Introduction to state-space models

nicolas.chopin@ensae.fr

based on a previous PG course with O. Papaspiliopoulos

Presentation of state-space models

Objectives

The sequential analysis of state-space models is the main (but not only) application of Sequential Monte Carlo.

Objectives

The sequential analysis of state-space models is the main (but not only) application of Sequential Monte Carlo.

The aim of this chapter is to define state-space models, give examples of such models from various areas of science, and discuss their main properties.

A first definition (with functions)

A time series model that consists of two discrete-time processes $\{X_t\} := (X_t)_{t\geq 0}$, $\{Y_t\} := (Y_t)_{t\geq 0}$, taking values respectively in spaces $\mathcal{X}$ and $\mathcal{Y}$, such that

$$\begin{aligned} X_t &= K_t(X_{t-1}, U_t, \theta), \quad t \geq 1 \\ Y_t &= H_t(X_t, V_t, \theta), \quad t \geq 0 \end{aligned}$$

where K_0, K_t, H_t, are determistic functions, $\{U_t\}$, $\{V_t\}$ are sequences of i.i.d. random variables (*noises*, or *shocks*), and $\theta \in \Theta$ is an unknown parameter.

A first definition (with functions)

A time series model that consists of two discrete-time processes $\{X_t\}:=(X_t)_{t\geq 0}$, $\{Y_t\}:=(Y_t)_{t\geq 0}$, taking values respectively in spaces $\mathcal{X}$ and $\mathcal{Y}$, such that

$$\begin{aligned} X_t &= K_t(X_{t-1}, U_t, \theta), \quad t \geq 1 \\ Y_t &= H_t(X_t, V_t, \theta), \quad t \geq 0 \end{aligned}$$

where K_0, K_t, H_t, are determistic functions, $\{U_t\}$, $\{V_t\}$ are sequences of i.i.d. random variables (*noises*, or *shocks*), and $\theta \in \Theta$ is an unknown parameter.

This is a popular way to define SSMs in Engineering. Rigorous, but not sufficiently general.

A second definition (with densities)

$$
\begin{aligned}
p_\theta(x_0) &= p_0^\theta(x_0) \\
p_\theta(x_t|x_{0:t-1}) &= p_t^\theta(x_t|x_{t-1}) \quad t \geq 1 \\
p_\theta(y_t|x_{0:t}, y_{0:t-1}) &= f_t^\theta(y_t|x_t)
\end{aligned}
\tag{1.1}
$$

A second definition (with densities)

$$\begin{aligned}
p_\theta(x_0) &= p_0^\theta(x_0) \\
p_\theta(x_t|x_{0:t-1}) &= p_t^\theta(x_t|x_{t-1}) \quad t \geq 1 \\
p_\theta(y_t|x_{0:t}, y_{0:t-1}) &= f_t^\theta(y_t|x_t)
\end{aligned} \tag{1.1}$$

Not so rigorous (or not general enough): some models are such that $X_t|X_{t-1}$ does not admit a probability density (with respect to a fixed dominating measure).

Examples of state-space models

Signal processing: tracking, positioning, navigation

X_t is position of a moving object, e.g.

$$X_t = X_{t-1} + U_t, \quad U_t \sim \mathcal{N}_2(0, \sigma^2 I_2),$$

and Y_t is a measurement obtained by e.g. a radar,

$$Y_t = \mathrm{atan}\left(\frac{X_t(2)}{X_t(1)}\right) + V_t, \quad V_t \sim \mathcal{N}_1(0, \sigma_Y^2).$$

and $\theta = (\sigma^2, \sigma_Y^2)$.

Signal processing: tracking, positioning, navigation

X_t is position of a moving object, e.g.

$$X_t = X_{t-1} + U_t, \quad U_t \sim \mathcal{N}_2(0, \sigma^2 I_2),$$

and Y_t is a measurement obtained by e.g. a radar,

$$Y_t = \mathrm{atan}\left(\frac{X_t(2)}{X_t(1)}\right) + V_t, \quad V_t \sim \mathcal{N}_1(0, \sigma_Y^2).$$

and $\theta = (\sigma^2, \sigma_Y^2)$.
(This is called the **bearings-only tracking** model.)

