Bayesian Estimation of Reproducibility/Repeatability in Early Drug Development

Dai Feng
Biometrics Research Dept.
Merck Research Lab
joint work with Richard Baumgartner and Vladimir Svetnik

NCB 2017

June 13, 2017

Outline

- Reproducibility/repeatability
- Why do we need a Bayesian method
- How to provide a Bayesian solution
- Potential issues and future work
- Summary

Reproducibility/Repeatability

- "Repeatability assesses pure random error due to "true" replications"
- "Reproducibility assesses closeness between observations made under condition other than pure replication, e.g., by different labs or observers." [Barnhart et al., 2007]

We have witnessed an increasing need to study the reproducibility/repeatability in early drug development due to the recent booming of promising new tools and technologies.

Within-Subject Coefficient of Variation (WSCV)

Definition—[Quan and Shih, 1996]

- Consider a one-way random effect model:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

$\alpha_i \sim N(0, \sigma_\alpha^2), \epsilon_{ij} \sim N(0, \sigma_\epsilon^2), \alpha_i \perp \epsilon_{ij}$

- Define WSCV as

$$\theta = \frac{\sigma_\epsilon}{\mu}$$

WSCV

Conjugate Prior

$$
\begin{aligned}
\mu &\sim N(0, 1e4) \\
\frac{1}{\sigma_\alpha^2} &\sim \Gamma(1e-2, 1e-2) \\
\frac{1}{\sigma_\epsilon^2} &\sim \Gamma(1e-2, 1e-2)
\end{aligned}
$$

Jeffreys Prior

The prior is proportional to the square root of the determinant of the Fisher information

$$\pi(\theta) = \sqrt{|I(\theta)|}$$

- Invariant under re-parameterization
  - For an alternate parameterization $\psi = h(\theta)$,

$$\pi(\psi) = \pi(\theta)|\frac{d\theta}{d\psi}|$$

  - The prior probability over any region will be invariant for all ways of choosing the parameters
- Generalization of flat priors in some parameterization
- Independent of data

Reference Prior

- History
  - Proposed by Jose Bernardo in 1979
  - Further developed by Jim Berger and others from 1980's till now
  - Brought about the concept of "Objective Bayesian"
- What is it
  - Quantify what exactly we mean by "non-informative" prior
  - Maximize some measures of distance between the posterior and prior to allow the data have the maximum effect on the posterior estimates
  - It is equal to Jeffreys prior under regularity conditions
- With nuisance parameters
  - For this case $\{(\mu, \sigma_\epsilon^2), \sigma_\alpha^2\}$, we are interested in (μ, σ_ϵ^2) and σ_α^2 is a nuisance parameter.
  - The prior is different than other setup, for example $\{(\mu, \sigma_\alpha^2), \sigma_\epsilon^2\}$.

WSCV

Jeffreys and Reference Prior

- The priors have the form

$$\left(\frac{1}{\sigma_\epsilon^2}\right)^a \left(\frac{1}{d\sigma_\alpha^2 + \sigma_\epsilon^2}\right)^b,$$

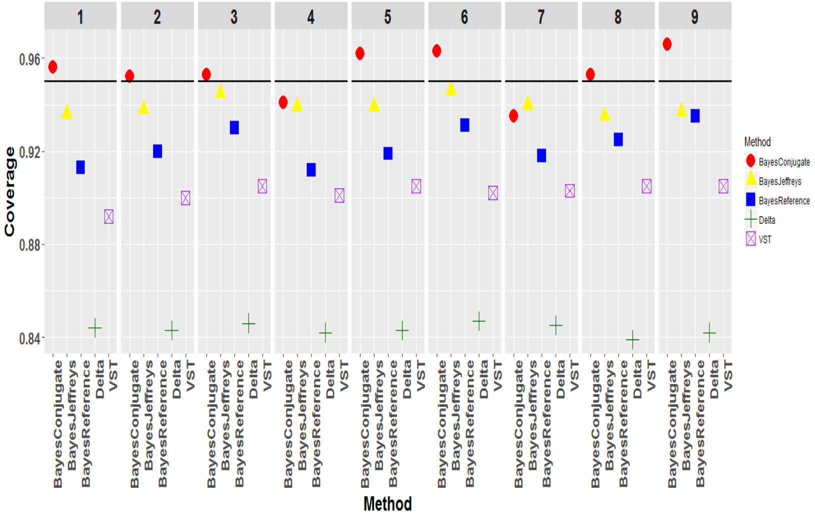
  with (a, b) equal to $(1, 3/2)$ and $(5/4, 1)$, respectively, for Jeffreys prior and reference prior.

