Cognitive State Prediction Using an EM Algorithm Applied to Gamma Distributed Data

Ali Yousefi[1,2], Angelique C Paulk[1], Thilo Deckersbach[3], Darin D Dougherty[3], Emad N Eskandar[1], Alik S. Widge[3,4,*], and Uri T. Eden[2,*]

[1] Department of Neurological Surgery, Massachusetts General Hospital and Harvard Medical School, Boston, MA; [2] Department of Mathematics and Statistics, Boston University, Boston, MA; [3] Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA; [4] Picower Institute for Learning & Memory, Massachusetts Institute of Technology, Cambridge, MA

Abstract— **Behavioral tests are widely used to quantify features of cognitive processing. For a large class of behavioral signals, the observed variables are non-Gaussian and dynamic; classical estimation algorithms are ill-suited to modeling such data. In this research, we propose a mathematical framework to predict a cognitive state variable related to behavioral signals, which are best modeled using a Gamma distribution. The proposed algorithm combines a Gamma Smoother and EM algorithm in the prediction process. The algorithm is applied to reaction time recorded from subjects performing a Multi-Source Interference Task (MSIT) to dynamically quantify their cognitive flexibility through the course of the experiment.**

I. Introduction

For a large class of behavioral data, observed variables are non-negative with an asymmetric distribution; moreover, their statistical measures change through the course of experiments. For example, many psychophysical tasks measure reaction time (RT) as a dependent variable that is influenced by trial parameters. The reaction times are positive random variables with right-skewed distributions. In repeated experiments, reaction times are often also dynamic; their statistics change as a subject's psychological or cognitive state evolves through an experiment. While many statistical analyses of reaction time data assume that the data are normally distributed, the actual structure of this class of behavioral signals can often be better described by a Gamma distribution. Additionally, state space methods have been used successfully to model dynamic features of neural and behavioral signals for a variety of cognitive tasks [1, 2, 3, 4]. In this work, we develop a mathematical algorithm to predict an underlying cognitive state variable using behavioral observations that are better fit by a Gamma model.

We have derived a complete state prediction algorithm for the class of linear state equations with Gamma distributed observations. The algorithm is based on maximum likelihood estimation using an approximate expectation-maximization (EM) algorithm; it sequentially estimates the model parameters and an unobserved cognitive state to maximize the likelihood of the observed data [5, 6].

* This research was supported by the Defense Advanced Research Projects Agency (DARPA) under contract number W911NF-14-2-0045. The opinions presented are those of the authors and not of DARPA or our institutions.

Address for reprint requests and other correspondence: Ali Yousefi, Department of Neurological Surgery, Massachusetts General Hospital and Harvard Medical School, Boston, MA (E-mail ayousefi@mgh.harvard.edu)

*: UTE and ASW contributed equally to this work.

For the state estimation step, a Gamma process filter analogous to the Kalman filter for Gaussian observations is derived. This state space framework with a Gamma observation model is well-suited to a large class of behavioral and neural data, and the prediction algorithm derives an adaptive filter solution for this class of data. We discuss the result of this new filter on RT data recorded during a multi-source interference task (MSIT) [7]. The algorithm is applied to the dataset to predict cognitive flexibility of subjects through the course of an experiment.

II. Dynamic Model of Gamma Distributed Data

We assume that the cognitive state evolves according to a simple linear state equation; the equation describes the temporal evolution of the cognitive state from one trial to the next through the course of an experiment [8, 2]. The observed reaction time data is modeled by a Gamma distribution, and its statistics are linked to a function of the cognitive state. The observed data is a sequence of continuous responses, and it is assumed that there are K trials for each experiment.

The cognitive state equation is modeled by a first order autoregressive process

$$x_{k+1} = a_1 * x_k + a_0 + \varepsilon_k \qquad (1)$$

where the (a_0, a_1) are model free parameters, the ε_k are independent zero mean Gaussian random variable with variance ${\sigma_\varepsilon}^2$. Parameter $0 \le a_1 < 1$ is the forgetting factor, and $a_0/(1-a_1)$ determines the steady-state value of the cognitive state.

