Mixed models: design of experiments

V. Fedorov

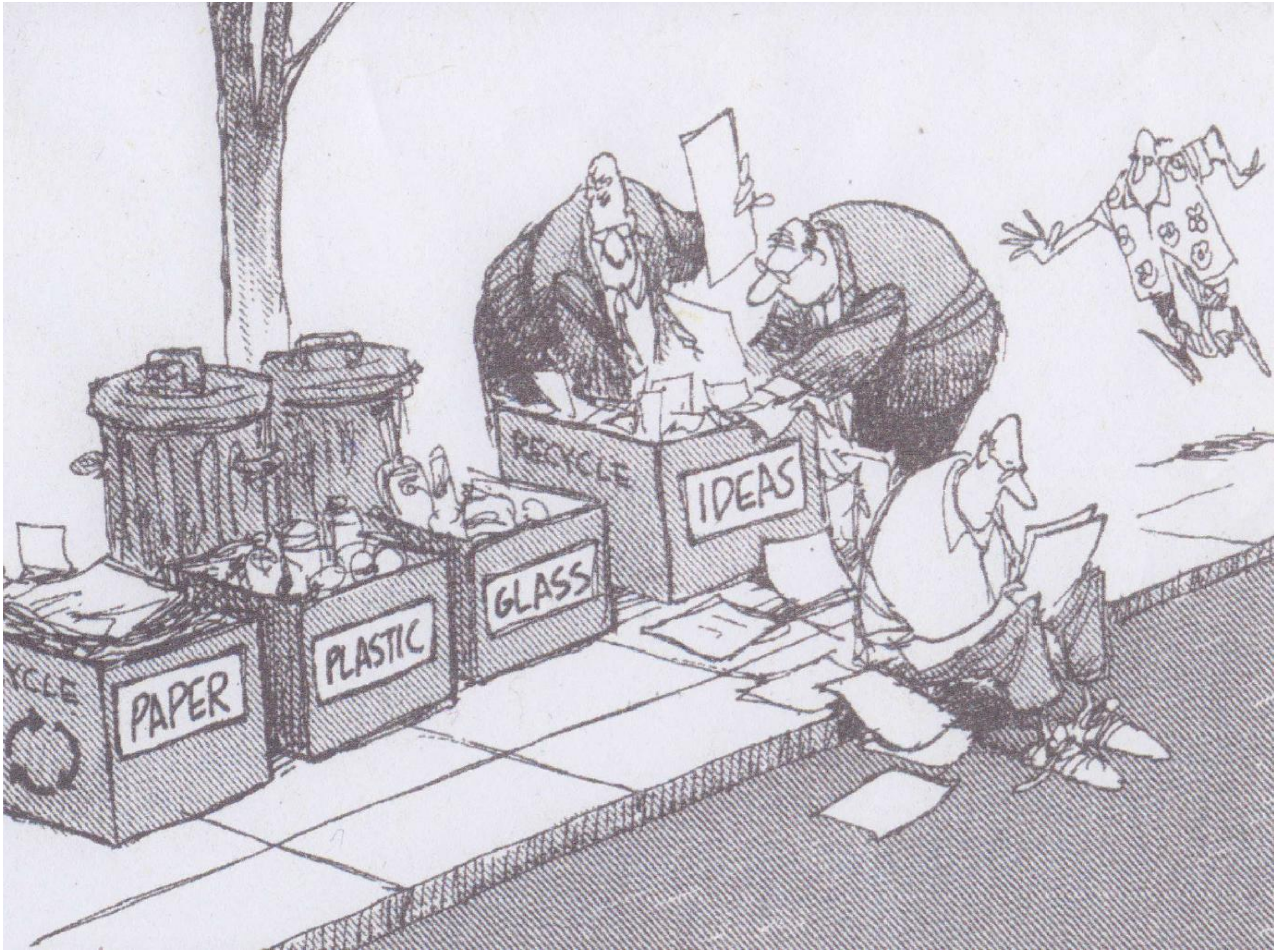
August, 2011

Selected references (estimation)