GPS

In GPS applications, the velocity v_t of the vehicle is observed, so motion model is (some variation of):

$$X_t = X_{t-1} + v_t + U_t, \qquad U_t \sim \mathcal{N}_2(0, \sigma^2 I_2).$$

Also Y_t usually consists of more than one measurement.

More advanced motion model

A random walk is too erratic for modelling the position of the target; assume instead its velocitity follows a random walk. Then define:

$$X_t = \begin{pmatrix} I_2 & I_2 \\ 0_2 & I_2 \end{pmatrix} X_{t-1} + \begin{pmatrix} 0_2 & 0_2 \\ 0_2 & U_t \end{pmatrix}, \quad U_t \sim \mathcal{N}_2(0, \sigma^2 I_2),$$

with obvious meanings for matrices 0_2 and I_2.

More advanced motion model

A random walk is too erratic for modelling the position of the target; assume instead its velocitity follows a random walk. Then define:

$$X_t = \begin{pmatrix} I_2 & I_2 \\ 0_2 & I_2 \end{pmatrix} X_{t-1} + \begin{pmatrix} 0_2 & 0_2 \\ 0_2 & U_t \end{pmatrix}, \quad U_t \sim \mathcal{N}_2(0, \sigma^2 I_2),$$

with obvious meanings for matrices 0_2 and I_2.

Note: $X_t(1)$ and $X_t(2)$ (position) are deterministic functions of X_{t-1}: no probability density for $X_t|X_{t-1}$.

multi-target tracking

Same ideas except $\{X_t\}$ now represent the position (and velocity if needed) of a set of targets (of random size); i.e. $\{X_t\}$ is a point process.

Time series of counts (neuro-decoding, astrostatistics, genetics)

- Neuro-decoding: Y_t is a vector of d_y counts (spikes from neuron k),

$$Y_t(k)|X_t \sim \mathcal{P}(\lambda_k(X_t)), \quad \log \lambda_k(X_t) = \alpha_k + \beta_k X_t,$$

  and X_t is position+velocity of the subject's hand (in 3D).
- astro-statistics: Y_t is number of photon emissions; intensity varies over time (according to an auto-regressive process)
- Y_t is the number of 'reads', which is a noisy measurement of the transcription level X_t at position t in the genome;

Time series of counts (neuro-decoding, astrostatistics, genetics)

- Neuro-decoding: Y_t is a vector of d_y counts (spikes from neuron k),

$$Y_t(k)|X_t \sim \mathcal{P}(\lambda_k(X_t)), \quad \log \lambda_k(X_t) = \alpha_k + \beta_k X_t,$$

  and X_t is position+velocity of the subject's hand (in 3D).
- astro-statistics: Y_t is number of photon emissions; intensity varies over time (according to an auto-regressive process)
- Y_t is the number of 'reads', which is a noisy measurement of the transcription level X_t at position t in the genome;

Note: 'functional' definition of state-space models is less convenient in this case.

Stochastic volatility (basic model)

Y_t is log-return of asset price, $Y_t = \log(p_t/p_{t-1})$,

$$Y_t | X_t = x_t \sim \mathcal{N}\left(0, \exp(x_t)\right)$$

where $\{X_t\}$ is an auto-regressive process:

$$X_t - \mu = \phi(X_{t-1} - \mu) + U_t, \quad U_t \sim \mathcal{N}(0, \sigma^2)$$

and $\theta = (\mu, \phi, \sigma^2)$.

Stochastic volatility (basic model)

Y_t is log-return of asset price, $Y_t = \log(p_t/p_{t-1})$,

$$Y_t|X_t = x_t \sim \mathcal{N}\left(0, \exp(x_t)\right)$$

where $\{X_t\}$ is an auto-regressive process:

$$X_t - \mu = \phi(X_{t-1} - \mu) + U_t, \quad U_t \sim \mathcal{N}(0, \sigma^2)$$

and $\theta = (\mu, \phi, \sigma^2)$.
Take $|\phi| < 1$ and $X_0 \sim N(\mu, \sigma^2/(1-\rho^2))$ to impose stationarity.