The observed data is modeled by a Gamma distribution with an offset-term, the observation model is

$$z_k = y_k + \alpha \qquad \alpha \ge 0 \qquad (2.a)$$

$$f_{y_k|x_k}(y) \sim \frac{1}{\Gamma(v)} * \left(\frac{v*y}{\mu_k}\right)^v * \frac{1}{y} * e^{-\frac{v*y}{\mu_k}} \qquad (2.b)$$

$$\mu_k = e^{b_1 * x_k + b_0} \qquad (2.c)$$

where the z_k are the observed data, the α is the offset-term and the y_k are samples of the Gamma distribution [9]. The μ_k is the mean of Gamma distribution at k-th trial, and it is linked to the cognitive state by a log function. The log link function is a common link for the Gamma distributed data, but it can be replaced by any real-valued continuous function. The (b_0, b_1) parameters determine how the cognitive state influences the distribution mean, and $1/v$ is the distribution dispersion. The $f_{y_k|x_k}(\cdot)$ relates the pdf of the Gamma distribution to μ_k and v.

978-1-4244-9270-1/15/$31.00 ©2015 IEEE

The observed data and its underlying cognitive states are also dependent on the behavioral test features. The (b_0, b_1) parameters, plus (a_0, a_1), can be defined as a function of current and previous test features to reflect this dependence. For the MSIT task, we will describe the relationship between (b_0, b_1) parameters and the test features.

The state prediction objective is to maximize the likelihood of the observed data given the test features. Because x is unobservable, and $\theta = (a_0, a_1, {\sigma_\varepsilon}^2, b_0, b_1, \alpha, v)$ is a set of unknown parameters, we utilize the EM algorithm for the state prediction. The EM algorithm is a well-established technique to perform maximum likelihood estimation when there is an unobserved process [5]. Using the EM algorithm, the x process and θ parameter will be estimated to maximize the observed data likelihood. The EM algorithm has been used to estimate state-space models with point-process or binary observations; this new EM algorithm allows for state-space model estimation on Gamma observations [1, 2, 10]. The algorithm can be extended to estimate state-space models with both binary and Gamma-distributed observation data.

III. STATE PREDICTION ALGORITHM

a. EM Algorithm, E-step

The EM algorithm computes the maximum likelihood estimates of θ by maximizing the expectation of the complete data log-likelihood. The complete likelihood is the joint probability distribution of $X = \{x_1, \cdots, x_K\}$ and $Y = \{y_1, \cdots, y_K\}$ given θ, which for the proposed model is

$$p(X, Y|\theta, x_0) = f_{x_0}(x_0) \times \prod_{k=1}^{K} (2\pi{\sigma_\varepsilon}^2)^{-\frac{1}{2}} * \exp\{-(2{\sigma_\varepsilon}^2)^{-1}(x_k - a_0 - a_1 * x_{k-1})^2\} \times \prod_{k=1}^{K} \frac{1}{\Gamma(v)} * \left\{\left(\frac{v * y_k}{\mu_k}\right)^v * \frac{1}{y_k} * e^{-\frac{v * y_k}{\mu_k}}\right\}_{\mu_k = e^{b_1 * x_k + b_0}} \tag{3}$$

The first term on the right side of Eq. 3 defines the probability density of x_0, the initial value of state-space variable. Here, we assume that x_0 is a known value, though this can be replaced by any distribution and estimated as a part of EM procedure. The second term is defined by a Gaussian probability distribution for the state model in equation (1), and the third term is the Gamma distribution defined in equation (2.c). For iteration $(l+1)$ of the E-step, we compute the expectation of the complete data log likelihood given the Y and the parameter estimate from iteration l, $\theta^{(l)} = \left({a_0}^{(l)}, {a_1}^{(l)}, \sigma_\varepsilon^{2(l)}, {b_0}^{(l)}, {b_1}^{(l)}, \alpha^{(l)}, v^{(l)}, {x_0}^{(l)}\right)$, which is defined as

$$E\left[\log(p(X, Y|\theta)) \| Y, \theta^{(l)}\right] = E\left[K\left(\log\frac{v^v}{\Gamma(v)}\right) + (v-1)\sum_{k=1}^{K}\log y_k - v\sum_{k=1}^{K}\left(\log\mu_k + \frac{y_k}{\mu_k}\right) \Big\| Y, \theta^{(l)}\right] + E\left[-\frac{K}{2}\log(2\pi{\sigma_\varepsilon}^2) - (2{\sigma_\varepsilon}^2)^{-1}\sum_{k=1}^{K}(x_k - a_0 - a_1 x_{k-1})^2 \| Y, \theta^{(l)}\right] \tag{4}$$