- *Note that there is no need to run MCMC to obtain the posteriors.*

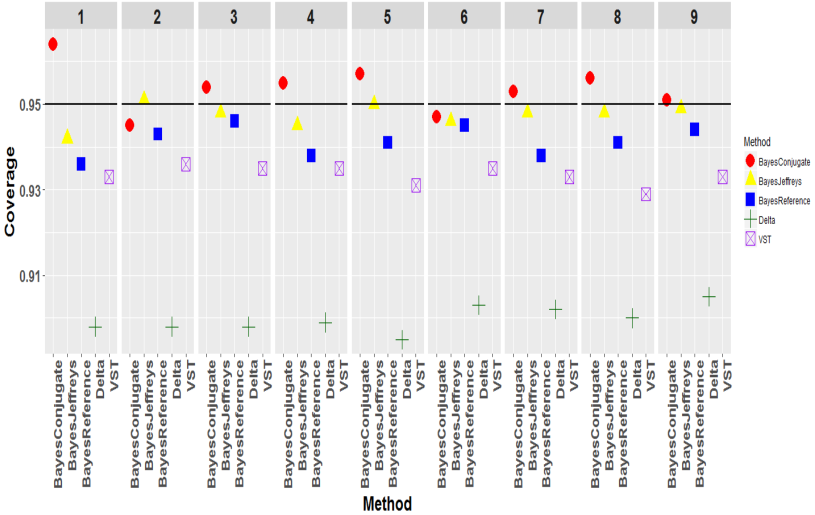
Setup for two raters normal data

Comb.	μ	σ_α	σ_ϵ
1	10	1	1
2	10	2	1
3	10	3	1
4	10	1	2
5	10	2	2
6	10	3	2
7	10	1	3
8	10	2	3
9	10	3	3

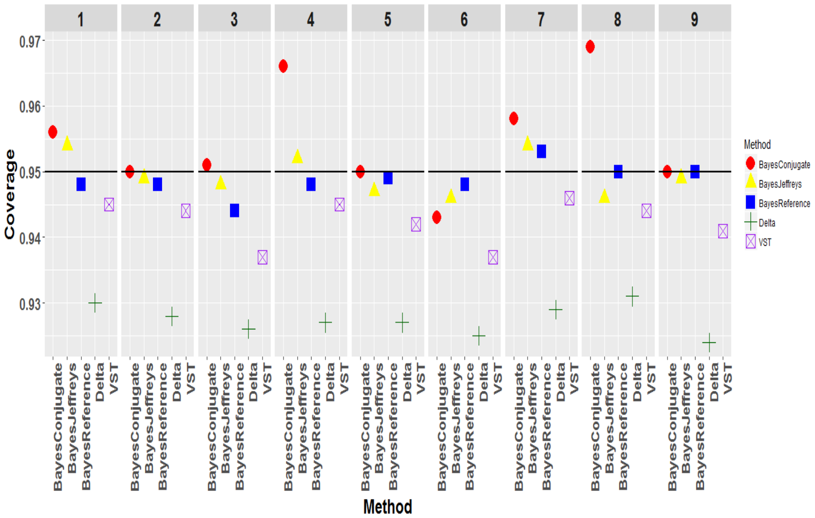
Coverage probability of nominal 95% credible/confidence intervals for WSCV when $n = 5$



