Approximations for Binary Responses in Dynamic Models

Charles C. Driver
Institute of Education, University of Zurich, Switzerland

Keywords: state-space model, binary data, dynamic system, continuous-time, vector autoregression.

Dynamic system models play a critical role in understanding complex processes that evolve over time. While these models are widely applicable in various fields, their significance becomes even more pronounced when applied to psychology. In psychology, the dynamics of individual and group behavior, cognitive processes, and emotional responses are essential areas of study. The accurate modeling of these dynamic systems is pivotal for unraveling the underlying mechanisms, and when these models are extended to incorporate binary data, their relevance becomes even more pronounced.

Dynamic systems models offer a powerful framework to capture the intricate interplay between various psychological variables. These models enable researchers to explore how behaviors, emotions, or cognitive processes change over time, providing insights into the temporal dependencies within the data. This approach recognizes that the past states of a psychological system can significantly impact its future behavior, reflecting the complex and dynamic nature of human cognition and behavior.

Fitting dynamic system models involves accounting for the inherent uncertainty associated with the system's latent state. The latent state, often representing an unobservable or hidden process, plays a pivotal role in understanding how the system evolves. To address this challenge, a range of estimation approaches has been developed. These approaches aim to provide estimates of the latent state that exhibit characteristics of accuracy and unbiasedness. By achieving these characteristics, they enhance our ability to make informed predictions about the system's behavior.

Among these estimation methods, the Kalman filter is a well-established and computationally efficient approach. The Kalman filter provides unbiased state estimates, meaning that, on average, the estimated latent states are centered around the true underlying values. In linear and Gaussian cases the Kalman filter is also efficient, meaning that the estimates from the filter have minimal variance, ensuring that they are not only centered around the true states but also have the least possible spread.

However, in real-world scenarios, many systems exhibit non-linear behavior and non-Gaussian noise. When faced with these complexities, the Kalman filter's performance may degrade. To address non-linear dynamics and non-Gaussian noise, several approximation techniques have been developed. These methods extend the applicability of the Kalman filter to a broader range of systems, enabling the estimation of latent states even when the underlying dynamics do not conform to linearity and Gaussian assumptions. In essence, they allow the Kalman filter to retain its utility in the face of challenging, real-world scenarios, where the system's behavior may deviate from the ideal conditions assumed by the classic Kalman filter. While such extensions to Kalman filters have been discussed and used before in psychological research, non-continuous response data (e.g., binary or categorical observations) have received relatively little focus in terms of methods and software to accommodate them in dynamic systems models. In this paper I focus specifically on the binary data case.

Examples of binary data in psychology include yes/no responses to questions, the presence or absence of specific behaviors, or choices made during decision-making tasks. Incorporating binary data into dynamic system models allows researchers to address questions such as how and why certain events occur, how they propagate through the system, and how they influence future decisions and behaviors. One particularly interesting application for binary data in state-space models may be for the purposes of modelling missingness mechanisms – in such cases the missingness probability may be modelled as a seperate process that interacts with the other system processes and measurements.

When models are difficult to estimate, Bayesian methods are often useful, and indeed there have been a variety of Bayesian approaches to including non-continuous data in state-space models (e.g., Asparouhov et al., 2018; Hecht et al., 2019; Li et al., 2016). Bayesian approaches are in general very flexible, but the flexibility comes with typically high

This work was funded by the Swiss National Science Foundation. Analyses were not preregistered, scripts to reproduce are available in the online supplementary material. Correspondence concerning this article should be addressed to Charles C. Driver, Institute of Education, University of Zurich, Freiestrasse 36, CH-8032 Zurich. E-mail: charles.driver2@uzh.ch

computational cost, which can limit possible use-cases or limit alternative model comparisons and sensitivity checks. A substantial reason Bayesian approaches are used in such circumstances often has nothing to do with the philosophy or use of priors, but simply for what are classically thought of as Bayesian sampling techniques. Sampling techniques have been used in frequentist contexts also however, though can suffer from the same computational costs (e.g., Durbin & Koopman, 2000). Particle filters are a specific form of sequential Monteo-Carlo sampling methodology, tailored for use in dynamic systems. These have been used in a range of works (e.g., Djuric et al., 2008; Henry et al., 2023), and generally offer a relatively performant sampling-based approach. However, while specifics differ depending on application, they are generally still orders of magnitude more costly, in computational terms, than extended Kalman filtering approaches (Daum & Huang, 2003).