- Johnson L. (1977) Stochastic parameter regression: an annotated bibliography, International Statistical Review, 45, 257-272; (1980) 48, 95-102.
- Demidenko E. (2004) Mixed models, Wiley.
- Kiefer J. and Wolfowitz J. (1956) Consistency of the MLE in the presence of infinitely many incidental parameters, AMS, 27, 887-906.
- Rao C.R. (1965) The theory of least squares when parameters are stochastic …, Biometrika, 52, 447-458.
- Harville D. ML approaches to variance component estimation and related problems, JASA, 72, 320-340.
- Spjotvoll E. (1977) Random coefficients regression models. A review. Statistics, 8, 69-93.
- Gladitz J. and Pilz J. (1982) Construction of optimal designs in random coefficient regression models, Statistics, 13, 371-385.
- Mallet A. (1986) A MLE method for random coefficient regression models, Biometrika, 73, 645-656.
- Lindstrom M. and Bates D. (1988) N-R and EM algorithms for linear mixed-effects models… , JASA, 83, 1014-1022.

Selected references (design)

- Jennrich R. and Schluchter M. (1986) Unbalanced repeated measures models with structured covariance matrices. Biometrics,42, 805-820.
- Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data, *Biometrics* 46, 673–687.
- Pilz J. (1991) Bayesian Estimation and Experimental Design in Linear Regression Models, Wiley.
- Fedorov V., Hackl P. and Mueller W. (1993) Estimation and experimental design for second kind regression models, Informatik, Biometrie und Epidemiologie in Medizin und Biologie, 24,134-151.
- Mentre F., Mallet A. and Baccar D. (1997) Optimal design in random-effects regression models, Biometrika, 84, 429-442.
- Fedorov V., Gagnon R. and Leonov S. (2002) Design of experiments with unknown parameters in variance, Applied Stochastic Models in Business and Industry, 18, 207-218.
- Fedorov V. and Leonov S. (2004) Parameter estimation for models with unknown parameters in variance. Communication in Statistics.
- Dragalin V. and Fedorov V. (2006) Design of multi-centre trials with binary response, Statistics in Medicine, 25, 2701-2719


RECYCLE
PAPER
PLASTIC
GLASS
RECYCLE
IDEAS

Type I mixing

Observations are independent and

Model 1

$$\mathbf{y}_i \sim p(\mathbf{y} | \mathbf{x}_i, \theta)$$

Model 2

$$\mathbf{y}_i \sim p(\mathbf{y} | \mathbf{x}_i, \theta_i)$$

$$\theta_i \sim p(\theta \mid \Theta)$$

For transition from (2) to (1) use: $\mathbf{y} \sim p(y|x, \Theta) = \int p(\mathbf{y}|\mathbf{x}, \theta) p(\theta|\Theta) d\theta$

Model 3

$$\mathrm{E}(\mathbf{y}_i) = \eta(\mathbf{x}_i, \theta)$$

$$\mathrm{Var}(\mathbf{y}_i) = \mathbf{S}(\mathbf{x}_i, \theta)$$

Model 4

$$\mathrm{E}(\mathbf{y}_i) = \eta(\mathbf{x}_i, \theta_i)$$

$$\mathrm{Var}(\mathbf{y}_i) = \mathbf{V}(\mathbf{x}_i, \theta_i)$$

$$\mathrm{E}(\theta_i) = \theta$$

$$\mathrm{Var}(\theta_i) = \Sigma(\theta)$$

For transition from (4) to (3) use:

$$\mathrm{E}(\mathbf{y}) = \mathrm{E}_\theta[\mathrm{E}_\mathbf{y}(\mathbf{y}|\theta)]$$

$$\mathrm{Var}(\mathbf{y}) = \mathrm{E}_\theta[\mathrm{Var}_\mathbf{y}(\mathbf{y}|\theta)] + \mathrm{Var}_\theta[\mathrm{E}_\mathbf{y}(\mathbf{y}|\theta)]$$