To evaluate the E-step, we have to calculate the following terms for $k = 1, \cdots, K$

$$x_{k|K} = E\left[x_k \| Y, \theta^{(l)}\right] \tag{5.a}$$

$$W_{k|K} = E\left[x_k^2 \| Y, \theta^{(l)}\right] \tag{5.b}$$

$$W_{k,k-1|K} = E\left[x_k x_{k-1} \| Y, \theta^{(l)}\right] \tag{5.c}$$

$$M_{k|K}(t) = E\left(\exp(t x_k) \| Y, \theta^{(l)}\right) \quad t \in R \tag{5.d}$$

The notation $k|K$ denotes the fixed interval smoothing, it is the expectation of state variable, or a function of the state variable, at k given the whole observation up to step K. To compute these expectations, we follow the strategy first proposed by Shumway and Stoffer [6] for linear Gaussian state and observation models. The strategy is decomposed into four steps: a recursive filtering algorithm to compute $x_{k|k}$, a fixed interval smoothing algorithm to calculate $x_{k|K}$ and $\sigma_{k|K}^2$, a state-space covariance algorithm to estimate $W_{k|K}$, $W_{k,k-1|K}$, and a moment generating function to compute $M_{k|K}(t)$.

b. Filter Algorithm

Given $\theta^{(l)}$, the one-step prediction of mean and variance of the state space variable is

$$x_{k|k-1} = a_1^{(l)} x_{k-1|k-1} + a_0^{(l)} \tag{6.a}$$

$$\sigma_{k|k-1}^2 = a_1^{(l)2} \sigma_{k-1|k-1}^2 + \sigma_\varepsilon^{2(l)} \tag{6.b}$$

The definitions for the mean and variance of the state space variable are sufficient to compute the Gaussian one-step prediction probability density. To compute the $x_{k|k}$, we write the posterior probability density of the state variable given the observed data and $\theta^{(l)}$, and we use a Gaussian approximation.

$$p\left(x_k|\theta^{(l)}, y_{1\cdots k}\right) \propto p\left(y_k|\theta^{(l)}, x_k\right) \times p\left(x_k|\theta^{(l)}, y_{1\cdots k-1}\right) \propto \frac{1}{\Gamma(v^{(l)})}\left(\frac{v^{(l)} * y_k}{{\mu_{k|k-1}}^{(l)}}\right)^{v^{(l)}} \frac{1}{y_k} e^{-\frac{v^{(l)} y_k}{{\mu_{k|k-1}}^{(l)}}} \times \left(2\pi\sigma_{k|k-1}^2\right)^{-\frac{1}{2}} \exp\left\{-\left(2\sigma_{k|k-1}^2\right)^{-1}\left(x_k - x_{k|k-1}\right)^2\right\} \propto \left(2\pi\sigma_{k|k}^2\right)^{-\frac{1}{2}} \exp\left\{-\left(2\sigma_{k|k}^2\right)^{-1}\left(x_k - x_{k|k}\right)^2\right\} \tag{7}$$

The last line in equation 7 is the Gaussian approximation of the posterior probability with parameters $x_{k|k}$ and $\sigma_{k|k}^2$ still to be determined. For a Gaussian distribution, the maximum of the pdf, and similarly the log of pdf, happens at its mean value. The mean of the Gaussian approximation is the point defined as

$$\left.\frac{\delta \log p\left(x_k|\theta^{(l)}, y_{1\cdots k}\right)}{\delta x_k}\right|_{x_{k|k}} = 0 \tag{8.a}$$

$$x_{k|k} = x_{k|k-1} + \sigma_{k|k-1}^2 * v * \left[-\frac{\partial \log \mu_k}{\partial x_k}\left(1 - \frac{y_k}{\mu_k}\right)\right]_{x_{k|k}} \tag{8.b}$$

The variance term is defined by the second derivative of the distribution log, which is

$$\left.\frac{\delta^2 \log p\left(x_k \middle| \theta^{(l)}, y_{1\cdots k}\right)}{\delta x_k{}^2}\right|_{x_{k|k}} = -\sigma^2_{k|k}{}^{-1} \quad (9.a)$$

$$\sigma^2_{k|k}{}^{-1} = \sigma^2_{k|k-1}{}^{-1} + v\left.\left(\left(1-\frac{y_k}{\mu_k}\right)\frac{\partial^2 \log\mu_k}{\partial x_k{}^2} + \left(\frac{\partial \log\mu_k}{\partial x_k}\right)^2 \frac{y_k}{\mu_k}\right)\right|_{x_{k|k}} \quad (9.b)$$

The $x_{k|k}$ computation normally requires multiple iterations of equation 8. However, this can be avoided with the Gaussian approximation technique used in Eden et al. [10]. The technique starts by taking the log of the posterior density and taking its derivative with respect to x_k.

