Assessment of measurement agreement for longitudinal/functional data

***Mina Yoo*[1], *Runze Li*[2]**

[1] *The Pennsylvania State University*; [2] *The Pennsylvania State University*

Summary

We consider an agreement index of multiple raters for densely measured longitudinal data. In order to accommodate a curvature feature in data and reduce ill-specified model error, nonparametric mean function is added in the linear mixed model. Based on the model, we proposed a time-varying agreement index, namely functional Concordance Correlation Coefficient(FCCC). The FCCC measures a degree of total-, inter-, intra-raters agreement at each measurement occasion. The indices are robust against a functional form of population characteristic and informative about the source of disagreements. FCCCs suggest directions on how to improve agreement among multiple raters based on a detail source of disagreement over time. We compare the performance of FCCCs with the Unified Concordance Correlation Coefficients for longitudinal/functional data.

Introduction

Agreement measures the "closeness" between readings, and contains both accuracy and precision. Assessment of agreement is essential in many fields. Of interest is to quantify agreement to decide

- whether or not new assay is acceptable with respect to the gold standard
- whether or not methodologies can be used interchangeably each other

Much work has been done on measurement agreement for independent data. For the repeated measurements, the Unified Concordance Correlation Coefficient(UCCC) is a popular agreement index for multiple raters.
The model for UCCC is a two way ANOVA :

$$y_{ijl} = \mu + a_i + \beta_j + \gamma_{ij} + \varepsilon_{ijl} \qquad (1)$$

where y_{ijl} is the l_{th} observation of subject i taken by rater j for $i = 1,\ldots,n,\ j = 1,\ldots,k$, $l = 1,\ldots,m$; μ is the overall mean; $a_i \sim \left(0, \sigma_a^2(t_l)\right)$; $\sum_{j=1}^{k} \beta_j = 0$; $\gamma_{ij} \sim \left(0, \sigma_\gamma\right)$; $\varepsilon_{ijk} \sim \left(0, \sigma_\varepsilon^2\right)$

Based on the model (1), Lin et al. (2007) proposed UCCC:

$$CCC_{total} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\beta^2 + \sigma_\varepsilon^2}, \qquad CCC_{intra} = \frac{\sigma_\alpha^2 + \sigma_\gamma^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\varepsilon^2}, \qquad CCC_{inter} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\gamma^2 + \sigma_\beta^2 + \frac{\sigma_\varepsilon^2}{m}}$$

where $\sigma_\beta^2 = \sum_{j=1}^{k-1} \sum_{j'=j+1}^{k} (\beta_j - \beta_{j'})^2 / k(k-1)$

- Each UCCC includes precision and accuracy.
- The assumptions in model (1), (1) a common set of occasions, (2) a balanced dataset, (3) discrete covariatese are difficult to satisfy in practice.
- Linear mixed model has flexibility in accommodating any degree of imbalance, but the impact of covariates are restricted in specified functions. Ill-specified model raises errors, and it leads to unreliable agreement indices.

Approach: we propose a FCCC for densely measured longitudinal data based on partially linear model with random covariates, robust against the number and timing of occasions and a functional form of time in data.

Model

For the sake of simplicity, balanced data is assumed. Suppose that $y_{ijk}(t_l)$ is the response of the i_{th} subject in j_{th} method and k_{th} possible other effect at time t_l; $i = i,\cdots,I,\ j = 1,\cdots,J,\ k = 1,\cdots,K,\ l = 1,\cdots,L$.

$$y_{ijk}(t_l) = \mu(t_l) + a_i + \alpha_j + \gamma_k + \varepsilon_{ijk}(t_l) \qquad (2)$$

where $\mu(t)$: smooth function; $\sum_{j=1}^{J} \alpha_j = 0$; $\sum_{k=1}^{K} \gamma_k = 0$; $a_i \sim \left(0, \sigma_a^2(t_l)\right)$; $\varepsilon_i(t) \sim \mathrm{GP}(0, R_i(t))$

The general expresion of model (2) is :

$$Y_i(t_l) = \mu(t_l) + X_i\beta + Zu_i + \varepsilon_i(t_l)$$

where $\mu(t)$ is smooth function; $\beta = \left[\alpha^T, \gamma^T\right]^T$; $u = [a_1,\cdots,a_I]^T \sim \left(0, \sigma_u^2 \times I_{I\times I}\right)$; $\varepsilon_i(t_l) \sim$GP(0, $R_i(t_l)$). $\varepsilon_{ijk}(t_l)$ and u_i are independent.