Type II mixing

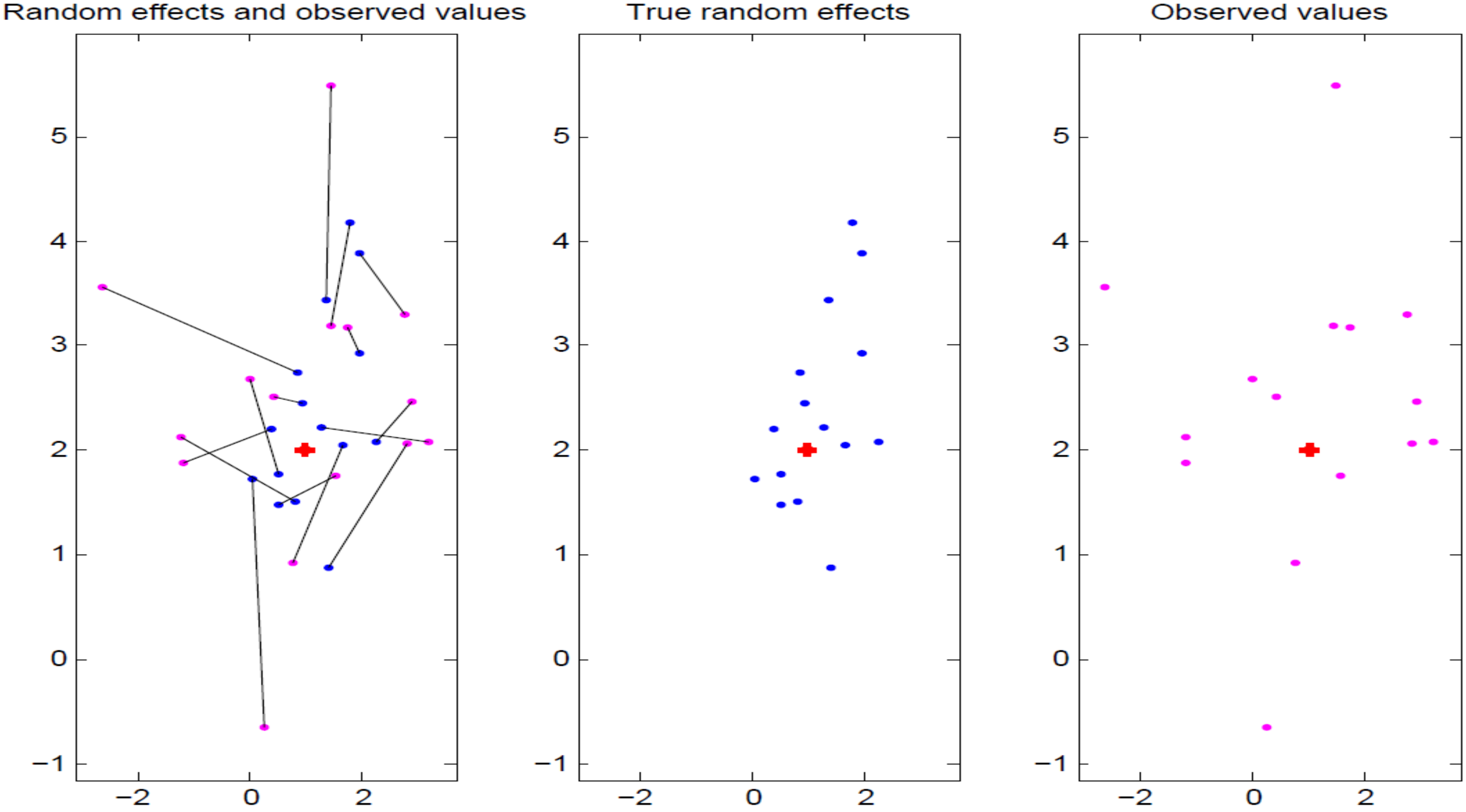
$$\mathbf{y}_{ij} \sim p(\mathbf{y} | \theta_i)$$
$$\theta_i \sim p(\theta \mid \vartheta(\mathbf{x}_i, \Theta))$$

Two major problems

Problem 1: Not much is known about population. A substantial number of observations is needed to understand its distribution.