$$\sigma^2_{k|k}{}^{-1}\left(x_k - x_{k|k}\right) \propto \sigma^2_{k|k-1}{}^{-1}\left(x_k - x_{k|k-1}\right) + \frac{\partial \log\mu_k}{\partial x_k} v\left(1-\frac{y_k}{\mu_k}\right) \quad (10)$$

Assuming the Gaussian estimate is valid, the relationship should be approximately true for all the values of x_k. Evaluating at $x_k = x_{k|k}$ gives

$$x_{k|k} = x_{k|k-1} + \sigma^2_{k|k} * v * \left[-\frac{\partial \log\mu_k}{\partial x_k} * \left(1-\frac{y_k}{\mu_k}\right)\right]_{x_{k|k-1}} \quad (11)$$

Differentiating equation 10 again, and replacing $x_k = x_{k|k-1}$ gives the posterior variance equation.

$$\sigma^2_{k|k}{}^{-1} = \sigma^2_{k|k-1}{}^{-1} + v\left.\left(\left(1-\frac{y_k}{\mu_k}\right)\frac{\partial^2 \log\mu_k}{\partial x_k{}^2} + \left(\frac{\partial \log\mu_k}{\partial x_k}\right)^2 \frac{y_k}{\mu_k}\right)\right|_{x_{k|k-1}} \quad (12)$$

c. Fixed Interval Smoother

Given the sequence of posterior one-step estimates, $\left(x_{k|k}, \sigma^2_{k|k}\right)$, we use the fixed-interval smoothing algorithm to compute $x_{k|K}$ and $\sigma^2_{k|K}$ [11]. The smoothing algorithm is

$$x_{k|K} = x_{k|k} + A_k\left(x_{k+1|K} - x_{k+1|k}\right) \quad (13.a)$$

$$A_k = \sigma^2_{k|k} * a_1^{(l)} * \sigma^2_{k+1|k}{}^{-1} \quad (13.b)$$

$$\sigma^2_{k|K} = \sigma^2_{k|k} + A_k^2 * \left(\sigma^2_{k+1|K} - \sigma^2_{k+1|k}\right) \quad (13.c)$$

for $k = K-1, \cdots, 1$ and initial conditions $x_{K|K}$ and $\sigma^2_{K|K}$.

d. State-space Covariance and Moment Generating Function

Given the Gaussian assumption for the posterior distribution of the hidden state, the expectation $M_{k|K}(t)$ for an arbitrary t is

$$M_x(t) = \exp\left(t * x_{k|K} + \tfrac{1}{2} * t^2 * \sigma^2_{k|K}\right) \quad (14)$$

which is the MGF for a normal distribution with mean $x_{k|K}$ and variance $W_{k|K}$. The covariance term in equation (5.c) is given by [8, 11]

$$\sigma_{k-1,k|K} = A_{k-1} * \sigma^2_{k|K} \quad (15)$$

The variance and covariance terms for the E-step will be

$$W_{k|K} = \sigma^2_{k|K} + x^2_{k|K} \quad (16.a)$$

$$W_{k-1,k|K} = \sigma^2_{k|K} + x_{k-1|K} * x_{k|K} \quad (16.b)$$

e. EM Algorithm, M-step

The model parameter θ is updated to maximize the observed maximum likelihood. The update rule for the observation model parameters, (b_0, b_1, α, v), is

$$\frac{\partial Q}{\partial v} \equiv K\left(\log v + 1 - \frac{\Gamma'(v)}{\Gamma(v)}\right) + \sum_{k=1}^{K} \log(y_k) - \mathrm{E}(\log\mu_k) - y_k \mathrm{E}(1/\mu_k) = 0 \quad (17.a)$$

$$\frac{\partial Q}{\partial b_0} \equiv \sum_{k=1}^{K} \frac{\partial \mathrm{E}(\log\mu_k)}{\partial b_0} + \sum_{k=1}^{K} y_k * \frac{\partial \mathrm{E}(1/\mu_k)}{\partial b_0} = 0 \quad (17.b)$$

$$\frac{\partial Q}{\partial b_1} \equiv \sum_{k=1}^{K} \frac{\partial \mathrm{E}(\log\mu_k)}{\partial b_1} + \sum_{k=1}^{K} y_k * \frac{\partial \mathrm{E}(1/\mu_k)}{\partial b_1} = 0 \quad (17.c)$$

$$\frac{\partial Q}{\partial \alpha} \equiv (1-v)\sum_{k=1}^{K} \frac{1}{y_k} + v\sum_{k=1}^{K} \mathrm{E}(1/\mu_k) = 0 \quad (17.d)$$

