Disentangling correlation between speed and ability at the subject level and between intensity and difficulty at the item level from psycholinguistic data: a joint modeling approach

Tom Loeys - Ghent University - Belgium
tom.loeys@ugent.be

1 Introduction

In psycholinguistic experiments multiple subjects are faced with multiple test items. Despite the early 70's paper of Clark (1973) arguing that averaging reaction times from such experiments over items for each subject and averaging over subjects for each item respectively, using these means in ANOVA-models (referred to as F1 and F2 statistics), and drawing inference from both statistics separately may be incorrect, the vast majority of published psycholinguistic results employed such techniques over the last decades. Baayen et al. (2008) explained in detail a mixed effects modeling approach with crossed random effects for subjects and items.
In addition to the reaction times, psycholinguistic literature is often describing accuracy as well. Accuracy is then summarized by simple frequency tables exploiting the binary outcomes (i.e. correct or incorrect response) measured for each subject-item combination. To improve on this and allow for estimation of covariates effects on the accuracy, Jaeger (2008) introduced in the psycholinguistic literature a model for the probability of a correct answer. More specifically, he proposed a mixed logistic regression model that allows for crossed random subject and item effects along the lines of Baayen et al. (2008).
Unfortunately, reaction times and accuracy are most often described separately without any concern being raised about their correlation. It is important to get a better understanding of the correlation between reaction times and accuracy, if any. The natural next step is therefore to consider a joint model for the reaction time and the accuracy. Joint modeling of these 2 outcomes can most easily be performed in a hierarchical framework. Van der Linden (2007) proposed an item-response theory model, a model for response time distribution and a higher-level structure accounting for the dependencies between the item and subjects parameters in these models. His hierachical framework is very flexible in that any item-response or response time model can be substituted.
Building on Van der Linden's work, we first provide a framework that combines the models introduced in the psycholinguistic literature by Baayen et al. (2008)

and Jaeger (2008), treats subjects and items as random, and allows for correlation between reaction time and accuracy. The main advantage of this framework is its ability to disentangle between correlation driven by subjects and correlation driven by items. Estimation of the model parameters in the joint model and model checking are performed in a Bayesian approach with Markov Chain Monte Carlo (MCMC). The performance of the proposed methodology is illustrated with a real-data example.

2 Methodology

We can typically consider two sources of information on a test in psycholinguistic experiments: (1) the reaction time resulting from the required time eliciting a response on a given trial, and (2) the response accuracy for each subject-item combination. Let subjects be indexed by $i = 1, \ldots, N$ and items by $j = 1, \ldots, K$. For subject i we measure a reaction time vector $T_i = (T_{i1}, \ldots, T_{iK})$ and a response vector $Y_i = (Y_{i1}, \ldots, Y_{iK})$. Further, let X_{1i} denote a subject specific characteristic and X_{2j} an item specific characteristic . For ease of notation, we will only consider a single subject and item covariate in the following, but models can straightforwardly be extended to have multiple covariates.

We start by introducing a model for the reaction time. Following Baayen et al. (2008) we assume that the reaction time T_{ij} for subject i on item j follows a mixed model with fixed effects α_0, α_1 and α_2 and random effects θ_{1i} and τ_{1j} for subject and item:

$$T_{ij} = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2j} + \theta_{1i} + \tau_{1j} + \epsilon_{ij} \tag{1}$$

with $\theta_{1i} \sim N(0, \sigma^2_{\theta 1})$, $\tau_{1j} \sim N(0, \sigma^2_{\tau 1})$ and $\epsilon_{ij} \sim N(0, \sigma^2)$.

Next we model the error rate. The probability that subject i answers item j incorrectly ($Y_{ij} = 1$) is assumed to follow a mixed effects logistic regression model with fixed effects β_0, β_1 and β_2 and random effects θ_{2i} and τ_{2j} for subject and item

$$\text{logit}(P(Y_{ij} = 1)) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2j} + \theta_{2i} + \tau_{2j} \tag{2}$$

with $\theta_{2i} \sim N(0, \sigma^2_{\theta 2})$ and $\tau_{2j} \sim N(0, \sigma^2_{\tau 2})$ subject and item deviation at the log odds ratio level respectively.