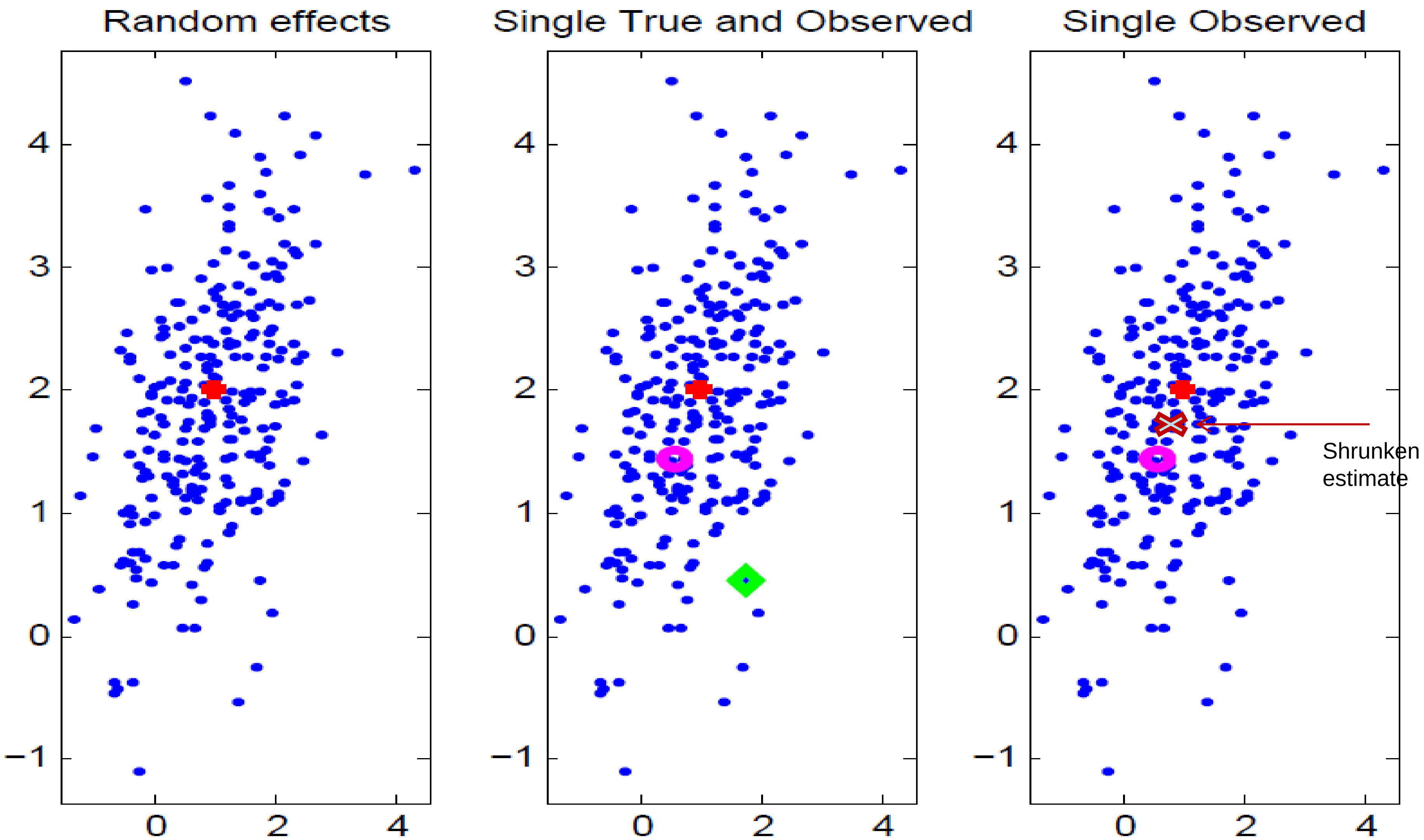
Problem 2: Population is well known, individual parameters for item “i” should be predicted. How to make it with relatively few observations