Fahrmeir (1992) provided an early seminal formulation of filtering algorithms with exponential family outcome data. Van Rijn (2008) extended on this but the approach requires fixing certain parameters such as state dynamics, the algorithms are complex, and there is a lack of software implementations (Henry et al., 2023).

In this work I describe a computationally fast, extended Kalman filter (EKF) approach to including binary response variables in state-space models, demonstrate the performance in comparison to a Bayesian sampling approach, and detail how such models may be estimated using the ctsem (Driver & Voelkle, 2021; Driver et al., 2017) software for R (R Core Team, 2014), which allows for complex individual differences (Driver & Voelkle, 2018a), continuous and discrete-time formulations, non-linear system and measurement dynamics, and input effects such as interventions (Driver & Voelkle, 2018b). While this approach is an approximation, the computational tractability and capacity for integrating with dynamic systems models of continuous data should prove useful for many modelling circumstances in psychological and related research.

Kalman Filter

The Kalman filter is a recursive, optimal estimation algorithm used for state estimation in linear dynamic systems. It maintains a state estimate ($\hat{\boldsymbol{\eta}}$) and an associated estimate covariance ($\hat{\mathbf{P}}$). The filter operates in two main steps: prediction and update.

Prediction Step

In the prediction step, the state estimate is projected forward in time using a linear transition model (i.e., matrix of regression coefficients and intercept):

$$\hat{\boldsymbol{\eta}}_{u|u-1} = \mathbf{F}_k\hat{\boldsymbol{\eta}}_{u-1|u-1} + \mathbf{b}$$

The prediction is accompanied by the propagation of the state covariance:

$$\hat{\mathbf{P}}_{u|u-1} = \mathbf{F}_k\hat{\mathbf{P}}_{u-1|u-1}\mathbf{F}_k^\top + \mathbf{Q}_k$$

Update Step

In the update step, measurements ($\mathbf{y}_k$) from the current observation occasion u are used to correct the state estimate. The Kalman gain ($\mathbf{K}$) is computed and is used to blend the predicted state estimate with the observed measurements according to the relative uncertainty of each, then the covariance estimate is adjusted again.

$$\mathbf{K}_k = \hat{\mathbf{P}}_{u|u-1}\mathbf{H}_k^\top(\mathbf{H}_k\hat{\mathbf{P}}_{u|u-1}\mathbf{H}_k^\top + \mathbf{R}_k)^{-1}$$

$$\hat{\mathbf{y}}_k = \mathbf{H}_k\hat{\boldsymbol{\eta}}_{u|u-1}$$

$$\hat{\mathbf{P}}_{u|u} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\hat{\mathbf{P}}_{u|u-1}$$

The updated state estimate update is given by:

$$\hat{\boldsymbol{\eta}}_{u|u} = \hat{\boldsymbol{\eta}}_{u|u-1} + \mathbf{K}_k(\mathbf{y}_k - \hat{\mathbf{y}}_k)$$

This classical Kalman filter formulation assumes that the state dynamics and measurements are linear, and that the noise is Gaussian.

Extended Kalman Filter

The Extended Kalman Filter is an extension of the traditional Kalman Filter designed to handle nonlinear dynamic systems. It achieves this by locally linearizing the system at each time step. The core idea behind the EKF is to assume that the system's dynamics and measurement models are locally linear around the current state estimate. This means that within a small neighborhood around the current estimate, the model can be approximated as linear. The EKF operates by performing this local linearization at each time step.

To linearize a nonlinear system, the EKF employs Jacobian matrices. These matrices capture the partial derivatives of the nonlinear functions with respect to the state variables. In the prediction step, the Jacobian matrix ($\mathbf{F}_k$) of the state transition function ($f(\cdot)$) is used. In the update step, the Jacobian matrix ($\mathbf{H}_k$) of the measurement function ($h(\cdot)$) is employed.