Estimation

Step 0 Set initial values of $\beta, u, \mu(\cdot)$

Step 1 Calculate $\hat r, \hat r^u$; $\quad \hat r(\cdot) = Y(\cdot) - \hat\mu(\cdot) - X^T\hat\beta,\ \ \hat r^u(\cdot) = Y(\cdot) - \hat\mu(\cdot) - X^T\hat\beta - Z^T\hat u$

Step 2 Estimate $\mathrm{var}(y_{ijk}(t_l)) \equiv V_{ijk}(t_l) = E(r_{ijk}^2(t)) = \sigma_\varepsilon^2(t) + \sigma_u^2,\ \ \forall i$ by following kernel estimator.

$$\widehat V_{ijk}(t_l) = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} \hat r_{ijkl}^2 K_{h_1}(t-t_l)}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} K_{h_1}(t-t_l)}$$

Because the variances only depends on time, $V_{ijk}(t_l)$ is the same for all subject at time t_l

step 3 An natural estimator for $\sigma_\varepsilon^2(t_l) = E\left(r_{ijk}^{u2}(t_l)\right)$ is the kernel estimator.

$$\hat\sigma_\varepsilon^2(t_l) = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} \hat r_{ijkl}^{u2} K_{h_2}(t-t_l)}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} K_{h_2}(t-t_l)}$$

step 4 Based on the fact, $V_{ijk}(t) = E(Var(y_{ijk}(t)|u)) + V(E(y_{ijk}(t)|u)) = \sigma_\varepsilon^2(t) + \sigma_u^2$

$$\hat\sigma_u^2 = \frac{1}{L}\sum_l^L\{\widehat V(t_l) - \hat R(t_l)\}, \quad \forall t = t_1,\cdots,t_L$$

step 5 Obtain $\hat\beta, \hat u, \hat\mu(t)$.

$$\begin{aligned} \hat\mu(\beta,u) &= \mathscr{S}(\hat r_i^u(t_l)) \\ \hat\beta &= \{(X^T(I-\mathscr{S})]\hat V^{-1}(I-\mathscr{S})X)\}^{-1}X^T(I-\mathscr{S})\hat V^{-1}(I-\mathscr{S})Y \\ \hat u &= \{(Z^T(I-\mathscr{S})\hat R^{-1}(I-\mathscr{S})Z + \hat G^{-1})\}^{-1}Z^T(I-\mathscr{S})\hat R^{-1}(I-\mathscr{S})\left(Y - X\hat\beta\right) \end{aligned}$$

where $\mathscr{S}$ stands for smoothing matrix.

Repeat step 2 through step 5 until all estimates converge.

Agreement indices

Agreements indexes for two effects, α_j, γ_k , are the same after switching corresponding variance of each effect. Here, agreement indices for γ at time t based on model (2) is following:

$$FCCC_{\text{total}}^{\gamma}(t_l) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2(t_l) + \sigma_\gamma^2}, \quad FCCC_{\text{intra}}^{\gamma}(t_l) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2(t_l)}, \quad FCCC_{\text{inter}}^{\gamma}(t_l) = \frac{\sigma_u^2}{\sigma_u^2 + \frac{1}{J}\sigma_\varepsilon^2(t_l) + \sigma_\gamma^2}$$

where $\sigma_\alpha^2 = \sum_{j=1}^{J-1}\sum_{j'=j+1}^{J} (\alpha_j - \alpha_{j'})^2_{J(J-1)}$ and $\sigma_\gamma^2 = \sum_{k=1}^{K-1}\sum_{k'=k+1}^{K} (\gamma_k - \gamma_{k'})^2_{K(K-1)}$

- Total-rater agreement: a measure of overall agreement for every occasion
- Intra-rataer agreement: how well is each rater in reproducing him/herself at each occasion
- Inter-rater agreement: how well are different raters in reproducing each other at each occasion

  ⇒ help to identify whether the disagreement is due to either intra-rater or inter-rater or both

  ⇒ provide directions on how to improve agreement among multiple methods
- Three indices decompose into precision and accuary
- Each index can be summarized into one numbers by the concept of reproducibility for functional data proposed by Li & Chow (2005).

$$\mathrm{FCCC}_{\text{total}}^{\gamma} = \int \mathrm{FCCC}_{\text{total}}^{\gamma}(t)w(t)dt \quad \text{given a weight function } w(\cdot)$$

.

Simulation Results

- $a_i \sim N(0,1)$, $\mu(t) = 3\times\sin(t/10)$, $\alpha_1 = 1$, $\alpha_2 = -1$, $\gamma_1 = 2$, $\gamma_2 = 1$, $\gamma_3 = -1$, $\gamma_4 = -2$, $\gamma_5 = 0.5$, $\gamma_6 = -0.5$ for $i = 1\cdots10$, $j = 1,2$, $k = 1,\cdots,6$, $l = 1,\cdots,100$.
- $\varepsilon_i(t) \sim \mathrm{GP}(0, R_i(t))$ where $\mathrm{cov}(\varepsilon_{t_l}, \varepsilon_{t_{l'}}) = \rho^{|t_l - t_{l'}|}$ for $\rho = 0, 0.2, 0.5, 0.7, 0.9$.

