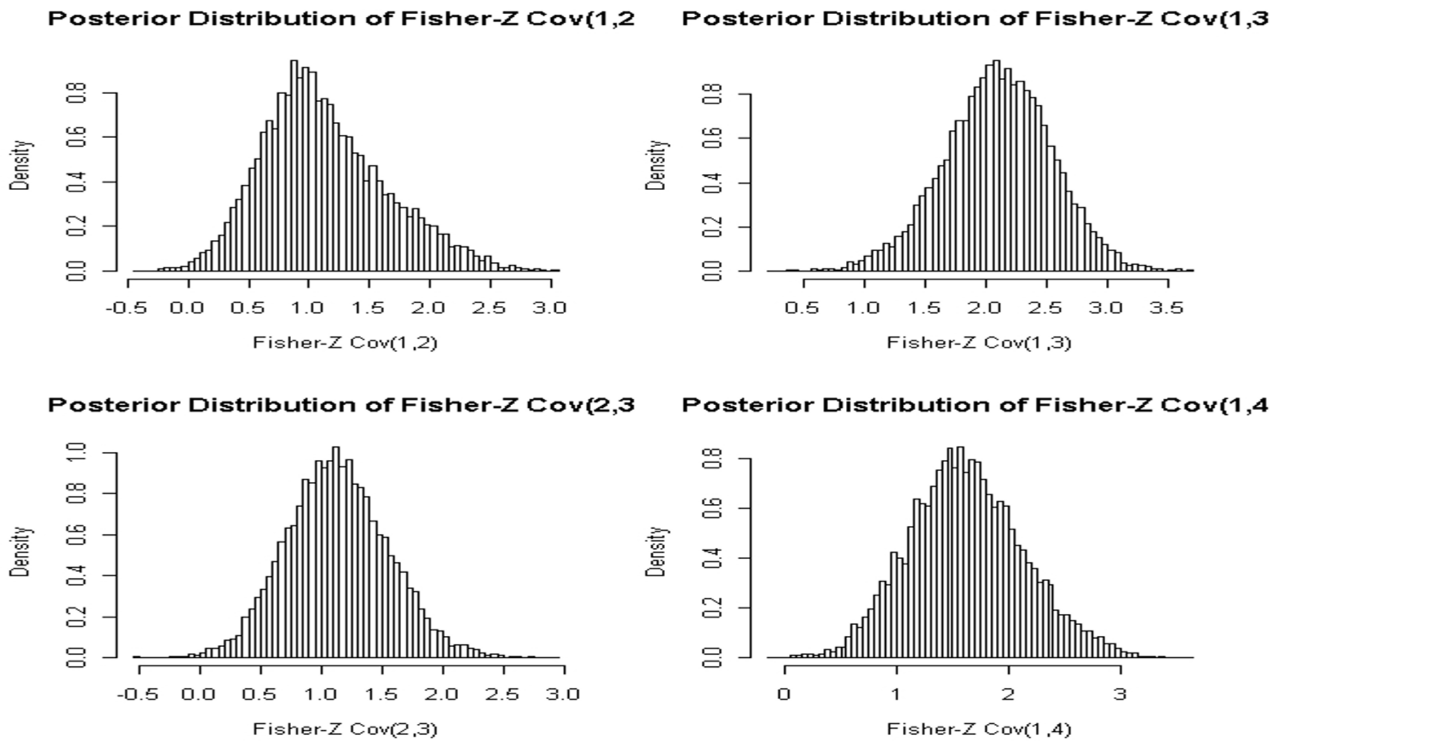


Figure 14.1B. Example 14.2. Sequences of Gibbs' Sampler Draws from the Fisher-Z Transformed Posterior Distributions of Selected Covariances. Below: Histograms of the Posterior Distributions for Each Parameter.

St Louis data results

- Gibbs' sampler sequences and histograms of draws look reasonably stable for the means
- Sequences for the covariances display some "jumpiness", reflecting lack of information to estimate some of these parameters.
- Prefer Bayesian results to ML, because they tend to average over plausible regions of the likelihood, and display the variability in the data.
- Bootstrap standard errors for the ML estimates (not shown here) are generally somewhat smaller than the posterior standard deviations, and are less reflective of the true variability in this sparse dataset.

General Location Model with Parameter Restrictions

- Categorical variables: loglinear models
- Continuous variables: anova-like restrictions on cell means – such as dropping higher-way interactions
- Within-cell covariance matrix of the continuous variables could be constrained (e.g. compound symmetry).
- As before, the E step of EM or the I step of Gibbs' is unaffected by these restrictions; M step of EM or P step of Gibbs' is for parameters of the constrained model

Summary

- The general location model is an interesting and flexible joint model that yields a variety of regression models, by conditioning on categorical or continuous variables
- The computational methods are elegant, but in practice this approach might be regarded as superceded by chained equation MI, given ready availability of software and the flexibility of this approach.