Learning about population

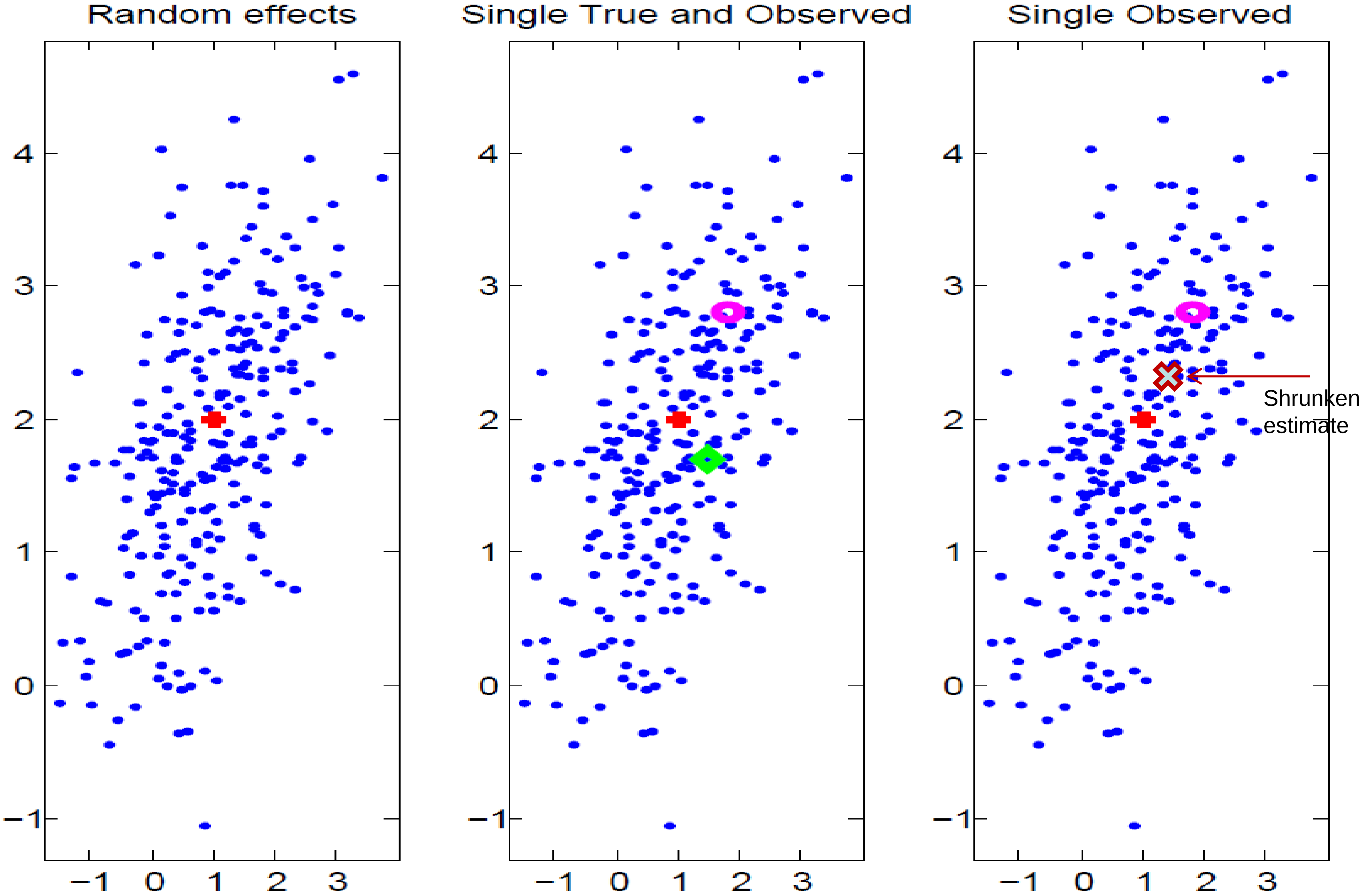


Should we spend more time making more “points” or making shorter “segments”?