Prediction Step

In the prediction step, the EKF propagates the state estimate ($\hat{\boldsymbol{\eta}}_{u|u-1}$) and any additional control variables $\mathbf{x}_k$ (e.g., intervention dummies) through the nonlinear state transition function ($f(\cdot)$). While this propagation is perfect (up to numerical constraints), because the state is uncertain, the covariance also needs to be propagated through the transition

function ($f(\cdot)$). While some more accurate but computationally costly approaches compute this by propagating many samples of possible states through the transition function (and then computing the covariance), the EKF uses the Jacobian matrix ($\mathbf{F}_k$) to locally approximate the nonlinear function as a linear one. With this, the covariance estimate can be computed using the same equation as in the classic filter – in the classic filter, the Jacobian does not need to be explicitly computed because the Jacobian of a linear transition function is just the transition matrix. The local linearization used by the EKF essentially assumes that the system behaves linearly over the short time interval between time steps. The forward-prediction of the state is:

$$\hat{\boldsymbol{\eta}}_{u|u-1} = f(\hat{\boldsymbol{\eta}}_{u-1|u-1}, \mathbf{x}_k)$$

The Jacobian matrix ($\mathbf{F}_k$) of the transition model function is computed as follows:

$$\mathbf{F}_k = \left.\frac{\partial f}{\partial \boldsymbol{\eta}}\right|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}_{u-1|u-1}}$$

The prediction covariance is adjusted based on the Jacobian:

$$\hat{\mathbf{P}}_{u|u-1} = \mathbf{F}_k \hat{\mathbf{P}}_{u-1|u-1} \mathbf{F}_k^\mathsf{T} + \mathbf{Q}_k$$

Update Step

In the update step, new measurements ($\mathbf{y}_k$) are used to correct the state estimate. The predicted measurement is obtained using a nonlinear measurement model function ($h(\cdot)$):

$$\hat{\mathbf{y}}_k = h(\hat{\boldsymbol{\eta}}_{u|u-1})$$

The Jacobian matrix ($\mathbf{H}_k$) of the measurement model function is computed as:

$$\mathbf{H}_k = \left.\frac{\partial h}{\partial \boldsymbol{\eta}}\right|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}_{u|u-1}}$$

The Kalman gain ($\mathbf{K}_k$) can then be computed as in the classical Kalman filter:

$$\mathbf{K}_k = \hat{\mathbf{P}}_{u|u-1} \mathbf{H}_k^\mathsf{T} (\mathbf{H}_k \hat{\mathbf{P}}_{u|u-1} \mathbf{H}_k^\mathsf{T} + \mathbf{R}_u$$

The EKF proceeds iteratively, continuously updating its state estimate based on new measurements. At each time step, it re-evaluates the Jacobian matrices and performs local linearization to account for the nonlinearities in the system.

It is important to note that the EKF's local linearization is an approximation. If the system undergoes significant nonlinear changes between time steps, the EKF's linearization may become inaccurate. This can lead to suboptimal estimates or divergence. As a result, the EKF is typically more effective when the nonlinearities are mild and the state changes are not too abrupt.

Likelihood for Continuous Variables

When dealing with continuous variables, the likelihood of data is often assumed to follow a Gaussian distribution due to its mathematical convenience and statistical properties. Given the expectation for y (the observation) ($\mathbb{E}(\mathbf{y})$), a measurement ($\mathbf{y}$), and an associated measurement covariance ($\mathbf{R}$), the likelihood of observing the measurement can be expressed as:

$$\mathcal{L}(\mathbf{y} \mid \mathbb{E}(\mathbf{y}), \mathbf{R}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{R})}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbb{E}(\mathbf{y}))^\mathsf{T} \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbb{E}(\mathbf{y}))\right)$$

Where: - $\mathcal{L}(\cdot)$ represents the likelihood function. - $\mathbf{y}$ is the observed measurement. - $\mathbb{E}(\mathbf{y})$ represents the expectation for y (the observation). - $\mathbf{R}$ is the measurement covariance. - $\mathbf{H}$ is the measurement model matrix.

In the case of continuous data, this likelihood describes how well the observed measurement aligns with the expected measurement, given the uncertainty represented by the measurement covariance $\mathbf{R}$.

Extended Kalman Filter for Binary Data

To deal with binary response variables, some specific adjustments can be made to the EKF. In the update step, the non-linear measurement function ($h(\cdot)$) needs to include an inverse logit to transform from any real value to a probability between 0 and 1 (inv_logit).