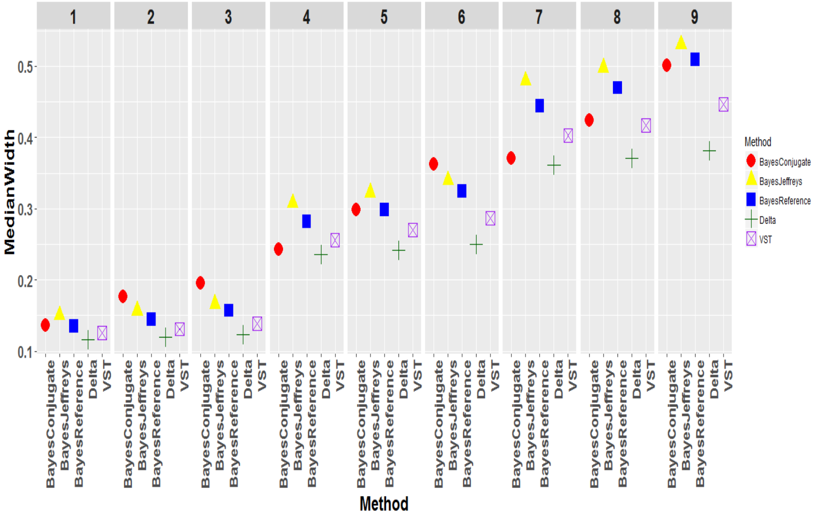
Coverage probability of nominal 95% credible/confidence intervals for WSCV when $n = 10$



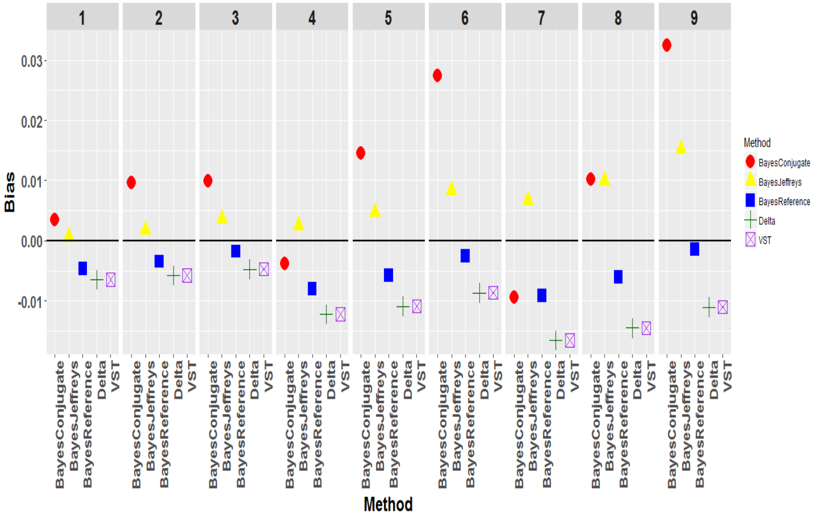
Coverage probability of nominal 95% credible/confidence intervals for WSCV when $n = 15$



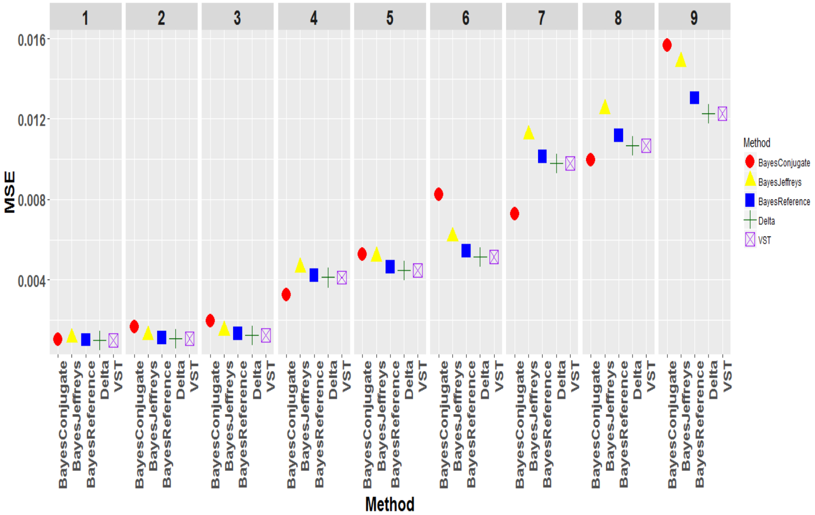
Median width of nominal 95% credible/confidence intervals for WSCV when $n = 5$



Bias of point estimates for WSCV when $n = 5$



MSE of point estimates for WSCV when $n = 5$



Intraclass Correlation Coefficient (ICC)

Definition—[Shrout and Fleiss, 1979]

The correlation between two readings from the same subject i, Y_{ij} and $Y_{ij'}$. There are different types of ICCs.