Equation set 17 can be solved numerically to find the new set of observation model parameters. The expectation terms of equation set 17, $\mathrm{E}(\log\mu_k)$ and $\mathrm{E}(1/\mu_k)$, are defined by

$$\mathrm{E}(\log\mu_k) = \mathrm{E}(b_1 * x_k + b_0) = b_1 * x_{k|K} + b_0 \quad (18.a)$$

$$\mathrm{E}(1/\mu_k) = \exp(-b_0)\mathrm{E}\left(\exp(-b_1 x_k)\right) \quad (18.b)$$

To compute (18.b), we use the MGF function already defined in equation (14).

The update rule for the state equation parameters, $(a_0, a_1, \sigma^2_\varepsilon, x_0)$, is

$$\begin{bmatrix} a_0^{(l+1)} \\ a_1^{(l+1)} \end{bmatrix} = \begin{bmatrix} K & \sum_{k=1}^{K} x_{k-1|K} \\ \sum_{k=1}^{K} x_{k-1|K} & \sum_{k=1}^{K} W_{k-1|K} \end{bmatrix}^{-1} * \begin{bmatrix} \sum_{k=1}^{K} x_{k|K} \\ \sum_{k=1}^{K} W_{k,k-1|K} \end{bmatrix} \quad (19.a)$$

$$\sigma_\varepsilon^{(l+1)2} = \frac{1}{K}\sum_{k=1}^{K}\left[W_{k|K} + a_0^{(l+1)2} + a_1^{(l+1)2} W_{k-1|K} \cdots - 2a_0^{(l+1)} x_{k|K} - 2a_1^{(l+1)} W_{k,k-1|K} + 2a_1^{(l+1)} a_0^{(l+1)} x_{k-1|K}\right] \quad (19.b)$$

$$x_0^{(l+1)} = \left(x_{1|K} - a_0^{(l+1)}\right)/a_1^{(l+1)} \quad (19.c)$$

the (a_0, a_1) parameters are updated first, and then $(\sigma^2_\varepsilon, x_0)$ parameters will be updated.

The complete state prediction algorithm is defined through equations 1 to 19. Table 1 describes the algorithm implementation steps. The algorithm terminates whenever the log-likelihood increment falls below a preset positive threshold, ε, $\log p\left(X, Y \middle| \theta^{(l)}\right) \leq \log p\left(X, Y \middle| \theta^{(l-1)}\right) + \varepsilon$.

Table 1 State Prediction Algorithm Implementation

1.	*Initialize model parameters:* $\theta^{(0)}$
2.	*State-space filtering: Equations (6.a), (6.b), (11), and (12)*
3.	*State-space smoothing: Equation set (13)*
4.	*Covariance and MGF computation: Equations (14), and (16)*
5.	*Parameter update: Equation set (17), (18), and (19)*
6.	*Iterate steps 2 to 5 till convergence*

IV. Multi-Source Interference Task

The MSIT, described in [7], is designed to evoke a high level of cognitive conflict. It is a successor to a long line of cognitive conflict tasks, originating with the Stroop color-word conflict task [12]. MSIT combines this with flanker and motor mapping effects to increase the cognitive load. Trials are classified as "non-interference" (easy, low load) or "interference" (hard, high load). In general, subjects rapidly acquire the task concept and are able to perform at nearly 100% accuracy [7]. The key behavioral readout is their RT in response to a sequence of trials. First, on interference trials, patients respond 200-400 milliseconds more slowly on average than on non-interference trials [7]. Second, when the effect of (non)interference is controlled for, an effect of trial sequencing emerges. If an interference trial is followed by another interference trial (I2I), the interference effect lessens on the second trial – the brain adapts. Conversely, if an interference trial follows a non-interference trial (N2I), the effect of interference is much larger, because the brain acquires an easy-trial "response set" that is disrupted by the switch. This is often referred to as the "Gratton effect" [13]. The exact nature of the effect is debated, with some authors arguing it to be primarily driven by stimulus/response switches [14] and others attributing it to competing mental processes [15]. It nevertheless does occur, and interestingly, is dependent on the cingulate cortex – surgical ablation of the cingulate also eliminates the Gratton effect specifically in MSIT [16]. Thus, we can use the effect of trial-type switches (after trial type itself is controlled for) as a marker for a subject's moment-to-moment cognitive flexibility.