Stochastic volatility (variations)

- Student dist' for noises
- skewness: $Y_t = \alpha X_t + \exp(X_t/2) V_t$
- leverage effect: correlation between U_t and V_t
- multivariate extensions

Nonlinear dynamic systems in Ecology, Epidemiology, and other fields

$Y_t = X_t + V_t$, where $\{X_t\}$ is some complex nonlinear dynamic system. In Ecology for instance,

$$X_t = X_{t-1} + \theta_1 - \theta_2 \exp(\theta_3 X_{t-1}) + U_t$$

where X_t is log of population size. For some values of θ, process is nearly chaotic.

Nonlinear dynamic systems: Lokta-Volterra

Predator-prey model, where $\mathcal{X} = (\mathbb{Z}^+)^2$, $X_t(1)$ is the number of preys, $X_t(2)$ is the number of predators, and, working in continuous-time:

$$\begin{aligned} X_t(1) &\overset{\theta_1}{\rightarrow} 2X_t(1) \\ X_t(1) + X_t(2) &\overset{\theta_2}{\rightarrow} 2X_t(2), \quad t \in \mathbb{R}^+ \\ X_t(2) &\overset{\theta_3}{\rightarrow} 0 \end{aligned}$$

(but Y_t still observed in discrete time.)

Nonlinear dynamic systems: Lokta-Volterra

Predator-prey model, where $\mathcal{X} = (\mathbb{Z}^+)^2$, $X_t(1)$ is the number of preys, $X_t(2)$ is the number of predators, and, working in continuous-time:

$$\begin{aligned} X_t(1) &\stackrel{\theta_1}{\rightarrow} 2X_t(1) \\ X_t(1) + X_t(2) &\stackrel{\theta_2}{\rightarrow} 2X_t(2), \quad t \in \mathbb{R}^+ \\ X_t(2) &\stackrel{\theta_3}{\rightarrow} 0 \end{aligned}$$

(but Y_t still observed in discrete time.)
Intractable dynamics: We can simulate from $X_t|X_{t-1}$, but we can't compute $p_t(x_t|x_{t-1})$.

Nonlinear dynamic systems: Lokta-Volterra

Predator-prey model, where $\mathcal{X} = (\mathbb{Z}^+)^2$, $X_t(1)$ is the number of preys, $X_t(2)$ is the number of predators, and, working in continuous-time:

$$\begin{aligned} X_t(1) &\overset{\theta_1}{\to} 2X_t(1) \\ X_t(1) + X_t(2) &\overset{\theta_2}{\to} 2X_t(2), \quad t \in \mathbb{R}^+ \\ X_t(2) &\overset{\theta_3}{\to} 0 \end{aligned}$$

(but Y_t still observed in discrete time.)
Intractable dynamics: We can simulate from $X_t|X_{t-1}$, but we can't compute $p_t(x_t|x_{t-1})$.
see also compartmental models in Epidemiology.

State-space models with an intractable or degenerate observation process

We have seen models such that $X_t|X_{t-1}$ is intractable; $Y_t|X_t$ may be intractable as well. Let

$$X_t' = (X_t, Y_t), \quad Y_t' = Y_t + V_t, \quad V_t \sim \mathcal{N}(0, \sigma^2)$$

and use $\{(X_t', Y_t')\}$ as an approximation of $\{(X_t, Y_t)\}$.

State-space models with an intractable or degenerate observation process

We have seen models such that $X_t|X_{t-1}$ is intractable; $Y_t|X_t$ may be intractable as well. Let

$$X_t' = (X_t, Y_t), \quad Y_t' = Y_t + V_t, \quad V_t \sim \mathcal{N}(0, \sigma^2)$$

and use $\{(X_t', Y_t')\}$ as an approximation of $\{(X_t, Y_t)\}$.
$\Rightarrow$ Connection with ABC (likelihood-free inference).

Finite state-space models (aka hidden Markov models)

$\mathcal{X} = \{1, \ldots, K\}$, uses in e.g.