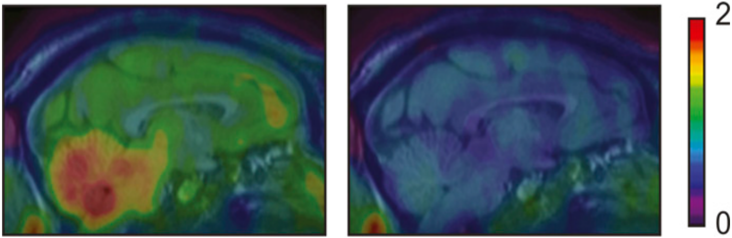
- ICC_1 is based on a one-way random effect model without observer effect

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

$$\alpha_i \sim N(0, \sigma_\alpha^2), \epsilon_{ij} \sim N(0, \sigma_\epsilon^2), \alpha_i \perp \epsilon_{ij}$$

$$ICC_1 = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}$$

A Positron Emission Tomography (PET) example

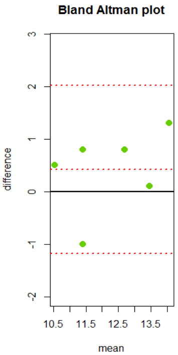


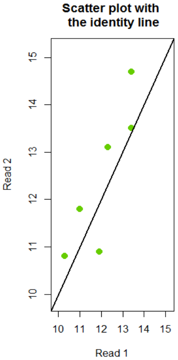
Averaged (0–90 minutes) PET images (overlaid on MRI image) of [^{11}C]MK-4232 in rhesus monkeys at baseline (left image) and after MK-3207 administration (right image). Color scale is in SUV [Hostetler et al., 2013].

Robust Bayesian Estimate

- It is well-known that even"high-quality" samples demonstrate small but noticeable deviation from the normal distribution in terms of having longer tails and slight skewness [Barnett and Lewis, 1984, Hampel et al., 2011].
- The t distribution provides a useful extension of the normal for robust modeling of data [Lange et al., 1989].

A Positron Emission Tomography (PET) Example (cont'd)





$$
\begin{array}{rrcl}
 & Y_{ij} & = & \mu + \alpha_i + \epsilon_{ij} \\
1 & \alpha_i & \sim & N(0, \sigma^2_\alpha) \\
2.1 & \epsilon_{ij} & \sim & N(0, \sigma^2_\epsilon) \\
2.2 & \epsilon_{ij} & \sim & t(0, \sigma^2_\epsilon, \nu)
\end{array}
$$

ICC_1

Method	Point	CI	DIC
normal-normal	0.80	(0.05, 0.99)	29.17
normal-t	0.78	(0.06, 0.99)	17.22

ICC_2

ICC_2 is based on a two-way random or mixed effect model with observer effect β_j

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

and

$$ICC_2 = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\epsilon^2}$$

where

- $\alpha_i \sim N(0, \sigma_\alpha^2), \epsilon_{ij} \sim N(0, \sigma_\epsilon^2), \alpha_i \perp \epsilon_{ij}$
- If β_j is fixed, then $\sigma_\beta^2 = \sum_{j=1}^k \beta_j^2/(k-1)$ is used with constraint of $\sum_{j=1}^k \beta_j = 0$
- If β_j is random, assume $\beta_j \sim N(0, \sigma_\beta^2)$ and $\alpha_i, \beta_j, \epsilon_{ij}$ are mutually independent.

Concordance Correlation Coefficient (CCC)

In a widely cited paper, Lin (1989), the CCC was proposed to evaluate the degree to which pairs fall on the 45° line

- Assume bivariate data was distributed under a distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Then define CCC as

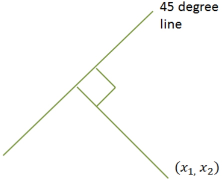
$$\begin{aligned} \rho_c &= 1 - \frac{E(x_1 - x_2)^2}{E(x_1 - x_2)^2 \quad \text{when } x_1 \text{ and } x_2 \text{ are uncorrelated}} \\ &= \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2} \end{aligned}$$

Concordance Correlation Coefficient (CCC)

In a widely cited paper, Lin (1989), the CCC was proposed to evaluate the degree to which pairs fall on the 45° line

- Assume bivariate data was distributed under a distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$