For the MSIT experiment, we model the RT, z_k in equation 2, by

$$b_0 = c_0 + c_1 I_i + c_2 Z_{k-1} \qquad (20.a)$$

$$b_1 = I_{n2i} + c_3 I_{i2n} \qquad (20.b)$$

where (c_0, c_1, c_2, c_3) are the model free parameters, and I_i is an indicator function for the interference trials. The I_{n2i} is the indicator function for the non-interference to interference trials, and the I_{i2n} is the indicator function for the interference to non-interference trials. The state-space variable (x_k) represents the influence of trial switching on RT, and it can be thought of as a measure of cognitive flexibility/rigidity.

V. State prediction in a sample Experiment

Figure 1 shows a sample MSIT analysis on a human subject, plus the cognitive state prediction result. This analysis is performed to demonstrate application of the proposed algorithm on a sample data. Future work will be focusing on a larger MSIT dataset which is comprised of more than 100 subjects. Data was collected under a protocol approved by the Massachusetts General Hospital Institutional Review Board.

The initial values of the model parameters, (c_0, c_1, c_2, c_3, v) are computed using a GLM regression algorithm, assuming $x_k = 1$ and $\alpha = 0$ [9]. For the state equation, a_1 is set to 1 and a_0 is set to zero. The σ_ε^2, the noise term in the state equation, is estimated by the EM algorithm.

Figure 1.a shows the distribution of RT for different sequence of trials. From the histogram we see that the RT interference effect is about 250 milliseconds, and the distribution is positively-skewed. Despite a similar RT distribution in non-interference-to-interference and interference-to-interference trials, their RTs change significantly through the experiment.

Figure 1.b shows the auto-correlation estimate for the RT. The graph result suggests the RTs are un-correlated or weakly correlated, when not accounting for interference effect.

Figure 1.c shows the GLM regression estimate of the coefficients. The GLM regression algorithm estimates a positive coefficient for Z_{k-1}, meaning that Z_{k-1} and Z_k are positively correlated when accounting for interference and switching effects. The auto-correlation analysis fails to extract existing correlation between Z_k and Z_{k-l} $l > 0$. This may reflect a model mis-specification, but may also imply that the dependence between Z_ks is unobservable if we exclude (I_i, I_{i2n}, I_{n2i}) terms. The I_i term defines the interference effect; the GLM estimate for the I_i coefficient is 0.25 meaning that interference trials take approximately 28% (μ_k definition in equation (2.c), $\exp(0.25) \approx 1.28$) longer than non-interference trials ignoring other terms of the model. The GLM estimate for the I_{n2i} coefficient is a positive value, supporting the possibility of observing an overall Gratton effect on non-interference to interference trials. The GLM estimate for the I_{i2n} coefficient is close to zero, suggesting non or minuscule Gratton effect on I_{i2n} trials of the experiment. For the dynamic model defined in equation (20), the I_{n2i} coefficient is equal to 1; the x_0 (initial value of the x_k) is asuumed to be equal to the I_{n2i} coefficient derived by the GLM regression estimate. The c_3 in equation (20.b) is the GLM estimate of the I_{i2n} coefficient normalized by the I_{n2i} coefficient estimate. Equations (2.c) and (20) define the relationship between the mean of RT and x_k, cognitive flexibility/rigidity. The mean is an increasing function of x_k; it means that the RT increment on the switching trials is linked to higher values of x_k, which is interpreted as cognitive rigidity. Conversely, the cognitive flexibility can be interpreted as the RT reduction, which corresponds to smaller or negative x_ks.

Figure 1.d shows the σ_ε^2 estimate through a successive EM algorithm. The EM estimate of σ_ε^2 significantly drops from its initial guess and converges to 0.004. This low variance estimate matches with a narrower confidence interval, or simply a more precise estimate, of $x_{k|K}$. The EM estimate can be applied to predict any other free parameters of the RT model.

Figure 1.e shows $x_{k|K}$, the cognitive flexibility/rigidity prediction. The $x_{k|K}$ graph shows how RT on switching trials changes in the time course of the experiment. There is about a 15% increment in the RT of a trial switch in the midst of the experiment, it then drops about 15% and increases again at the end of the experiment.

Figure 1.f shows observed RT plus its prediction with and without the cognitive state (x_k). The result supports the hypothesis of a dynamic cognitive flexibility/rigidity state. The modeling procedure presented here is able to estimate this dynamic slow process and leads to an improved RT prediction.