In this context, the observations follow a Bernoulli distribution, which is inherently non-Gaussian.The innovation (or residual) is calculated as the difference between the actual binary response ($\mathbf{y}_k$) and the predicted observation ($\hat{\mathbf{y}}_u$). At this point, the residual covariance will vary as a function of the latent state η_u – when $h(\eta_u)$ is close to 0 or 1, error due to the binary data will be close to zero, but when $h(\eta_u)$ is a value more in the middle of 0 and 1, the expected residual will be higher. Based on the logit link, the residual covariance matrix is:

$$\mathbf{R}_k = \mathbf{R}_k + \text{diag}(\text{abs}(1 - \hat{\mathbf{y}}_u) \cdot \hat{\mathbf{y}}_u)$$

Because there may be other sources of measurement error (beyond the logit link) it is also possible to view the above equation as an addition to another, pre-specified or estimated, residual covariance matrix. While the size of the residual covariance is correct here, the Gaussian form of it, inherent to the EKF equations, is not, and as such the update step for the next latent state will only be approximate. In particularly difficult cases or when accuracy is more desirable than computational tractability, alternative approaches like particle filters or full Bayesian modeling may offer more robust solutions.

Likelihood With Binary Variables

When incorporating one or more binary variables into the likelihood model, the approach changes to handle binary responses. In this scenario, the likelihood can be defined using the Bernoulli distribution. Each binary variable (z_i) can be considered as a Bernoulli-distributed random variable, and the likelihood of observing these binary variables given the predicted probabilities (p_i) can be expressed as:

$$\mathcal{L}(z_i \mid p_i) = p_i^{z_i}(1 - p_i)^{1-z_i}$$

The likelihood for all binary variables can be computed as the product of individual Bernoulli likelihoods:

$$\mathcal{L}(\mathbf{y} \mid \mathbf{p}) = \prod_{i=1}^{N} p_i^{z_i}(1 - p_i)^{1-z_i}$$

Where: - $\mathcal{L}(z_i \mid p_i)$ represents the Bernoulli likelihood for a single binary variable. - $\mathbf{y}$ is a vector of observed binary responses. - $\mathbf{p}$ is a vector of predicted probabilities for binary responses.

This likelihood models how well the predicted probabilities align with the observed binary responses.

References

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Daum, F., & Huang, J. (2003). Curse of dimensionality and particle filters. *2003 IEEE Aerospace Conference Proceedings (Cat. No. 03TH8652)*, *4*, 4_1979–4_1993.

Djuric, P. M., Vemula, M., & Bugallo, M. F. (2008). Target Tracking by Particle Filtering in Binary Sensor Networks. *IEEE Transactions on Signal Processing*, *56*(6), 2229–2238. https://doi.org/10.1109/TSP.2007.916140

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with r package ctsem. *Journal of Statistical Software*, *77*(5), 1–35. https://doi.org/10.18637/jss.v077.i05

Driver, C. C., & Voelkle, M. C. (2018a). Hierarchical Bayesian continuous time dynamic modeling. *Psychological Methods*, *23*(4), 774. https://doi.org/http://dx.doi.org/10.1037/met0000168

Driver, C. C., & Voelkle, M. C. (2018b). Understanding the time course of interventions with continuous time dynamic models. In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous time modeling in the behavioral and related sciences*. Springer International Publishing.

Driver, C. C., & Voelkle, M. C. (2021, January). Hierarchical continuous time modeling. In J. F. Rauthmann (Ed.), *The Handbook of Personality Dynamics and Processes* (pp. 887–908). Academic Press. https://doi.org/10.1016/B978-0-12-813995-0.00034-0

Durbin, J., & Koopman, S. J. (2000). Time Series Analysis of Non-Gaussian Observations Based on State Space Models from Both Classical and Bayesian Perspectives. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *62*(1), 3–56. https://doi.org/10.1111/1467-9868.00218

Fahrmeir, L. (1992). Posterior Mode Estimation by Extended Kalman Filtering for Multivariate Dynamic Generalized Linear Models. *Journal of the American Statistical Association*, *87*(418), 501–509. https://doi.org/10.1080/01621459.1992.10475232

Hecht, M., Hardt, K., Driver, C. C., & Voelkle, M. C. (2019). Bayesian continuous-time rasch models. *Psychological methods*. https://doi.org/10.1037/met0000205

Henry, T. R., Slipetz, L. R., Falk, A., Qiu, J., & Chen, M. (2023, May). Ordinal Outcome State-Space Models for Intensive Longitudinal Data. https://doi.org/10.48550/arXiv.2305.13444

Li, Q., Pan, J., & Belcher, J. (2016). Bayesian inference for joint modelling of longitudinal continuous, binary and ordinal events. *Statistical Methods in Medical Research*, *25*(6), 2521–2540. https://doi.org/10.1177/0962280214526199

R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Van Rijn, P. W. (2008). Categorical time series in psychological measurement. *Psychometrika*, *62*, 215–236.