# How hierarchical models improve point estimates of model parameters at the individual level

## Kentaro Katahira

Department of Psychology, Graduate School of Environmental Studies, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601, Japan

## HIGHLIGHTS

- We investigate how hierarchical models improve point estimates of subject-level parameters.
- One accuracy measure used is the correlation between the model parameter and the subject's trait variable.
- Another accuracy measure used is the root mean square error from the true parameter.
- For both measures, the conditions under which the hierarchical model is superior to other non-hierarchical methods are clarified.

## ARTICLE INFO

## ABSTRACT

Computational models have been used to analyze the data from behavioral experiments. One objective of the use of computational models is to estimate model parameters or internal variables for individual subjects from behavioral data. The estimates are often correlated with other variables that characterize subjects in order to investigate which computational processes are associated with specific personal or physiological traits. Although the accuracy of the estimates is important for these purposes, the parameter estimates obtained from individual subject data are often unreliable. To solve this problem, researchers have begun to use hierarchical modeling approaches to estimate parameters of computational models from multiple-subject data. It is widely accepted that the hierarchical model provides reliable estimates compared to other non-hierarchical approaches. However, how and under what conditions the hierarchical models provide better estimates than other approaches has yet to be systematically investigated. This study attempts to investigate these issues, focusing on two measures of estimation accuracy: the correlation between estimates of individual parameters and subject trait variables and the absolute measures of error (root mean squared error, RMSE) of the estimates. An analytical calculation based on a simple Gaussian model clarifies how the hierarchical model improves the point estimates of these two measures. We also performed simulation studies employing several realistic computational models based on the synthesized data to confirm that the theoretical properties hold in realistic situations.

## 1. Introduction

Computational models that provide trial-by-trial predictions of subjects' behavior have been recognized as a valuable tool for investigating underlying neural, cognitive, and psychological processes (Corrado & Doya, 2007; Daw, 2011; Lee & Wagenmakers, 2014; O'Doherty, Hampton, & Kim, 2007; Yechiam, Busemeyer, Stout, & Bechara, 2005). One purpose of using these computational models is to estimate latent variables underlying the computational processes. The estimates are often correlated with neural signals (e.g., from functional magnetic resonance imaging, fMRI) to find brain regions that represent the internal variables (O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; O'Doherty et al., 2004, 2007; Tanaka et al., 2004). Other applications use the estimates of model parameters to characterize individual subjects. In such applications, the parameter estimates of individual model parameters are correlated with trait variables of individual subjects[1]

[1] If the individual difference is the phasic state caused by an experimental manipulation, it should be called "state" variable rather than trait variable. However, we use "trait" throughout this paper for brevity.

(e.g., the characteristics of physiological or neural activities, personality trait, and the degree of mental disorders) to investigate which parameter is related to which personal trait (Huys et al., 2012; Katahira, Fujimura, Matsuda, Okanoya, & Okada, 2014; Kunisato et al., 2012; Lindström, Selbing, Molapour, & Olsson, 2014; Otto, Raio, Chiang, Phelps, & Daw, 2013; Sakamoto, Somatori, Okubo, & Kunisato, 2015; Suzuki et al., 2012; Yechiam et al., 2005). For both applications, obtaining accurate parameter estimates for individual subjects is an important first step.

A typical experiment entails data from multiple subjects. One method of addressing multiple-subject data in model-based analysis is to separately estimate different parameter sets for each subject's data using maximum likelihood estimation (MLE). This method is called single-subject MLE. Single-subject MLE is a straightforward approach for considering individual differences. However, such individual-level independent analysis often yields unreliable estimates, especially when the number of trials is limited. Recently, a hierarchical modeling approach that includes the group-level population distribution of model parameters has been used to estimate parameters at the individual level (Ahn, Krawitz, Kim, Busemeyer, & Brown, 2011; Ahn et al., 2014; Guitart-Masip et al., 2012; Huys et al., 2011, 2012; Suzuki, Adachi, Dunne, Bossaerts, & O'Doherty, 2015). A hierarchical model can incorporate group tendencies as well as individual differences (e.g., Ahn et al., 2011). It is widely recognized that the hierarchical model approach can provide reliable estimates compared to parameter estimates that are separately estimated for each subject (i.e., single-subject MLE).