Then define CCC as

$$\rho_c = 1 - \frac{E(x_1 - x_2)^2}{E(x_1 - x_2)^2 \text{ when } x_1 \text{ and } x_2 \text{ are uncorrelated}}$$

$$= \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2}$$

Concordance Correlation Coefficient (CCC)

In a widely cited paper, Lin (1989), the CCC was proposed to evaluate the degree to which pairs fall on the 45° line

- Assume bivariate data was distributed under a distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

  Then define CCC as

$$\begin{aligned} \rho_c &= 1 - \frac{E(x_1 - x_2)^2}{E(x_1 - x_2)^2 \quad \text{when } x_1 \text{ and } x_2 \text{ are uncorrelated}} \\ &= \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2} \end{aligned}$$

- Bivariate normal was assumed to facilitate inference.

CCC (cont'd)

CCC has been generated to multiple observers for data with replications

- $$\frac{2\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\sigma_{ij}}{(p-1)\sum_{i=1}^{p}\sigma_i^2+\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}(\mu_i-\mu_j)^2} \quad p>2$$
- total CCC, inter-CCC, and intra-CCC, etc.

Prior for CCC under Normal Assumption

- Conjugate prior

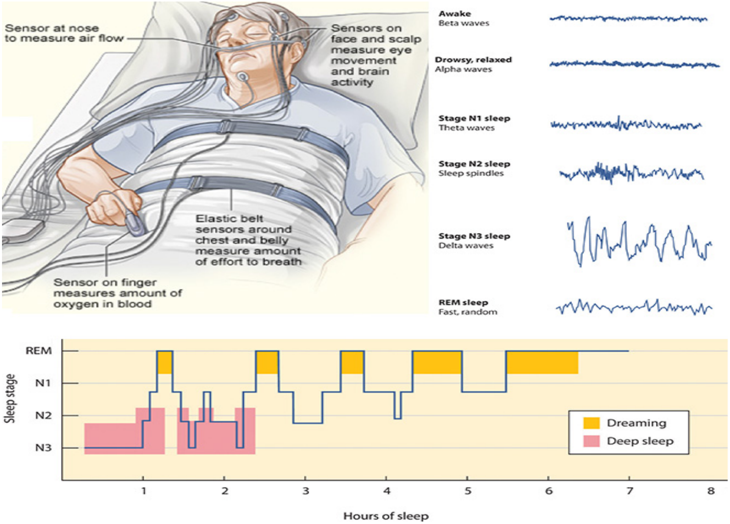
$$\begin{aligned} \boldsymbol{\mu} &\sim N(\mathbf{0}, \boldsymbol{I}) \\ \Sigma^{-1} &\sim Wishart(\boldsymbol{I}, k) \end{aligned}$$

- Jeffreys prior for bivariate normal

$$\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)^2}$$

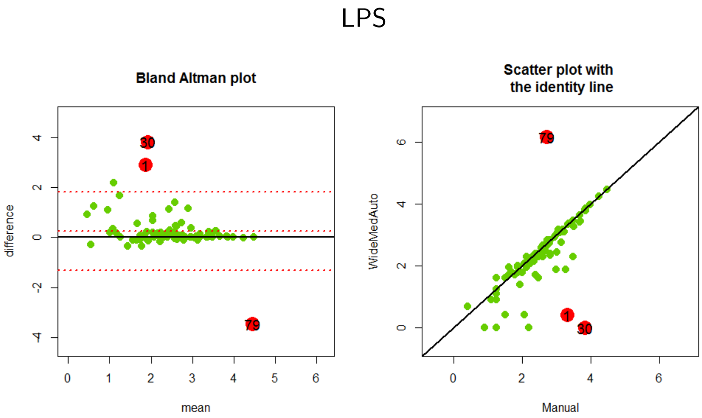
Note that it is straightforward to obtain independent samples from the posteriors.

An Electroencephalogram (EEG) Example



https://www.medicwiz.com/medtech/diagnostics/9-types-of-eeg-tests-everything-about-brainwave-monitoring

An EEG Example



Robust Bayesian Estimate of the CCC by Multivariate t Distribution

- The t distribution is widely used in statistics for robust inference especially when the data contain values that look like outliers [Liu, 1994, Berger, 1994].
- The Bayesian estimate is more reliable inferentially than the MLEs specially with small sample size [Liu, 1994].