Predicting for patients



Predicting for patients



Short intro to optimal design

- Model $\mathbf{Y} \sim p(\mathbf{y} | \mathbf{t}, \mathbf{x}, \theta)$

  observed uncontrolled controlled unknown

- Utility $\mathbf{U}(\mathbf{t}, \mathbf{x}, \theta)$

- Cost $\varphi(\mathbf{t}, \mathbf{x}, \theta)$

- Design region $\mathbf{x} \in \mathscr{X}$

Estimation: MLE

Likelihood function:

$$\mathscr{L}_N(\theta) = \prod_{i=1}^{n} \prod_{j=1}^{r_i} p(\mathbf{y}_{ij}|\mathbf{x}_i, \theta)$$

Maximum likelihood estimator:

$$\hat{\theta}_N = \arg\max_{\theta \in \Omega} \prod_{i=1}^{n} \prod_{j=1}^{r_i} p(\mathbf{y}_{ij}|\mathbf{x}_i, \theta)$$

Asymptotic normality:

$$(\hat{\theta}_N - \theta_0)/\sqrt{N} \sim \mathscr{N}(0, D(\xi, \theta_0))$$

Normalized var-cov matrix:

$$D(\xi, \theta) = M^{-1}(\xi, \theta)$$

Design:

$$\xi = \{\mathbf{x}_i,\ w_i\}_1^n \quad , \quad w_i = r_i/N$$

Total cost:

$$Cost = N\Phi(\xi) = N\sum_{i=1}^{n} w_i \varphi(\mathbf{x}_i)$$

Estimation: Information matrix

$$NM(\xi,\theta_0)=\sum_{i=1}^{n} n_i\mu(\mathbf{x}_i,\theta_0)=N\sum_{i=1}^{n} w_i\mu(\mathbf{x}_i,\theta_0)$$

Information matrix of a “single” observation:

$$\mu(\mathbf{x},\theta_0)=\mathrm{E}\{\zeta(\mathbf{y}|\mathbf{x},\theta)\zeta^T(\mathbf{y}|\mathbf{x},\theta)\}_{\theta=\theta_0}$$

Score function: $\zeta(y|x,\theta)=\frac{\partial}{\partial\theta}\log p(y|x,\theta)$

For transformed parameters:

$$M(\xi,\theta)=JM(\xi,\vartheta(\theta))J^T$$

$$J=\frac{\partial\vartheta^T}{\partial\theta}=\left\|\frac{\partial\vartheta_\alpha}{\partial\theta_\beta}\right\|_{1,1}^{m',m}$$

Main optimization problem

Optimal design:

$$\xi^* = \arg \min_{\xi \in \Xi(\mathscr{X})} \Psi(N(\xi)M(\xi,\theta))$$
$$\text{s.t.} \quad N(\xi)\Phi(\xi) \leq C$$

Equivalently:

$$\xi^* = \arg \min_{\xi \in \Xi(\mathscr{X})} \Psi\left(\frac{M(\xi,\theta)}{\Phi(\xi)}\right)$$

The same but with “prior” information

Optimal design:

$$\xi^* = \arg \min_{\xi \in \Xi(\mathscr{X})} \Psi(\mathbf{M}_0 + N(\xi)M(\xi,\theta))$$
$$\text{s.t.} \quad N(\xi)\Phi(\xi) \le C$$

Equivalently:

$$\xi^* = \arg \min_{\xi \in \Xi(\mathscr{X})} \Psi\left(\frac{\mathbf{M}_0}{C} + \frac{M(\xi,\theta)}{\Phi(\xi)}\right)$$

Popular optimality criteria

- D-criterion: $$\Psi(D) = |D|$$

- Linear criterion: $$\Psi(D) = \mathrm{tr}\, AD = \mathrm{tr}\, LDL^T$$

- E-criterion: $$\Psi(D) = \lambda_{max}(D)$$

For any design:

$$|D|^{1/m} \le m^{-1}\, \mathrm{tr}\, D \le \lambda_{max}(D)$$

$$|D|^{1/m} \le m^{-1}\, \mathrm{tr}\, AD, \quad |A| = 1$$

$$L^T DL \le \lambda_{max}(D), \quad L^T L = 1$$

Basic “design” formulae

Necessary and sufficient conditions (D-criterion):

$$\mathrm{tr}[\mu(\mathbf{x},\theta)M^{-1}(\xi^*,\theta)] \le m\varphi(\mathbf{x})/\Phi(\xi^*)$$

First order algorithm (D-criterion)