A joint modeling framework

Following the hierarchical framework introduced by Van der Linden (2007), we invoke in this presentation a joint modeling approach by a imposing a joint multivariate distribution on the vector of all random effects for subject and item. More specifically, we assume that the subjects parameters θ_{1i} and θ_{2i} follow a bivariate normal distribution with mean 0 and a covariance structure specified by

$$\Sigma_S = \begin{pmatrix} \sigma_{\theta 1}^2 & \rho_\theta \sigma_{\theta 1} \sigma_{\theta 2} \\ \rho_\theta \sigma_{\theta 1} \sigma_{\theta 2} & \sigma_{\theta 2}^2 \end{pmatrix} \tag{3}$$

Similarly the item parameters τ_{1i} and τ_{2i} are assumed to follow a bivariate normal distribution with mean 0 and a covariance structure specified by

$$\Sigma_I = \begin{pmatrix} \sigma_{\tau 1}^2 & \rho_\tau \sigma_{\tau 1} \sigma_{\tau 2} \\ \rho_\tau \sigma_{\tau 1} \sigma_{\tau 2} & \sigma_{\tau 2}^2 \end{pmatrix} \tag{4}$$

In (3) and (4), ρ_θ measures the correlation between speed and ability (or equivalently between slowness and disability) at the subject level and ρ_τ measures the corrrelation between time intensity and difficulty at the item level respectively.

Parameter Estimation

Following Van der Linden (2007) we will take a Bayesian approach with Markov Chain Monte Carlo (MCMC) computation. To keep the impact of the choice of the priors minimal, low informative priors are proposed here. Independent normal distributions with zero mean and large variances are used as priors for the fixed effect parameters α_0 through α_2 (or $\tilde{\alpha}_0$ through $\tilde{\alpha}_2$), and β_0 through β_2. The inverse Wishart distribution is used as the conjugate prior for the covariance matrix of a multivariate normal distributions, i.e.

$$\Sigma_S \sim \text{Inverse} - \text{Wishart}(\Sigma_{S0}^{-1}, \kappa_{S0}) \text{ and } \Sigma_I \sim \text{Inverse} - \text{Wishart}(\Sigma_{I0}^{-1}, \kappa_{I0}) \tag{5}$$

where κ_{S0} and κ_{I0} are scalar degrees-of-freedom parameters and Σ_{S0} and Σ_{I0} are 2×2 (positive definite symmetric) scale matrices. This prior provides information about the scale of the random effects. We choose $\kappa = 2$ because it is least informative, and

$$\Sigma^{-1} = \omega \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The off-diagonals equal to zero reflect the lack of knowledge on the correlation, while the equal diagonal elements reflect the lack of knowledge on the relative size of reaction time and accuracy variability. The prior is informative however through the value of ω, representing a scale factor on the variance. The larger ω, the more mass is concentrated on large values of variances. To explore the sensitivity of the results on the choice of ω, we analyzed our example with 3 different values of ω ($\omega = 0.2, 1, 5$).

A Gamma(η_1, η_2) prior can be assumed for the measurement precision $1/\sigma$ of the normal distributed reaction times, with η_1 and η_2 small for a low informative prior.

3 Empirical Example

The example from Baten et al. (2011) is embedded in a psycholinguistic research tradition dealing with the process of visual word recognition in bilinguals. Word targets (English-Dutch homographs and controls) were put in final position of low-constraint English sentences presented through serial visual presentation. The focus was to investigate whether or not the presence of a sentence context automatically lead to the activation of the appropriate meaning of the homograph? For instance, in "he was looking for something special and saw a new BRAND" is the English meaning of BRAND automatically activated or does it still compete with the less plausible Dutch meaning?
In the 2 experiments described below, Dutch-English bilinguals at two different levels respectively, i.e. one high proficiency group and one intermediate proficiency group, participated. Participants were asked to perform an L2-lexical decision task. 32 highly proficient and 31 intermediate proficient test takers responded to 29 test items. The proportions of incorrect responses were 12.9% and 25.5% for the high and intermediate proficiency samples respectively. For both experiments a mixed linear model for the reaction time (expressed in ms) was fitted as in (1) with fixed item effects for homographs, frequency, overlap and time quartile) and crossed random effects for subject and item, while for the probability of an incorrect answer a logistic regression model with the same factors as in (2) was used. The multivariate normal distributions (3) and (4) on the subject and item random effects were superimposed in the joint modeling framework.