Bayesian Treatment of the Multivariate t Distribution

We adopt the widely used representation as a scale mixture of normal distributions. Let $\boldsymbol{X}_i$ denote a vector of p dimensions

$$p(\boldsymbol{X_i}|\boldsymbol{\mu}, \Sigma, \lambda_i) \sim N(\boldsymbol{\mu}, \lambda_i \Sigma^{-1}) \quad (1)$$

$$p(\lambda_i|\nu) \sim \Gamma(\nu/2, \nu/2) \quad (2)$$

Then the marginal distribution of $\boldsymbol{X}_i$ has central multivariate t distribution with ν degrees of freedom and parameters $\boldsymbol{\mu}$ and Σ with the density function

$$f(\boldsymbol{X}_i|\boldsymbol{\mu}, \Sigma, \nu) = \frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\nu/2)(\nu\pi)^{p/2}|\Sigma|^{1/2}[1 + \frac{1}{\nu}(\boldsymbol{X} - \boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})]^{(\nu+p)/2}}$$

Bayesian Treatment of the Multivariate t Distribution (cont'd)

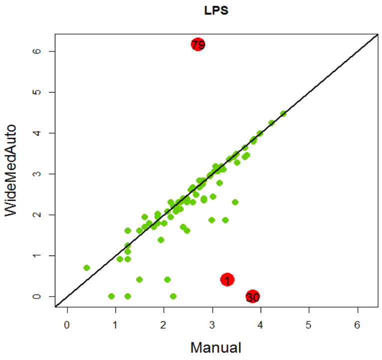
$$p(\boldsymbol{X_i}|\boldsymbol{\mu}, \Sigma, \lambda_i) \sim N(\boldsymbol{\mu}, \lambda_i \Sigma^{-1}) \tag{3}$$

$$p(\lambda_i|\nu) \sim \Gamma(\nu/2, \nu/2) \tag{4}$$

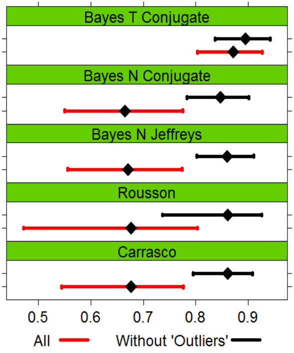
- Assign conjugate priors on parameters $\boldsymbol{\mu}$, Σ and uniform prior on ν
- Initial values are obtained by Expectation/Conditional Maximization Either (ECME) algorithm
- Use slice sampling [Neal, 2003] to generate samples from the posterior of ν

An EEG Example—With and Without "Outliers"