Step forward:

$$\mathbf{x}_{s+1}^{\oplus} = \arg\max_{\mathbf{x}\in\mathscr{X}} \{\mathrm{tr}[\mu(\mathbf{x},\theta)M^{-1}(\xi_s,\theta)] - m\varphi(\mathbf{x})/\Phi(\xi_s)\}$$

Step backward:

$$\mathbf{x}_{s+1}^{\ominus} = \arg\min_{\mathbf{x}\in\mathscr{X}_s} \{\mathrm{tr}[\mu(\mathbf{x},\theta)M^{-1}(\xi_s,\theta)] - m\varphi(\mathbf{x})/\Phi(\xi_s)\}$$

Note. If there is a prior information then use:

$$\mathrm{tr}\{[\mu(\mathbf{x},\theta)+\mathbf{M}_0\varphi(\mathbf{x})/C][M(\xi^*,\theta)+\mathbf{M}_0\Phi(\xi^*)/C]^{-1}\} \le m\varphi(\mathbf{x})/\Phi(\xi^*)$$

Specific cases from 1988, A&F

Table 1: Dual problems for the most used criteria.

$\Psi(M)$	$\varphi(x, \xi)$
$\ln\|M^{-1}\|$	$m - tr\ M^{-1} m(x)$
$\ln\|A^T M^{-1} A\|$	$s - tr\ M^{-1} A [A^T M^{-1} A]^{-1} A^T M^{-1} m(x)$, $s = rank\ A$
$tr\ A^T M^{-1} A$	$tr\ A^T M^{-1} A - tr\ M^{-1} A\ A^T M^{-1} m(x)$
$\int_{X_0} w(x) d(x, \xi) dx$ $d(x, \xi) = tr\ M^{-1} m(x)$	$\int_{X_0} w(x) d(x, \xi) dx - \int_{X_0} w(x') tr\ M^{-1} m(x') M^{-1} m(x) dx'$
$m^{-1} (tr\ M^{-p})^{1/p}$	$tr\ M^{-p} - tr\ M^{-p-1} m(x)$

Optimal Design Construction

Information matrix of a single observation, cost function, design region (candidate points), optimality criterion

↓

DESIGN  *ENGINE*

↓

Optimal design, comparison of different designs, sample size determination, …

Linear case with known **V** and ∑

$$E(\mathbf{y}_i) = \eta(\mathbf{x}_i, \theta_i) = \mathbf{F}^T(\mathbf{x}_i)\theta_i, \quad \mathrm{Var}(\mathbf{y}_i) = \mathbf{V}$$

$$E(\theta_i) = \theta, \quad \mathrm{Var}(\theta_i) = \Sigma$$

$$\hat{\theta} = \left[\sum_i [\mathbf{M}^{-1}(\mathbf{x}_i) + \Sigma]^{-1}\right]^{-1} \sum_i [\mathbf{M}^{-1}(\mathbf{x}_i) + \Sigma]^{-1} \bar{\theta}_i$$

$$\mathbf{M}(\mathbf{x}_i) = \mathbf{F}(\mathbf{x})\mathbf{V}^{-1}\mathbf{F}^T(\mathbf{x})$$

$$\bar{\theta}_i = \mathbf{M}^{-1}(\mathbf{x}_i)\mathbf{F}(\mathbf{x}_i)\mathbf{y}_i$$

$$\hat{\theta}_i = [\mathbf{M}(\mathbf{x}_i) + \Sigma^{-1}]^{-1}[\mathbf{M}(\mathbf{x}_i)\bar{\theta}_i + \Sigma^{-1}\theta]$$

$$\hat{\hat{\theta}}_i = [\mathbf{M}(\mathbf{x}_i) + \Sigma^{-1}]^{-1}[\mathbf{M}(\mathbf{x}_i)\bar{\theta}_i + \Sigma^{-1}\hat{\theta}]$$

Estimation of population parameters: information matrix for a "single" observation

$$\mu_{\alpha\beta}(\mathrm{x},\theta) = \frac{\partial \eta^T(\mathrm{x},\theta)}{\partial \theta_\alpha} \mathrm{S}^{-1}(\mathrm{x},\theta) \frac{\partial \eta(\mathrm{x},\theta)}{\partial \theta_\beta} + \frac{1}{2}\,\mathrm{tr}\left[\mathrm{S}^{-1}(\mathrm{x},\theta)\frac{\partial \mathrm{S}(\mathrm{x},\theta)}{\partial \theta_\alpha}\mathrm{S}^{-1}(\mathrm{x},\theta)\frac{\partial \mathrm{S}(\mathrm{x},\theta)}{\partial \theta_\beta}\right]$$