VI. Conclusion

We proposed an EM algorithm to predict state variables of a state-space model with a Gamma distributed observation. The algorithm can be applied to a large class of behavioral signals, for which the observed variables are dynamic with a Gamma distribution. We have applied this to the problem of estimating a time-varying cognitive flexibility process on data derived from MSIT. This early result demonstrates the proposed method's potential to quantify cognitive dynamics through the course of the experiment. This is a powerful general framework for capturing moment-to-moment variation in human behavior. Its most immediate use is in better quantifying subjects' behavior during psychophysical experiments that involve reaction time, much as state-space models have done for learning [2]. Further, this modeling technique has clinical implications. Because it is sensitive to immediate changes in behavior, it could quantify sharp changes in subjects' mental state, e.g. in response to a brain stimulation intervention. Finally, because the extracted state is defined at all times, it can be a substrate for neural decoding analyses similar to those commonly used in brain-computer interface applications. Thus, this single algorithm has uses throughout the basic and clinical neurosciences.

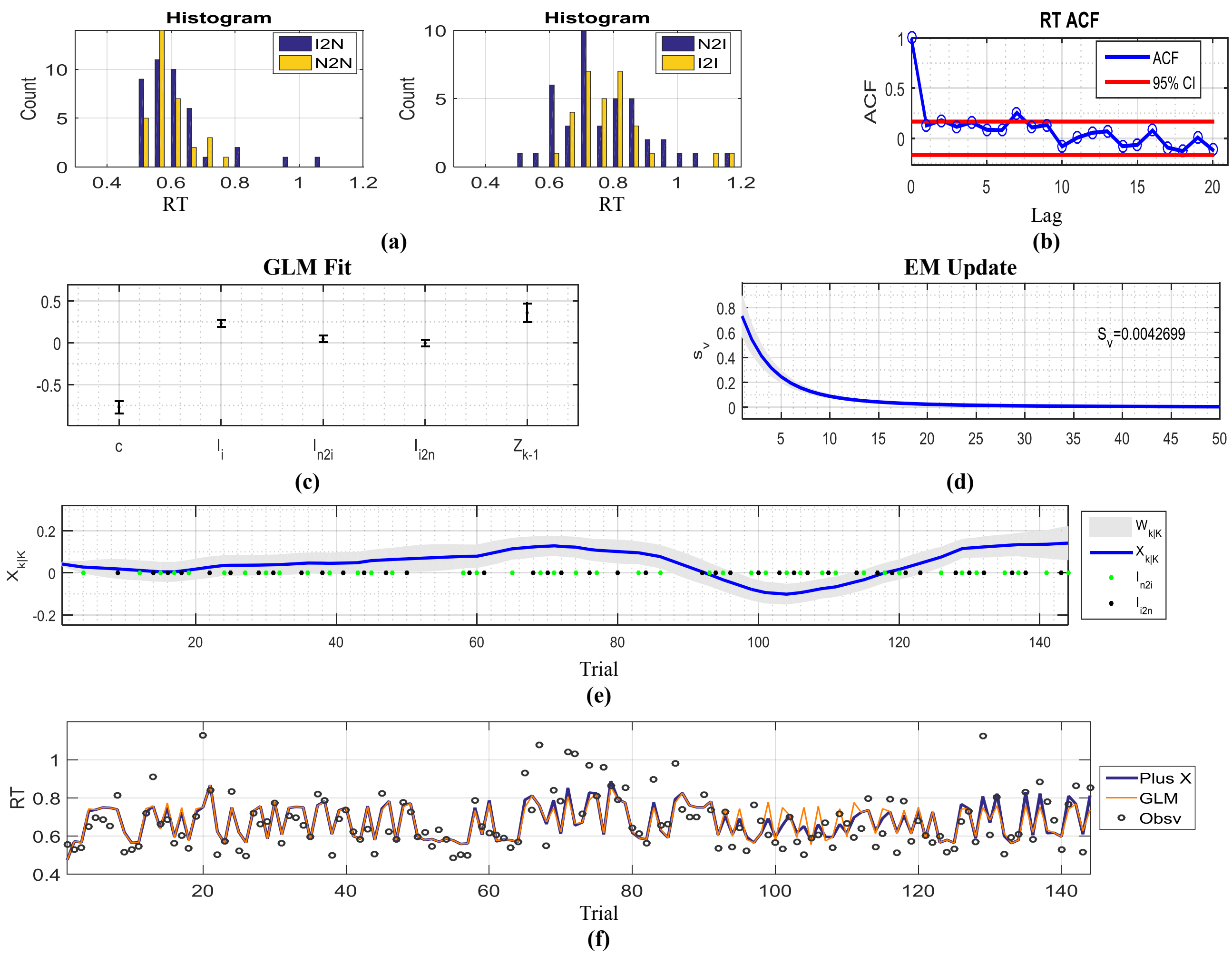

Figure 1 Sample MSIT data and cognitive state prediction. **a.** RT histogram for different sequence of trials – *I* means an interference trial and N means a non-interference trial, thus *I2N* signifies an N trial that follows an *I* trial. **b.** The Auto-Correlation Function (ACF) graph shows the auto-correlation estimate for RT with its confidence bound. The result shows that RTs are generally un-correlated. **c.** GLM regression estimate of coefficients. The I_i coefficient defines the interference effect, and c (c_0) relates to the mean RT. The initial estimate for the I_{n2i} coefficient is a positive term, and the estimate for the I_{i2n} is close to zero. When the sequence of interference trials is accounted, the RT positively correlates to the previous reaction time which is defined by Z_{k-1} coefficient. **d.** S_v (σ_ε^2) update using the EM algorithm. The S_v converges to 0.004 through EM iteration update. **e.** The $x_{k|K}$, cognitive flexibility, for the experiment. The $x_{k|K}$ increases in the midst of the block which correlates with a higher RT for I_{n2i} trials. It declines afterward, which corresponds to a lower RT for I_{n2i} trials. **f.** The observed RT plus the prediction result. The red line shows the GLM prediction without the cognitive state term, and the blue line shows the RT prediction using the cognitive term. The blue shows a better prediction accuracy than the red line.