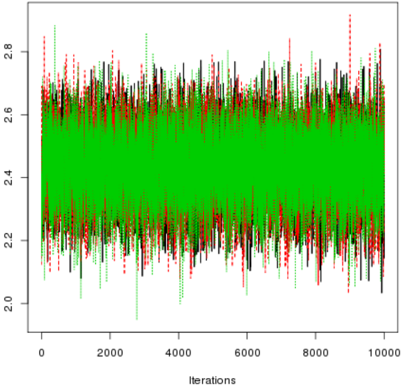
Scatter Plot



Point estimate and CI

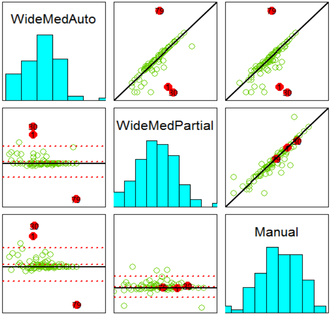


An EEG Example—Convergence Diagnostics

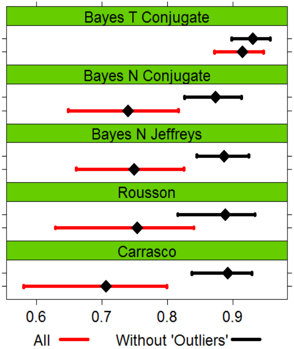


An EEG Example—Three Raters

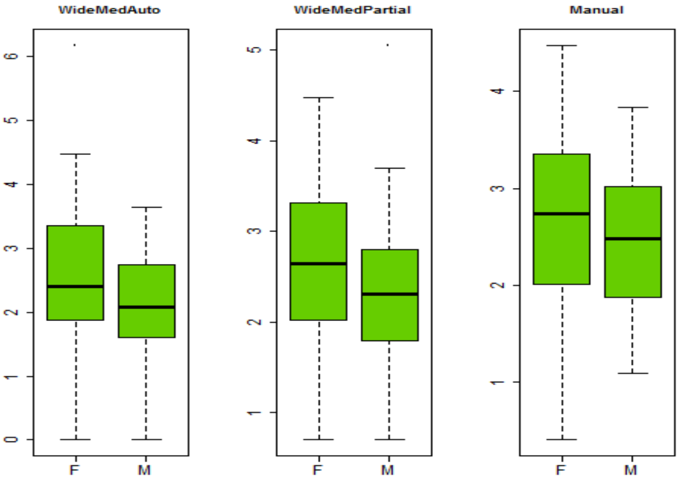
Data



Point estimate and CI



An EEG Example—Gender Effect



Bayesian Treatment of the Multivariate t Distribution–Adjustment of Covariates

- Assume a multivariate linear model with t distribution as follows

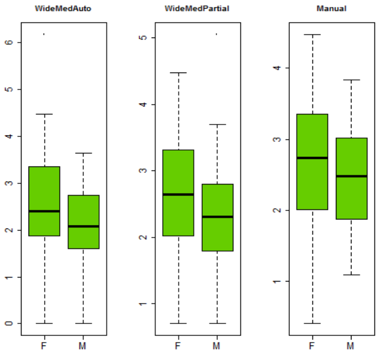
$$Y_i = X_i\beta + \epsilon_i,$$

where $\epsilon_i, i = 1, \ldots, n$, are i.i.d. $MVT(0, \Sigma, \nu)$.

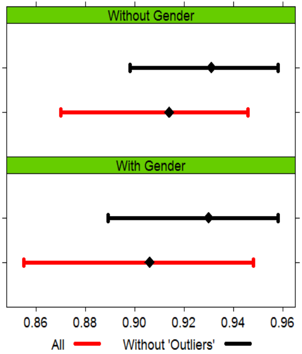
  - Assign non-informative priors on Σ and β, and uniform prior on ν.

An EEG example—Gender Effect (cont'd)

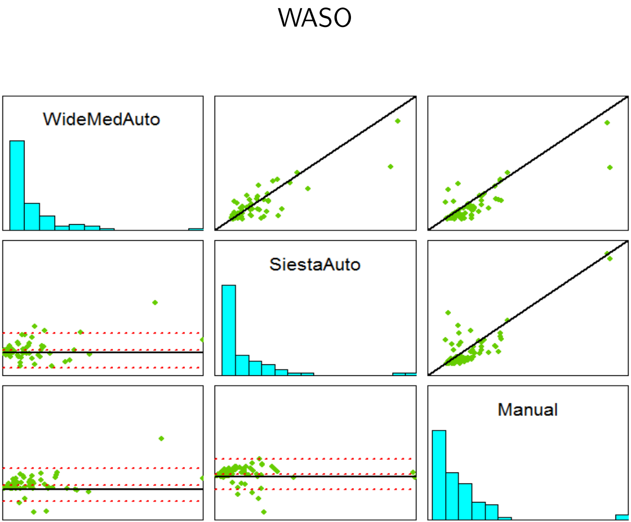
Box plot



Point estimate and CI



Another EEG Example



Bayesian Treatment of a Skewed Distribution

$$Y_{ij} = \beta_0 + \alpha_i + \beta_j + \epsilon_{ij}$$

- β_0 is the overall mean
- β_j is the fixed rater effect
- ϵ_{ij} is the random error and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$
- α_i is the random subject effect and $\log(\alpha_i) \sim N(\mu_\alpha, \sigma_\alpha^2)$

Bayesian Treatment of a Skewed Distribution (cont'd)

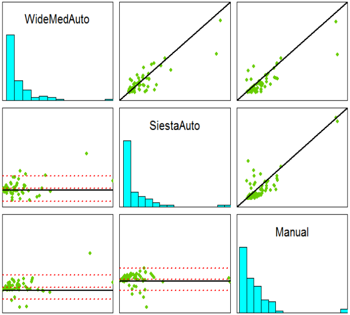
- Introduce α_i as latent variables [Van Dyk and Meng, 2001]
- Let

$$\begin{aligned}\log(\alpha_i)|\mu_\alpha, \sigma^2_\alpha &\sim N(\mu_\alpha, \sigma^2_\alpha) \\ Y_{ij}|\beta_0, \beta_j, \alpha_i, \sigma^2_\epsilon &\sim N(\beta_0 + \alpha_i + \beta_j, \sigma^2_\epsilon)\end{aligned}$$

