Kalman Filter

Tracking a stochastic process with noisy observation

Generative model

Dynamics: $x_n = x_{n-1} + \eta_n$

$\eta_n \sim \mathcal{N}(0,\sigma_\eta^2)$

Gaussian random walk

Observation: $y_n = x_n + \varepsilon_n$

$\varepsilon_n \sim \mathcal{N}(0,\sigma_\varepsilon^2)$

Gaussian noise

Independence: $\perp\{\eta_n,\varepsilon_n | n\in\mathbb{N}\}$

Causal graphical model

Conjugate prior

Gaussian likelihood, parameterized by the mean: $y_n \sim \mathcal{N}(x_n, \sigma_\varepsilon^2)$

Gaussian prior, parameterized by mean and variance

$x_n \sim \mathcal{N}(a,b)$

Posterior

$$p(x_n|y_n) \propto e^{-\frac{1}{2b}(x_n-a)^2} \cdot e^{-\frac{1}{2\sigma_\varepsilon^2}(x_n-y_n)^2}$$

$$\propto e^{-\left[\left(\frac{1}{2b}+\frac{1}{2\sigma_\varepsilon^2}\right)x_n^2-2\left(\frac{1}{2b}a+\frac{1}{2\sigma_\varepsilon^2}y_n\right)x_n\right]}$$

$$\propto e^{-\left(\frac{1}{2b}+\frac{1}{2\sigma_\varepsilon^2}\right)\left(x_n-\frac{\frac{1}{b}a+\frac{1}{\sigma_\varepsilon^2}y_n}{\frac{1}{b}+\frac{1}{\sigma_\varepsilon^2}}\right)^2}$$

Precision-weighted averaging

$$x_n|y_n \sim \mathcal{N}\left(\frac{\frac{1}{b}a+\frac{1}{\sigma_\varepsilon^2}y_n}{\frac{1}{b}+\frac{1}{\sigma_\varepsilon^2}}, \frac{1}{\frac{1}{b}+\frac{1}{\sigma_\varepsilon^2}}\right)$$

Iterative prior

$x_n|y_n \sim \mathcal{N}(c,d)$

$x_{n+1}|y_n = x_n + \eta_n \sim \mathcal{N}(c, d+\sigma_\eta^2)$

Convolution: $p(\alpha+\beta = X) = \int p(\alpha = Z)p(\beta = X - Z)\mathrm{d}Z$

Update rules

$x_n|\mathbf{y}_{n-1} \sim \mathcal{N}(\mu_n, s_n^2)$

$$\mu_{n+1} = \frac{\frac{1}{s_n^2}\mu_n+\frac{1}{\sigma_\varepsilon^2}y_n}{\frac{1}{s_n^2}+\frac{1}{\sigma_\varepsilon^2}} = \frac{\sigma_\varepsilon^2\mu_n+s_n^2 y_n}{\sigma_\varepsilon^2+s_n^2}$$

$$s_{n+1}^2 = \frac{1}{\frac{1}{s_n^2}+\frac{1}{\sigma_\varepsilon^2}} + \sigma_\eta^2 = \frac{\sigma_\varepsilon^2 s_n^2}{\sigma_\varepsilon^2+s_n^2} + \sigma_\eta^2$$

Graphical models

Bayes nets, acyclic directed/causal graphical models

Nodes for variables, arrows for dependencies

Markov property

each variable $x_i$ depends directly only on its immediate parents $\mathrm{Pa}(x_i)$

conditionally independent of all other variables

Joint distribution determined by conditional distributions

$p(\mathbf{x}) = \prod_i p(x_i|x_{\mathrm{Pa}(i)})$

Inference in complex models

Posterior over unobserved variables given observed variables

Prior and likelihood generally easy

e.g., conditional probabilities in graphical models

Normalization term (marginal probability of evidence) often intractable

Or marginalizing out intermediate variables

MCMC – Markov chain Monte Carlo

Design a Markov chain with stationary distribution matching desired posterior
Simulate it and use trajectory as samples

Markov chains and stationary distributions

Transition matrix: $T_{ij} = \Pr[s_{t+1} = S_i \mid s_t = S_j]$

Stationary distribution $p$:

$Tp = p$

$\Sigma_j T_{ji} p_j = p_i$

Eigenvector with eigenvalue 1 (unique if $T$ ergodic)