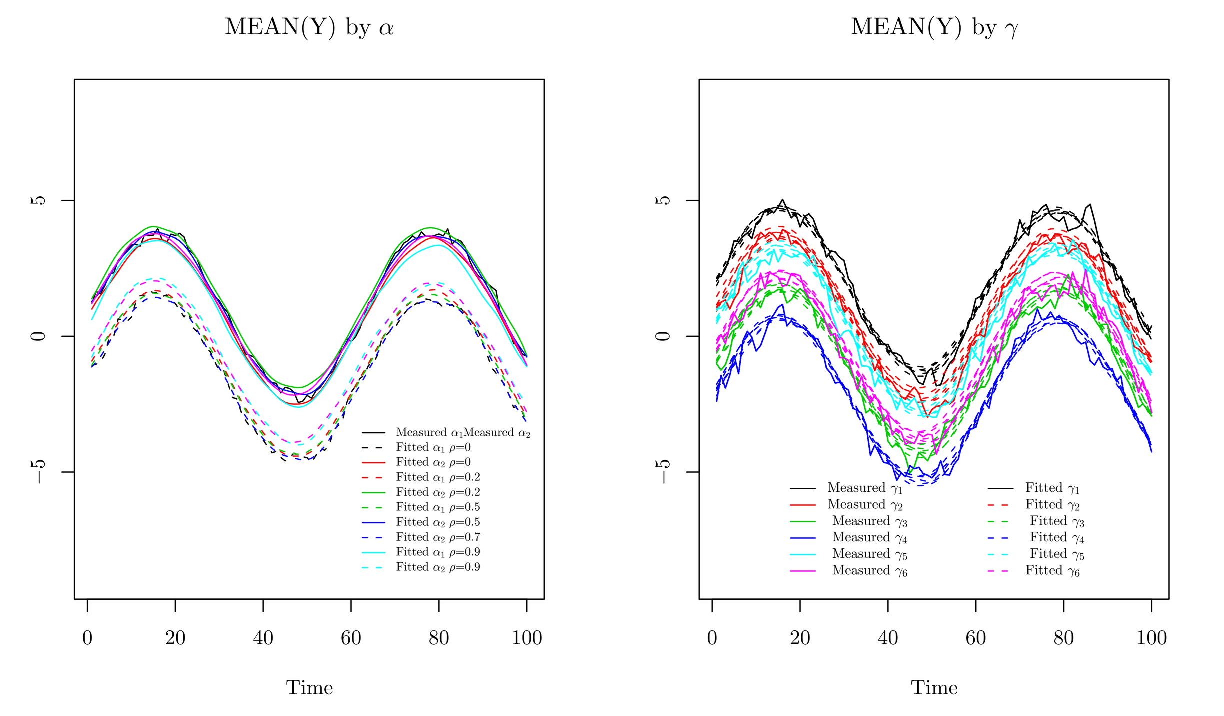

Figure 1: *Estimates.*
The left plot shows the average of estimates and true values belong to α_1 and α_2 by time for $\rho = 0, 0.2, 0.5, 0.7, 0.9$. The right plot depicts the average of estimates and true values corresponding to $\gamma_1, \cdots, \gamma_6$ over time by different correlations.

Table 1: *The results of Monte Carlo simulation; The first column is true FCCCs and others are estimates of FCCC and UCCC by ρ with equal weight.*

Mean (std)		$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$
Inter	UCCC	0.18 (0.12)	0.18 (0.12)	0.18 (0.12)	0.19 (0.13)	0.18 (0.13)
0.28	FCCC	0.23 (0.07)	0.23 (0.08)	0.23 (0.08)	0.23 (0.09)	0.23 (0.08)
Intra	UCCC	0.13 (0.08)	0.13 (0.10)	0.13 (0.10)	0.14 (0.11)	0.14 (0.10)
0.25	FCCC	0.22(0.10)	0.21 (0.10)	0.21 (0.10)	0.21 (0.11)	0.22 (0.11)
Total	UCCC	0.13 (0.10)	0.13 (0.10)	0.13 (0.10)	0.14 (0.11)	0.13 (0.10)
0.24	FCCC	0.20 (0.07)	0.20 (0.07)	0.20 (0.07)	0.20 (0.08)	0.22 (0.08)

FCCCs are close to true indeices while every UCCC underestimates the magnitude of agreements.

Conclusions

- FCCCs outperform UCCCs when there is non-linear time trend in population.
- FCCCs are a function of time which fits to the characteristic of longitudinal data.
- FCCCs are close to true values, but slightly underestimate. An optimal bandwidth proposed by Ruppert et al. (1995) was applied to estimate a nonparametric function via local linear smoother. Based on our another simulation result, the optimal bandwidth tends to oversmooth data.

  ⇒ The bias leads to smaller quantity of agreement.
- The model (2) explaines time trend in measurements via $\mu(\cdot)$ while the impact of the explanatory variables on responds are time invariant.

  ⇒ Functional mixed effects model can be applied to assess a measurement agreement index.

References

Li, R., & Chow, M. (2005). Evaluation of reproducibility for paired functional data. *Journal of Multivariate Analysis*, *93*(1), 81–101.

Lin, L., AS, H., & W., W. (2007). A Unified Approach for Assessing Agreement for Continuous and Categorical Data. *Journal of Biopharmaceutical Statistics*, *17*(4), 629–652.

Ruppert, D., Sheather, S., & Wand, M. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, (pp. 1257–1270).