The upper left panel of figure 1 shows a scatter plot of the estimated subject random intercepts for reaction time versus subject random intercepts for error rate (on the log odd ratio scale) in the high proficiency group, together with the estimated correlation from the joint modeling framework. A similar plot is provided for item random intercepts (right panel) in the high proficiency group in the upper right panel, while results for the intermediate proficiency group are presented in the lower panels. There is no evidence of a strong correlation between slowness and disability at the subject level in either the high or intermediate proficiency group. In contrast, a substantial positive correlation ($\rho_\tau = 0.88$) is estimated between time intensity and difficulty at the item side in the high proficiency group: items requiring larger reaction times tend to have larger error rates. As can be seen from the lower right panel of figure 2 the item correlation in the intermediate proficiency group is poorly estimated due to the apparent small item variability on the reaction time, and may be rather unreliable

While a range of checks on the reaction time distribution did not reveal major deviation from normality in our specific example, there is a growing literature on modeling of reaction times arguing that alternative distributional assumptions may often be required. As the focus of this presentation is on the correlation between reaction time and accuracy at the subject and at the item level, we assess in the following the robustness of our findings against such distributional misspecification on these correlations. More specifically, we assume next that the reaction time distribution follows a shifted Weibull distribution

$$f(t_{ij} \mid \psi_i, \lambda_{ij}, \gamma_i) = \lambda_{ij}\gamma(t_{ij} - \psi_i)^{\gamma_i - 1} \exp\left[-\lambda_{ij}(t_{ij} - \psi_i)^{\gamma_i}\right], \quad t_{ij} \geq \psi \tag{6}$$

with participant specific shifts $\psi_i \in \Re^+$ and shapes $\gamma_i \in \Re^+$, and a participant and item specfic rate parameter λ_{ij} (Rouder and colleagues, 2008).The following model can be assumed for the rate on the log scale:

$$\log \lambda_{ij} = -\tilde{\alpha}_0 - \tilde{\alpha}_1 x_{1i} - \tilde{\alpha}_2 x_{2j} - \tilde{\theta}_{1i} - \tilde{\tau}_{1j} \tag{7}$$

with $\tilde{\theta}_{1i} \sim N(0, \sigma^2_{\tilde{\theta}1})$ and $\tilde{\tau}_{1j} \sim N(0, \sigma^2_{\tilde{\tau}1})$, and $\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\theta}_{1i}, \tilde{\tau}_{1j} \in \Re$. The results presented in figure 2, mimicking figure 1 under the Weibull assumption, yield very similar correlations between speed and ability, and between time intensity and difficulty, and illustrate the robustness of our findings.

4 Discussion

In this presentation we proposed a relatively straightforward but important extension to the current modeling approaches for reaction time and accuracy from psycholinguistic experiments. Its major advantage is the ability to directly estimate the correlation between reaction time and error rate at the subject and item level. Several modeling assumptions are imposed though. The assumption of normal distribution for the reaction times may be violated for example, but as shown the Bayesian framework is very flexible in that (shifted) Weibull or other distributions can easily be used instead. The methodology presented offers the psycholinguist an additional tool to gain further insight into his/her data, albeit at an additional level of model complexity...

References:

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items. *Journal of Memory and Language, 59,* 390-412.

Baten,K., Hofman,F.,&, Loeys,T. (2011) Cross-lingual activation in bilingual sentence processing: the role of word class meaning. *Bilingualism: Language and Cognition*

Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335-359.

Jaeger,F. T. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59,* 434-446.

Rouder, J., Tuerlinckx, F., Speckman, P., Lu, J., & Gomez, P. (2008). A Hierarchical Approach for Fitting Curves to Response Time Measurements. *Psychonomic Bulletin & Review, 15 (6),* 1201-1208.

