An Approximate Kalman Filter for a Binomial Model with Gaussian Latent State

Aaron John Danielson

February 18, 2025

1 Simple Univariate Rate

1.1 Introduction

We consider a state-space model for a sequence of binomial successes $\{y_1, \ldots, y_T\}$ and trials $\{n_1, \ldots, n_T\}$ where the latent state $\theta_t \in \mathbb{R}$ represents the parameter of the binomial distribution at time t. The objective of this project is to estimate θ_t for all t with an approximate Kalman filter given a sequence of time-indexed binomial observations.

1.2 Model Assumptions

1.2.1 State Transition

For each epoch t, the latent state evolves according to

$$\theta_t = \theta_{t-1} + \epsilon_t \tag{1}$$

where

$$\epsilon_t = \sqrt{\tau(t) - \tau(t-1)}\, \epsilon \tag{2}$$

for $\epsilon \sim N(0, \sigma_\theta^2)$ and $\tau(t)$ denotes the time (in days) at which the game indexed by t is played, and ϵ represents the increase in variance per unit time. This formulation implies that the prior variance of θ_t is

$$\mathrm{Var}[\epsilon_t] = \omega_t^2 = (\tau(t) - \tau(t-1))\, \sigma_\theta^2.$$

Thus, we have

$$\theta_t \sim \mathcal{N}\Big(\theta_{t-1}, \omega_t^2\Big). \tag{3}$$

1.2.2 Observation Model

The observation model is given by

$$y_t \sim \mathrm{Binomial}(n_t, \sigma(\theta_t)), \tag{4}$$

with the sigmoid function defined as

$$\sigma(\theta) = \frac{1}{1 + e^{-\theta}}. \tag{5}$$

1.2.3 Prior

Finally, let the initial state of the model parameter be distributed according to a prior π,

$$\theta_0 \sim \pi(\theta).$$

1.3 Derivation of the Laplace Approximation

Let $\mathcal{D}_t = \{(y_1, n_1, \tau(1)), (y_2, n_2, \tau(2)), \ldots, (y_t, n_t, \tau(t))\}$ denote the set of observations, trial counts, and corresponding time stamps up to time t. Given a new observation y_t, we wish to update our belief about θ_t by approximating the posterior distribution

$$p(\theta_t \mid \mathcal{D}_t) \propto p(y_t \mid \theta_t)\, p(\theta_t \mid \mathcal{D}_{t-1}).$$

This section derives closed-form solutions to our problem by approximate the posterior with a Gaussian distribution

$$p(\theta_t \mid \mathcal{D}_t) \approx \mathcal{N}(\theta_t; \mu_t, \sigma^2_{t,\text{post}})$$

where μ_t and $\sigma^2_{t,\text{post}}$ denote the posterior mode and variance, respectively.

1.3.1 Log-Posterior Formulation

In our state-space model, the latent state evolves according to a Markov process, which implies that the current state θ_t depends only on the immediately preceding state θ_{t-1} (and not directly on earlier states or observations). Formally, this means

$$p(\theta_t \mid \mathcal{D}_{t-1}) = p(\theta_t \mid \theta_{t-1}),$$

where $\mathcal{D}_{t-1}$ denotes all observations, trial counts, and time stamps up to time $t-1$. Thus, the Gaussian prior from the state transition can be written as

$$p(\theta_t \mid \mathcal{D}_{t-1}) \propto \exp\left\{-\frac{1}{2}\frac{(\theta_t - \theta_{t-1})^2}{\omega_t^2}\right\}.$$

The observation model is given by the binomial likelihood with log-likelihood

$$\ell(\theta_t; y_t) = y_t \ln \sigma(\theta_t) + (n_t - y_t) \ln\Big(1 - \sigma(\theta_t)\Big).$$

Combining the log-likelihood with the Gaussian prior, the unnormalized log-posterior is

$$Q(\theta_t) = \ell(\theta_t; y_t) - \frac{1}{2}\frac{(\theta_t - \theta_{t-1})^2}{\omega_t^2}.$$

1.3.2 Derivatives of the Log-Likelihood

Computation of the Laplace approximation to the log-posterior requires expressions for the first derivative (gradient) of the log-likelihood function. These are given by

$$g_t = \left.\frac{\partial \ell(\theta_t; y_t)}{\partial \theta_t}\right|_{\theta_t=\theta_t^-} = y_t - n_t\,\sigma(\theta_t^-) \tag{6}$$

and

$$h_t = \left.\frac{\partial^2 \ell(\theta_t; y_t)}{\partial \theta_t^2}\right|_{\theta_t=\theta_t^-} = -n_t\,\sigma(\theta_t^-)\Big(1 - \sigma(\theta_t^-)\Big). \tag{7}$$

For the Laplace approximation, we evaluate these at the prior mean $\theta_t^- = \theta_{t-1}$.