Linear case: $\eta(\mathbf{x},\theta) = \mathbf{F}^T(\mathbf{x})\theta$

$$\mathbf{S}(\mathbf{x},\Sigma,\mathbf{V}(\mathbf{x})) = \mathbf{F}^T(\mathbf{x})\Sigma\mathbf{F}(\mathbf{x}) + \mathbf{V}(\mathbf{x})$$

$$\mu(\mathbf{x}) = \mathbf{F}(\mathbf{x})[\mathbf{F}^T(\mathbf{x})\Sigma\mathbf{F}(\mathbf{x}) + \mathbf{V}(\mathbf{x})]^{-1}\mathbf{F}^T(\mathbf{x})$$

If **M** is regular then:

$$\mu(\mathbf{x}) = [\mathbf{M}^{-1}(\mathbf{x}) + \Sigma]^{-1}$$

Note: Go to slides 16-17 to learn how to build an optimal design

Estimation of individual parameters

- *The best unbiased prediction for i-th item:*

maximize $|\mathbf{M}(\mathbf{x})|$

$$\mu(x) = \mathbf{f}(x)\mathbf{f}^T(x)$$
$$\psi(x) = \operatorname{tr}\mu(x)\mathbf{M}(\xi) - m$$

- *For the best individual prediction on average:*

maximize $|\mathbf{M}^{-1}(\mathbf{x}) + \Sigma|^{-1}$

$$\mu(x) = \mathbf{f}(x)\mathbf{f}^T(x)$$
$$\psi(x) = \operatorname{tr}\{[\mu(\mathbf{x}) + n\Sigma^{-1}][M(\xi^*) + n\Sigma^{-1}]^{-1}\} - m$$

Unknown variances

Parameters $\Theta = \{\theta, \Sigma, \mathbf{V}\}$, where $\mathbf{V} = \sigma^2 \mathbf{I}$ are unknown.

Information matrix of a “single” observation:

$$\mu(\mathbf{x}, \Theta) = \begin{pmatrix} \mu_{\theta\theta}(\mathbf{x}) & \mathbf{0}_{m,m} & \mathbf{0}_{m,1} \\ \mathbf{0}_{m,m} & \mu_{\Sigma\Sigma}(\mathbf{x}) & \mu_{\Sigma\sigma}(\mathbf{x}) \\ \mathbf{0}_{1,m} & \mu_{\Sigma\sigma}^T(\mathbf{x}) & \mu_{\sigma\sigma}(\mathbf{x}) \end{pmatrix}$$

$$\mu_{\theta\theta}(\mathbf{x}) = \mathbf{F}(\mathbf{x})\mathbf{S}^{-1}(\mathbf{x})\mathbf{F}^T(\mathbf{x})$$

$$\{\mu_{\Sigma\Sigma}(\mathbf{x})\}_{\alpha\beta} = \frac{1}{2}\left[\mathbf{F}_\alpha(\mathbf{x})\mathbf{S}^{-1}\mathbf{F}_\beta^T(\mathbf{x})\right]^2 \qquad \{\mu_{\Sigma\sigma}(\mathbf{x})\}_\alpha = \frac{1}{2}\mathbf{F}_\alpha(\mathbf{x})\mathbf{S}^{-2}\mathbf{F}_\alpha^T(\mathbf{x})$$

$$\mu_{\sigma\sigma}(\mathbf{x}) = \frac{1}{2}\,\mathrm{tr}\left[\mathbf{S}^{-2}\right]$$

$$\mathbf{F}(\mathbf{x}) = [f(x_1), \ldots, f(x_k)]$$

Summary

- What will the mixed effects model be used for?
- Definition of a “single” observation.
- Derivation of its information matrix.
- Understanding the cost structure.
- Computing
- Benchmarking