State Space Models!

$$y_t = a x_t + v_t \quad \leftarrow \text{observation equation}$$

$$x_t = \phi x_{t-1} + w_t \quad \leftarrow \text{state equation}$$

$$v_t \overset{iid}{\sim} N(0, \sigma_v^2), \quad w_t \overset{iid}{\sim} N(0, \sigma_w^2), \quad x_1 = \mu$$

Allows us to define a joint probability distribution for $\underset{\sim}{x}$ and $\underset{\sim}{y}$

$$\begin{pmatrix} \underset{\sim}{y} \\ \underset{\sim}{x} \end{pmatrix} \sim N\left(\begin{pmatrix} a\,\mathbb{E}[\underset{\sim}{x}] \\ \mathbb{E}[\underset{\sim}{x}] \end{pmatrix}, \begin{pmatrix} a^2\mathbb{V}[\underset{\sim}{x}] + \sigma_v^2 I & a\mathbb{V}[\underset{\sim}{x}] \\ a\mathbb{V}[\underset{\sim}{x}] & \mathbb{V}[\underset{\sim}{x}] \end{pmatrix} \right)$$

* Smoother: $\underset{\sim}{x} \mid \underset{\sim}{y}$ } involves inverting & multiplying $n \times n$ matrices
* Filtered values: $x_t \mid y_1, \ldots y_t$] n diff. cond. distrib.
* Forecasts: $x_t \mid y_1, \ldots, y_{t-1}$] n-1 conditional distrib.

can derive all of these from the joint distribution

Kalman Filter allows us to quickly and recursively compute forecasts, forecast variances, filtered values, and filter variances without wasting any computations!

Start at $t=1$

$$* \ \mathbb{E}[x_t \mid y_1,\dots,y_{t-1}] = \phi\,\mathbb{E}[x_{t-1} \mid y_1,\dots,y_{t-2}] + \left(\frac{a\,\mathbb{V}[x_{t-1} \mid y_1,\dots,y_{t-2}]}{a^2\,\mathbb{V}[x_{t-1} \mid y_1,\dots,y_{t-2}] + \sigma_v^2}\right) \times \left(y_{t-1} - a\,\mathbb{E}[x_{t-1} \mid y_1,\dots,y_{t-2}]\right)$$

$$* \ \mathbb{V}[x_t \mid y_1,\dots,y_{t-1}] = \sigma_w^2 + \phi^2\,\mathbb{V}[x_{t-1} \mid y_1,\dots,y_{t-2}]\left(1 - \frac{a\,\mathbb{V}[x_{t-1} \mid y_1,\dots,y_{t-2}]}{a^2\,\mathbb{V}[x_{t-1} \mid y_1,\dots,y_{t-2}] + \sigma_v^2}\right)$$

$$* \ \mathbb{E}[x_t \mid y_1,\dots,y_t] = \mathbb{E}[x_t \mid y_1,\dots,y_{t-1}] + \left[\frac{a\,\mathbb{V}[x_t \mid y_1,\dots,y_{t-1}]}{a^2\,\mathbb{V}[x_t \mid y_1,\dots,y_{t-1}] + \sigma_v^2}\right]\left(y_t - a\,\mathbb{E}[x_t \mid y_1,\dots,y_{t-1}]\right)$$

$$* \ \mathbb{V}[x_t \mid y_1,\dots,y_t] = \mathbb{V}[x_t \mid y_1,\dots,y_{t-1}]\left(1 - \frac{a\,\mathbb{V}[x_t \mid y_1,\dots,y_{t-1}]}{a^2\,\mathbb{V}[x_t \mid y_1,\dots,y_{t-1}] + \sigma_v^2}\right)$$

Output of Kalman Filter (specifically the forecasts)

gives us what we need to easily compute forecasts & forecast variances for the y_t's

$$\mathbb{E}[y_t \mid y_1, \ldots, y_{t-1}] = a\,\mathbb{E}[x_t \mid y_1, \ldots, y_{t-1}]$$

$$\mathbb{V}[y_t \mid y_1, \ldots, y_{t-1}] = a^2\,\mathbb{V}[x_t \mid y_1, \ldots, y_{t-1}] + \sigma_v^2$$

