State Space Models:

$$y_t = a x_t + v_t$$ ← observation equation

$$x_t = \phi x_{t-1} + w_t$$ ← state equation

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

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 nxn 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

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

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, but 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 $p(\underset{\sim}{y}) = \int \underbrace{p(\underset{\sim}{x}, \underset{\sim}{y})}_{\text{average over what we don't know}} d\underset{\sim}{x}$

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

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

] likelihood using this rep. is diff. to compute

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

Use conditional representation

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

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

Direct Maximization

* works well if starting a_1, ϕ_1, μ, σ_v^2, σ_w^2 are "good"

Take an iterative approach...

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

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

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

Indirect Maximization of $p(\underset{\sim}{y})$

$$p(\underset{\sim}{y}) = \int p(\underset{\sim}{x}, \underset{\sim}{y})\, d\underset{\sim}{x}$$
$$= \int p(\underset{\sim}{y} \mid \underset{\sim}{x})\, p(\underset{\sim}{x})\, d\underset{\sim}{x} = \mathbb{E}\left[p(\underset{\sim}{y} \mid \underset{\sim}{x})\right]$$

depend on the parameters $\phi, a, \sigma_w^2, \sigma_v^2, \mu$ in a "nicer" way than $p(\underset{\sim}{y})$ itself

Expectation Maximization, Nan Laird, Arthur Dempster, Don Rubin

↓

EM · Algorithm

Maximizing $p(\underset{\sim}{y})$ is equivalent to maximizing $\mathbb{E}\left[\log\left(p(\underset{\sim}{y} \mid \underset{\sim}{x})\, p(\underset{\sim}{x})\right) \mid \underset{\sim}{y}\right]$

We will iteratively maximize:

$$\mathbb{E}[\log(p(\underset{\sim}{y} \mid \underset{\sim}{x})\, p(\underset{\sim}{x}) \mid y_1, \dots, y_n] =$$

$$\mathbb{E}[\log(p(y \mid x)) \mid \underset{\sim}{y}] \quad + \mathbb{E}[\log(p(\underset{\sim}{x})) \mid \underset{\sim}{y}]$$

$$\mathbb{E}[\log(p(y \mid x)) \mid \underset{\sim}{y}] =$$

$$\mathbb{E}\left[\log\left(\left(\frac{1}{\sqrt{2\pi\sigma_v^2}}\right)^n \exp\left\{-\frac{1}{2\sigma_v^2}\sum_{i=1}^{n}(y_t - a x_t)^2\right\}\right) \mid \underset{\sim}{y}\right] =$$

$$K_1 + \quad \frac{n}{2}\log(\sigma_v^2) - \frac{1}{2\sigma_v^2}\sum_{i=1}^{n}\mathbb{E}\left[(y_t - a x_t)^2 \mid \underset{\sim}{y}\right] =$$

$$K_1 + \frac{n}{2}\log(\sigma_v^2) - \frac{1}{2\sigma_v^2}\sum_{i=1}^{n} y_t^2 - 2a y_t \mathbb{E}[x_t \mid \underset{\sim}{y}] + a^2\mathbb{E}[x_t^2 \mid y]$$

We will iteratively maximize:

$$\mathbb{E}[\log(p(y|\underset{\sim}{x})p(\underset{\sim}{x})) \mid y_1, \dots, y_n] =$$

$$\mathbb{E}[\log(p(y|x)) \mid y] + \mathbb{E}[\log(p(\underset{\sim}{x})) \mid y]$$

$$\mathbb{E}[\log(p(\underset{\sim}{x})) \mid y] =$$

$$\mathbb{E}\left[\log\left(p(x_1)\prod_{t=2}^{n} p(x_t | x_1, \dots, x_{t-1})\right) \Big| y\right] =$$

$$\mathbb{E}\Bigg[\log\Bigg(\frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left\{-\frac{1}{2\sigma_w^2}(x_2 - \phi\mu)^2\right\} \times \frac{1}{(\sqrt{2\pi\sigma_w^2})^{n-2}} \exp\left\{\frac{-1}{2\sigma_w^2}\sum_{t=3}^{n}(x_t - \phi x_{t-1})^2\right\}\Bigg) \Bigg| y\Bigg]$$

$$= K_2 + \frac{(n-1)}{2}\log(\sigma_w^2) - \mathbb{E}\left[\frac{1}{2\sigma_w^2}\left((x_2 - \phi\mu)^2 + \sum_{t=3}^{n}(x_t - \phi x_{t-1})^2\right) \Big| y\right]$$

$$= K_2 + \frac{(n-1)}{2}\log(\sigma_w^2) - \frac{1}{2\sigma_w^2}\Bigg[\mathbb{E}[x_2^2|y] - 2\phi\mu\,\mathbb{E}[x_2|y] + \phi^2\mu^2 + \sum_{t=3}^{n}\mathbb{E}[x_t^2|y] - 2\phi\,\mathbb{E}[x_t x_{t-1}|y] + \phi^2\,\mathbb{E}[x_{t-1}^2|y]\Bigg]$$

$\mathbb{E}[\log(p(y|\underset{\sim}{x})p(\underset{\sim}{x})) \mid y_1, \ldots, y_n] =$

$\mathbb{E}[\log(p(y|x)) \mid \underset{\sim}{y}] + \mathbb{E}[\log(p(\underset{\sim}{x})) \mid \underset{\sim}{y}] =$

$$K_1 + \frac{n}{2}\log(\sigma_v^2) - \frac{1}{2\sigma_v^2}\sum_{i=1}^{n} y_t^2 - 2a y_t \mathbb{E}[x_t|y] + a^2\mathbb{E}[x_t^2|y] + K_2 + \frac{(n-1)}{2}\log(\sigma_w^2) - \frac{1}{2\sigma_w^2}\Big[\mathbb{E}[x_2^2|y] - 2\phi\mu\mathbb{E}[x_2|y] + \phi^2\mu^2 + \sum_{t=3}^{n}\mathbb{E}[x_t^2|y] - 2\phi\mathbb{E}[x_t x_{t-1}|y] + \phi^2\mathbb{E}[x_{t-1}^2|y]\Big)$$

EM Algorithm:

a nice thing about EM is that starting vals can be bad

Start with initial values $a^{(0)}, \phi^{(0)}, \sigma_v^{2(0)}, \sigma_w^{2(0)}, \mu^{(0)}$

(a) Compute $\mathbb{E}[x_t|\underset{\sim}{y}]^{(i-1)}, \mathbb{E}[x_t^2|\underset{\sim}{y}]^{(i-1)}, \mathbb{E}[x_t x_{t-1}|\underset{\sim}{y}]^{(i-1)}$

for $a^{(i-1)}, \phi^{(i-1)}, \sigma_v^{2(i-1)}, \sigma_w^{2(i-1)}, \mu^{(i-1)}$

(b) Maximize expected joint log likelihood to get

$\sigma_v^{2(i)}, \sigma_w^{2(i)}, a^{(i)}, \phi^{(i)}, \mu^{(i)}$ fixing the expectations

Iterate (a) and (b) until parameters stop changing (can take a long time)

In practice, it is common to:

* First optimize using EM with low convergence threshold to get initial $\mu^{(EM)}, \phi^{(EM)}, a^{(EM)}, \sigma_N^{2\,(EM)}, \sigma_W^{2\,(EM)}$

* Second, use direct maximization of $p(\underline{y})$ starting from EM estimates to get our final estimates

$\hat{\mu} \quad \hat{\phi} \quad \hat{a} \quad \hat{\sigma}_N^2 \quad \hat{\sigma}_W^2$

may not have a global maximum :/