1.3.3 Laplace Approximation

The goal of the Laplace approximation is to approximate the (possibly non-quadratic) log-posterior $Q(\theta_t)$ by a quadratic function. Expand $Q(\theta_t)$ with a second-order Taylor series around the prior mean θ_t^- such that

$$Q(\theta_t) \approx Q(\theta_t^-) + (\theta_t - \theta_t^-)\,Q'(\theta_t^-) + \frac{1}{2}(\theta_t - \theta_t^-)^2\,Q''(\theta_t^-).$$

Since $Q(\theta_t)$ is the sum of the log-likelihood and the log-prior, the first and second derivatives are the sum of the first and second derivatives of the log-likelihood and log-prior functions. When we expand the prior term around $\theta_t = \theta_{t-1}$, its first derivative is

$$-\frac{\theta_t - \theta_{t-1}}{\omega_t^2}$$

so that, when evaluated at $\theta_t^{=}\theta_{t-1}$,

$$Q'(\theta_t^-) = g_t.$$

The second derivative of the prior term is constant

$$-\frac{1}{\omega_t^2},$$

which implies that

$$Q''(\theta_t^-) = h_t - \frac{1}{\omega_t^2}.$$

Thus, the overall Taylor expansion of the unnormalized log-posterior around $\theta_t = \theta_{t-1}$ becomes

$$Q(\theta_t) \approx \text{const} + g_t(\theta_t - \theta_{t-1}) + \frac{1}{2}(\theta_t - \theta_{t-1})^2\left(h_t - \frac{1}{\omega_t^2}\right).$$

1.3.4 Posterior Mode

The mode μ_t of the approximate quadratic log-posterior satisfies

$$g_t + (\mu_t - \theta_{t-1}) \left(h_t - \frac{1}{\omega_t^2} \right) = 0.$$

Solving for μ_t, we obtain

$$\mu_t = \theta_{t-1} - \frac{g_t}{h_t - \frac{1}{\omega_t^2}}.$$

Multiplying numerator and denominator by ω_t^2 gives

$$\mu_t = \theta_{t-1} - \frac{\omega_t^2 \, g_t}{\omega_t^2 \, h_t - 1}.$$

Since h_t is negative for the binomial likelihood, we rewrite the denominator as

$$\omega_t^2 \, h_t - 1 = -\Big(1 - \omega_t^2 \, h_t\Big),$$

so that the update becomes

$$\mu_t = \theta_{t-1} + \frac{\omega_t^2 \, g_t}{1 - \omega_t^2 \, h_t}.$$

This quadratic approximation of the log-posterior is the essence of the Laplace method?it allows us to approximate a non-Gaussian posterior with a Gaussian whose mean is the mode μ_t and whose variance is given by the inverse of the negative second derivative of $Q(\theta_t)$.

1.3.5 Posterior Variance

The Laplace approximation also estimates the posterior variance as the negative inverse of the second derivative of Q at the mode:

$$\sigma_t^2 \approx -\frac{1}{Q''(\mu_t)} \approx \frac{1}{\frac{1}{\omega_t^2} - h_t} = \frac{\omega_t^2}{1 - \omega_t^2 \, h_t}.$$

1.4 Final Update Equations

The recursive update equations for the latent state are therefore:

$$\theta_t = \theta_{t-1} + \frac{\omega_t^2 \, g_t}{1 - \omega_t^2 \, h_t}, \tag{8}$$

$$\sigma_t^2 = \frac{\omega_t^2}{1 - \omega_t^2 \, h_t}, \tag{9}$$

where, evaluated at the prior mean θ_{t-1}, the derivatives are

$$g_t = y_t - n_t \, \sigma(\theta_{t-1}), \tag{10}$$

$$h_t = -n_t \, \sigma(\theta_{t-1}) \Big(1 - \sigma(\theta_{t-1})\Big). \tag{11}$$