Our primary focus in this paper is related to the second purpose of using computational models, i.e., estimating the correlation between the model parameter for individual subjects and the trait variables associated with the subjects. It is well known that a hierarchical model induces "shrinkage" of the individual parameter estimates towards the population mean (Efron & Morris, 1977). Thus, it appears to somehow distort the relative parameter estimates among individuals. Such a distortion could worsen the estimates of correlation between the parameter estimates and trait variables. Considering this, a natural question is how and under what conditions the hierarchical modeling approach can improve the estimates of the correlation. Although some studies have evaluated the correlations between the parameter estimates of HB and true parameters by using simulations (e.g., Farrell & Ludwig, 2008), this question has not yet been examined either systematically or analytically. We address this by combining theoretical analysis of a simple Gaussian model and systematic simulations based on several practical models.

In addition, we examine the absolute-estimation error of estimates for individual parameters. Whereas the absolute values of the parameter estimates do not affect the estimates of correlation given the relative values among subjects, if the model estimates are used to construct regressors for neural or physiological activity data, the absolute values of the estimates do matter. For example, the regressor obtained from different parameter values (e.g., temporal discount rate) is correlated with blood-oxygen-level dependent (BOLD) signals from different brain regions (Tanaka et al., 2004; see Wilson & Niv, 2015 for a discussion on robustness against mistuning of model parameters). For this purpose, single-parameter sets that maximize the likelihood for aggregated entire subject data have often been used to obtain stable regressors. This approach is called the fixed-effect MLE. How and under what conditions the fixed-effect MLE outperforms the single-subject MLE with respect to absolute errors in the estimates and what advantages are provided by hierarchical modeling for this purpose are investigated in this paper.

We discuss these issues for point estimates rather than for the entire posterior distribution of the model parameters. Practically, many methods in cognitive neuroscience and psychology still rely on point estimates, though full Bayesian analysis is becoming popular. The following reasons describe why point estimates are useful: first, when parameter estimates are used for fMRI analysis, regression is performed for each voxel. The target voxels can easily become greater than tens of thousands in number. Thus, incorporating the full posterior distribution of the model parameter is often infeasible. Although recent studies have attempted to combine the BOLD signal and behavioral data in a unified full Bayesian framework (Turner et al., 2013; Turner, Van Maanen, & Forstmann, 2015), these methods are currently not scalable to whole-brain, massive voxel data. Second, the point estimates are easily visualized using correlation plots. In addition, they can be easily submitted to traditional, mature analysis that many researchers are still familiar with, although the full Bayesian approach might be superior, in principle (Kruschke, 2013, 2014). Because we focus on the point estimates, we do not claim that the present study concerns the general properties of Bayesian inference in computational models. However, our theoretically derived results will also contribute to an understanding of the properties of hierarchical (full) Bayesian approaches. We also focus on estimates of parameters at the individual level, rather than those at the group level. In many situations in psychology, the population (the group level) distribution is the main concern, rather than the parameters of individual subjects. However, there are cases in which individual model parameters matter, as discussed above.

This paper is organized as follows. In Section 2, we formally describe the problem settings. We then analyze a simple Gaussian model as a model for individual subject responses in Section 3. This model is analytically tractable, and we gain several general insights. Next, in Section 4, we perform numerical simulations to confirm that the properties clarified by the analysis of the Gaussian model also hold in various models that have been influential in psychology and cognitive neuroscience. The models include psychophysical functions, signal detection theory (SDT), a response time (RT) distribution model, and a reinforcement learning (RL) model (Q-learning model). In Section 5, we discuss several implications of the results in terms of the fit of the computational model to multiple-subject data.

## 2. Problem formulation

Here, we formally describe the scenario that this paper considers. Suppose that we obtain the behavioral data of $N$ subjects. Each subject experiences a total of $T_i$ trials. The response (or action) at the $t$th trial of the $i$th subject is denoted by $x_{it}$. Computational models represent the internal computational processes of the subjects and predict the responses of the subjects. Specifically, the models predict how likely the subjects are to exhibit a response at trial $t$. This prediction is represented as $p(x_{it}; \theta_i)$. This is a probability density function if the response $x_{it}$ takes on continuous values but is a probability mass function if the response is defined as a discrete set. $\theta_i = (\theta_{i,1}, \ldots, \theta_{i,M})$ denotes the parameter set of the computational model for the $i$th subject, where the $m$th parameter for the $i$th subject is denoted by $\theta_{i,m}$. The response of the $i$th subject is denoted as $x_i = (x_{i1}, \ldots, x_{it})$, and all of the subjects' responses are denoted as $x = (x_1, \ldots, x_N)$.

The main task considered in this paper is estimating the parameter set $\theta_i$ from the given behavioral data. To focus on the accuracy of the parameter estimation, we assume that the fitted model includes the *true model* and can exactly represent the underlying computational processes with an appropriate parameter set. The issues of mis-specification of the model and model comparison are not addressed in this paper.