Van der Linden, W. (2007). A Hierarchical Framework for Speed and Accuracy on Test Items. *Psychometrika, 72 (3),* 287-308.

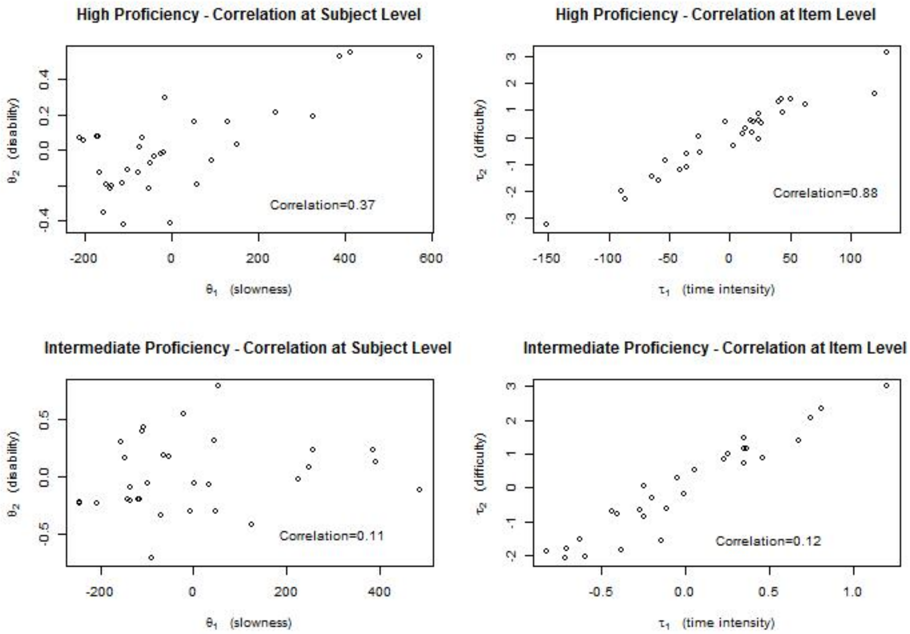

Figure 1: Correlation at Subject and Item Level under the assumption that reaction times follow a normal distribution

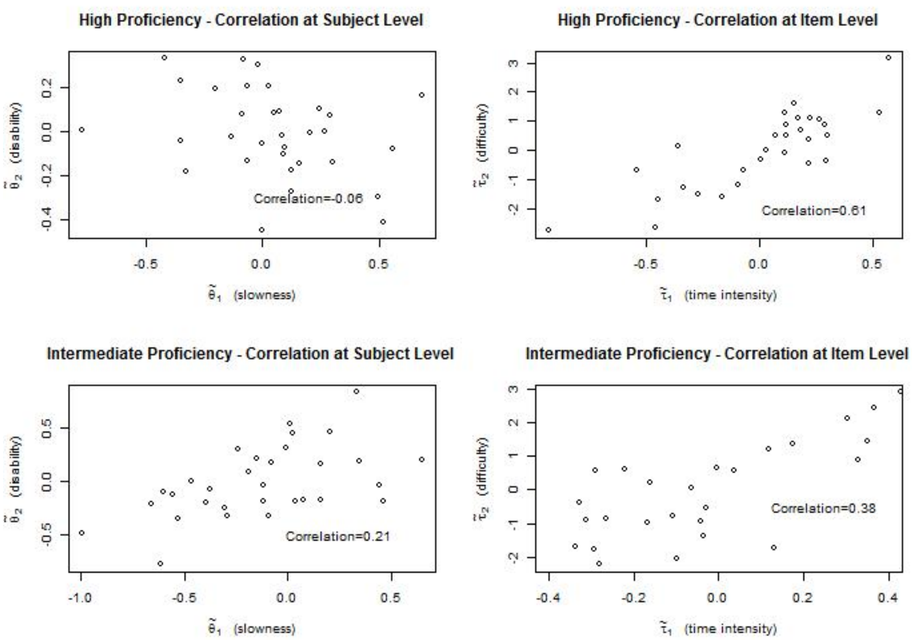

Figure 2: Correlation at Subject and Item Level under the assumption that reaction times follow a shifted Weibull distribution