REFERENCES

[1] Smith, A., and E. Brown. "Estimating a state-space model from point process observations." *Neural Computation* 15.5 (2003): 965-991.

[2] Smith, Anne C., et al. "Dynamic analysis of learning in behavioral experiments." *The journal of neuroscience* 24.2 (2004): 447-461.

[3] Srinivasan, Lakshminarayan, et al. "A state-space analysis for reconstruction of goal-directed movements using neural signals." *Neural computation* 18.10 (2006): 2465-2494.

[4] Kang, X., Sarma, S., Santaniello, S., Schieber, M., and Thakor, N., "Task-Independent Cognitive State Transition Detection From Cortical Neurons During 3D Reach-to-Grasp Movements," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* , vol.PP, no.99, pp.1,1

[5] Bilmes, Jeff A. "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models."*International Computer Science Institute* 4.510 (1998): 126.

[6] Shumway, Robert H., and David S. Stoffer. "An approach to time series smoothing and forecasting using the EM algorithm." *Journal of time series analysis* 3.4 (1982): 253-264.

[7] Bush, George, and Lisa M. Shin. "The Multi-Source Interference Task: an fMRI task that reliably activates the cingulo-frontal-parietal cognitive/attention network." *NATURE PROTOCOLS-ELECTRONIC EDITION-* 1.1 (2006): 308.

[8] Prerau, Michael J., et al. "Characterizing learning by simultaneous analysis of continuous and binary measures of performance." *Journal of neurophysiology*102.5 (2009): 3060-3072.

[9] McCullagh, Peter, John A. Nelder, and P. McCullagh. *Generalized linear models*. Vol. 2. London: Chapman and Hall, 1989.

[10] Eden, U., et al. "Dynamic analysis of neural encoding by point process adaptive filtering." *Neural computation* 16.5 (2004): 971-998.

[11] Mendel, Jerry M. *Lessons in estimation theory for signal processing, communications, and control*. Pearson Education, 1995.

[12] Stroop, J. Ridley. "Studies of interference in serial verbal reactions." *Journal of experimental psychology* 18.6 (1935): 643.

[13] Gratton, Gabriele, Michael GH Coles, and Emanuel Donchin. "Optimizing the use of information: strategic control of activation of responses." *Journal of Experimental Psychology: General* 121.4 (1992): 480.

[14] Schmidt, James R., and Jan De Houwer. "Now you see it, now you don't: Controlling for contingencies and stimulus repetitions eliminates the Gratton effect." *Acta psychologica* 138.1 (2011): 176-186.

[15] Davelaar, Eddy J. "When the ignored gets bound: sequential effects in the flanker task." *Frontiers in psychology* 3 (2012).

[16] Sheth, Sameer A., et al. "Human dorsal anterior cingulate cortex neurons mediate ongoing behavioral adaptation." *Nature* 488.7410 (2012): 218-221.