Gives us a quick & efficient way of evaluating

$$p(\underline{y}) = \underbrace{p(y_1) \prod_{t=2}^{n} p(y_t \mid y_1, \ldots, y_{t-1})}$$

No matrix inversions, just the product of n normal densities

Kalman Smoother computes $E[x_t \mid \underset{\sim}{y}]$, $\mathbb{V}[x_t \mid \underset{\sim}{y}]$ using output from the Kalman filter but in reverse,

Start at $t=n$ $\underbrace{E[x_n \mid y_1, \ldots, y_n]}_{\text{filter at } x_n}$, $\underbrace{\mathbb{V}[x_n \mid y_1, \ldots, y_n]}_{\text{filter variance at } x_n}$

Recursively computes

$$\mathbb{E}[x_t \mid y_1, \ldots, y_n] \text{ and } \mathbb{V}[x_t \mid y_1, \ldots, y_n]$$

using $\mathbb{E}[x_{t+s} \mid y_1, \ldots, y_n]$ and $\mathbb{V}[x_{t+s} \mid y_1, \ldots, y_n]$

Takes n steps, <u>but</u> each step is simple, doesn't involve inverting matrices

Estimating the State-Space Model Parameters

We have a joint distribution of $\begin{pmatrix} \underset{\sim}{y} \\ \underset{\sim}{x} \end{pmatrix}$

Need maximize $p(\underset{\sim}{y}) = \int p(\underset{\sim}{x}, \underset{\sim}{y}) p(\underset{\sim}{x}) d\underset{\sim}{x}$

average over what we don't know

Fortunately!

If $\begin{pmatrix} \underset{\sim}{y} \\ \underset{\sim}{x} \end{pmatrix}$ are normally distributed, then $\underset{\sim}{y}$ is normally distributed

maximize w.r.t. $a, \phi, \sigma^2_{\sim}, \sigma^2_{\sim}, \mu$

$$p(\underset{\sim}{y}) = p(y_1) \prod_{t=2}^{n} p(y_t \mid y_1, \ldots, y_{t-1})$$

we can compute this using the kalman filter

This is much easier to work with than

$$\underset{\sim}{y} \sim N\left(a\mathbb{E}[\underset{\sim}{x}],\ a^2\mathbb{V}[\underset{\sim}{x}] + \sigma^2_{\sim} I_n\right)$$

likelihood using this rep. is diff. to compute

Maximizing $p(\underset{\sim}{y})$ in Practice

Direct Maximization

Use conditional representation

$$p(\underset{\sim}{y}) = p(y_1)\prod_{t=2}^{n} p(y_t \mid y_1, \ldots, y_{t-1})$$

$$(\ast) = \frac{1}{\sqrt{2\pi}^{\,n}}\frac{1}{\sqrt{\mathbb{V}[y_1]}}\exp\left\{-\frac{(y_1-\mathbb{E}[y_1])^2}{2\mathbb{V}[y_1]}\right\} \times \prod_{t=2}^{n}\frac{1}{\sqrt{\mathbb{V}[y_t \mid y_1, \ldots, y_{t-1}]}}\exp\left\{-\frac{(y_t-\mathbb{E}[y_t \mid y_1, \ldots, y_{t-1}])^2}{2\mathbb{V}[y_t \mid y_1, \ldots, y_{t-1}]}\right\}$$

* works well if starting a_1, ϕ_1, μ_1, $\sigma_{v_1}^2$, σ_w^2 are "good"

Take an iterative approach...

Define $r_t = y_t - \mathbb{E}[y_t \mid y_1, \ldots, y_{t-1}]$

$$(\heartsuit) = \frac{1}{\sqrt{2\pi}^{\,n}}\frac{1}{\sqrt{\mathbb{V}[y_1]}}\exp\left\{-\frac{r_1^2}{2\mathbb{V}[y_1]}\right\} \times \prod_{t=2}^{n}\frac{1}{\sqrt{\mathbb{V}[y_t \mid y_1, \ldots, y_{t-1}]}}\exp\left\{-\frac{r_t^2}{2\mathbb{V}[y_t \mid y_1, \ldots, y_{t-1}]}\right\}$$

* Fix variances and maximize with respect to $\mathbb{E}[y_t \mid y_1, \ldots, y_{t-1}]$ using $(\ast)$
* Compute r_t
* Maximize $(\heartsuit)$ fixing r_t's