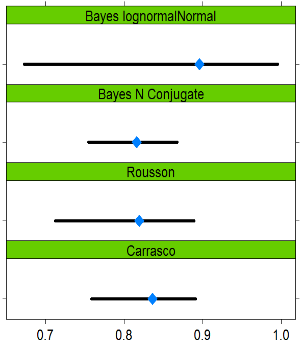
- Assign non-informative priors to $\beta_0, \beta_j, \mu_\alpha, \sigma^2_\alpha, \sigma^2_\epsilon$

Another EEG Example (cont'd)

Scatter Plot



Point estimate and CI



Bayesian Model Checking

Posterior predictive p-values: the probability that the replicated data could be more extreme than the observed data

$$\begin{aligned} p_B &= P(T(y^{rep},\theta) \leq T(y,\theta)|y) \\ &= \int\int I_{T(y^{rep},\theta)\leq T(y,\theta)} p(y^{rep}|\theta)p(\theta|y)dy^{rep}d\theta \end{aligned}$$

The posterior predictive p-value for maximum values of each measure is 0.51, 0.45, and 0.46, respectively, for WideMedAuto, SiestaAuto and Manual method.

An R Package

agRee: Various Methods for Measuring Agreement
`https://CRAN.R-project.org/package=agRee`

Potential Issues and Future Work

- A parametric distributional assumption underpins each model. The less appropriate the assumption, the worse the results.
  - The goodness-of-fit of each model can be checked and based on the DIC, different distributional assumptions can be compared to each other.
  - A more accurate estimate may be obtained through some non-parametric Bayesian approaches. Sample size?!
- Different choice of prior
  - Adoption of informative prior
  - Objective prior
- Computational issues

Summary

- The Bayesian approaches can provide very compelling results even from a frequentist point of view such as accurate coverage probabilities.
- Practically relevant issues, such as accommodation of covariates and model diagnostics and comparison, can all be addressed coherently within the Bayesian framework.
- Some issues need further investigation.

Thanks!

References I

Barnett, V. and Lewis, T. (1984).
Outliers in statistical data.
Wiley, 2nd edition.

Barnhart, H. X., Haber, M. J., and Lin, L. I. (2007).
An overview on assessing agreement with continuous measurements.
Journal of biopharmaceutical statistics, 17(4):529–569.

Berger, J. (1994).
An overview of robust bayesian analysis [with discussion]. test 3: 5-124.
Berger53Test1994.

References II

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (2011).
Robust statistics: the approach based on influence functions, volume 114.
John Wiley & Sons.

Hostetler, E. D., Joshi, A. D., Sanabria-Bohórquez, S., Fan, H., Zeng, Z., Purcell, M., Gantert, L., Riffel, K., Williams, M., O'Malley, S., et al. (2013).
In vivo quantification of calcitonin gene-related peptide receptor occupancy by telcagepant in rhesus monkey and human brain using the positron emission tomography tracer [11c] mk-4232.

References III

Journal Of Pharmacology and Experimental Therapeutics, 347(2):478–486.

Lange, K. L., Little, R. J., and Taylor, J. M. (1989).
Robust statistical modeling using the t distribution.
Journal of the American Statistical Association, 84(408):881–896.

Liu, C. (1994).
Statistical analysis using the multivariate t distribution.
PhD thesis, Harvard University.

Neal, R. M. (2003).
Slice sampling.
Annals of statistics, pages 705–741.

References IV

Quan, H. and Shih, W. J. (1996).
Assessing reproducibility by the within-subject coefficient of variation with random effects models.
Biometrics, 52(4):1195–1203.

Shrout, P. E. and Fleiss, J. L. (1979).
Intraclass correlations: Uses in assessing rater reliability.
Psychological Bulletin, 86(2):420–428.

Van Dyk, D. A. and Meng, X.-L. (2001).
The art of data augmentation.
Journal of Computational and Graphical Statistics, 10(1):1–50.