Example

$T = [.6\ .1\ .1;\ .3\ .8\ 0;\ .1\ .1\ .9]$

$p = [.2;\ .3;\ .5]$

Gibbs sampling

Version of MCMC

Yields joint distribution $\mathrm{p}(\mathbf{x}) = \mathrm{p}(x_1,\ldots,x_n)$

Possibly conditioned on some observables: $\mathrm{p}(\mathbf{x}_{\text{unobserved}} \mid \mathbf{x}_{\text{observed}})$

Cycle repeatedly through unknown variables ($i$)

Sample $x_i \sim \mathrm{p}(x_i \mid \mathbf{x}_{-i})$, where $\mathbf{x}_{-i} = (x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$

Doesn't matter which variables are observed or unobserved; all are held fixed except $x_i$

Bayes net: $\mathrm{p}(x_i \mid \mathbf{x}_{-i}) = \mathrm{p}(x_i \mid \mathbf{x}_{\mathrm{An}(i)}, \mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{x}_{\mathrm{Ch}(i)}, \mathbf{x}_{\mathrm{De}(i)})$ [Ancestors, Parents, Children, Descendants]

$$\propto \mathrm{p}(x_i \mid \mathbf{x}_{\mathrm{An}(i)}, \mathbf{x}_{\mathrm{Pa}(i)}) \cdot \mathrm{p}(\mathbf{x}_{\mathrm{Ch}(i)}, \mathbf{x}_{\mathrm{De}(i)} \mid x_i, \mathbf{x}_{\mathrm{An}(i)}, \mathbf{x}_{\mathrm{Pa}(i)})$$
$$= \mathrm{p}(x_i \mid \mathbf{x}_{\mathrm{Pa}(i)}) \cdot \mathrm{p}(\mathbf{x}_{\mathrm{Ch}(i)}, \mathbf{x}_{\mathrm{De}(i)} \mid x_i)$$
$$= \mathrm{p}(x_i \mid \mathbf{x}_{\mathrm{Pa}(i)}) \cdot \prod_{j\in \mathrm{Ch}(i)\cup \mathrm{De}(i)} \mathrm{p}(x_j \mid \mathbf{x}_{\mathrm{Pa}(j)})$$
$$\propto \mathrm{p}(x_i \mid \mathbf{x}_{\mathrm{Pa}(i)}) \cdot \prod_{j\in \mathrm{Ch}(i)} \mathrm{p}(x_j \mid \mathbf{x}_{\mathrm{Pa}(j)})$$

Stationary distribution is $\mathrm{p}(x_1,\ldots,x_n)$

Preserved under each update step

Exercises

1. Compare the Kalman filter to simple RL (with no cue). Look at their updating rules and explain how they relate. Extra challenge: building on this connection, try to derive a Bayesian version of Rescorla-Wagner (hint – assume the weights follow Gaussian random walks).

2. Generate data from a Kalman filter, meaning the sequence of mean predictions across trials, for some interesting sequence of observations. Fit the Kalman and RL models to the data and compute AIC for each model. If you want more, create data from an RL model on the same observation sequence, and then fit Kalman and RL models to these data and compute AICs.

3. Prove that $\mathrm{p}(\mathbf{x}_{\text{unobserved}}|\mathbf{x}_{\text{observed}})$ is the stationary distribution for Gibbs sampling. That is, let $\mathbf{z}$ represent the sample at any step in the Markov chain, and treat $\mathbf{z}$ as a random variable with distribution matching $\mathrm{p}(\mathbf{x}_{\text{unobserved}}|\mathbf{x}_{\text{observed}})$. Then define $\mathbf{z}'$ as the next sample, where $\mathbf{z}'_i$ is drawn from $\mathrm{p}(x_i \mid \mathbf{x}_{-i} = \mathbf{z}_{-i})$ for some unobserved variable $x_i$, and all other components of $\mathbf{z}'$ are unchanged (i.e., $\mathbf{z}'_j = \mathbf{z}_j$ for $j \neq i$). Show that the distribution of $\mathbf{z}'$ also matches $\mathrm{p}(\mathbf{x}_{\text{unobserved}}|\mathbf{x}_{\text{observed}})$.
[Hint – let $\mathbf{y}$ stand for any possible value of $\mathbf{x}_{\text{unobserved}}$. You know $\mathrm{p}(\mathbf{z}=\mathbf{y}) = \mathrm{p}(\mathbf{x}_{\text{unobserved}}=\mathbf{y}|\mathbf{x}_{\text{observed}})$ for any $\mathbf{y}$. Using this fact, show that the same statement holds about $\mathrm{p}(\mathbf{z}'=\mathbf{y})$.]