- speech processing; X_t is a word, Y_t is an acoustic measurement (possibly the earliest application of HMMs). Note K is quite large.
- time-series modelling to deal with heterogenity (e.g. in medecine, X_t is state of patient)
- rediscovered in Economics as Markov-switching models; there X_t is the state of the Economy (recession, growth), and Y_t is some economic indicator (e.g. GDP) which follows an ARMA process (with parameters that depend on X_t)
- also related: change-point models

Finite state-space models (aka hidden Markov models)

$\mathcal{X} = \{1, \ldots, K\}$, uses in e.g.

- speech processing; X_t is a word, Y_t is an acoustic measurement (possibly the earliest application of HMMs). Note K is quite large.
- time-series modelling to deal with heterogenity (e.g. in medecine, X_t is state of patient)
- rediscovered in Economics as Markov-switching models; there X_t is the state of the Economy (recession, growth), and Y_t is some economic indicator (e.g. GDP) which follows an ARMA process (with parameters that depend on X_t)
- also related: change-point models

Note: Not of direct interest to us, as sequential analysis may be performed *exactly* using Baum-Petrie filter.

A quick note on the generality of the definition

Consider a GARCH model, i.e. $Y_t \sim \mathcal{N}(0, \sigma_t^2)$, with

$$\sigma_t^2 = \alpha + \beta Y_{t-1}^2 + \gamma \sigma_{t-1}^2.$$

If we replace $\theta = (\alpha, \beta, \gamma)$ by Markov process (θ_t), do we obtain a state-space model?

Sequential analysis of state-space models

Definition

The phrase *state-space models* refers not only to its definition (in terms of $\{X_t\}$ and $\{Y_t\}$) but also to a particular 'inferential scenario': $\{Y_t\}$ is observed (data denoted $y_0, \ldots$), $\{X_t\}$ is not, and one wishes to recover the X_t's given the Y_t's, often sequentially (over time).

Filtering, prediction, smoothing

Conditional distributions of interest (at every time t)

- Filtering: $X_t|Y_{0:t}$
- Prediction: $X_t|Y_{0:t-1}$
- data prediction: $Y_t|Y_{0:t-1}$
- fixed-lag smoothing: $X_{t-h:t}|Y_{0:t}$ for $h \geq 1$
- complete smoothing: $X_{0:t}|Y_{0:t}$
- likelihood factor: density of $Y_t|Y_{0:t-1}$ (so as to compute the full likelihood)

Parameter estimation

All these tasks are usually performed for a fixed θ (assuming the model depends on some parameter θ). To deal additionally with parameter uncertainty, we could adopt a Bayesian approach, and consider e.g. the law of (θ, X_t) given $Y_{0:t}$ (for filtering). But this is often more involved.

Our notations (spoiler!)

- $\{X_t\}$ is a Markov process with initial law $P_0(\mathrm{d}x_0)$, and Markov kernel $P_t(x_{t-1}, \mathrm{d}x_t)$.
- $\{Y_t\}$ has conditional distribution $F_t(x_t, \mathrm{d}y_t)$, which admits probability density $f_t(y_t|x_t)$ (with respect to common dominating measure $\nu(\mathrm{d}y_t)$).
- when needed, dependence on θ will be made explicit as follows: $P_t^\theta(x_{t-1}, \mathrm{d}x_t)$, $f_t^\theta(y_t|x_t)$, etc.

Our notations (spoiler!)

- $\{X_t\}$ is a Markov process with initial law $P_0(\mathrm{d}x_0)$, and Markov kernel $P_t(x_{t-1}, \mathrm{d}x_t)$.
- $\{Y_t\}$ has conditional distribution $F_t(x_t, \mathrm{d}y_t)$, which admits probability density $f_t(y_t|x_t)$ (with respect to common dominating measure $\nu(\mathrm{d}y_t)$).
- when needed, dependence on θ will be made explicit as follows: $P_t^\theta(x_{t-1}, \mathrm{d}x_t)$, $f_t^\theta(y_t|x_t)$, etc.

Algorithms, calculations, etc may be extended straightforwardly to non-standard situations such that $\mathcal{X}$, $\mathcal{Y}$ vary over time, or such that $Y_t|X_t$ also depends on $Y_{0:t-1}$, but for simplicity, we will stick to these notations.

problems with a structure similar to the sequential analysis of a state-space model

Consider the simulation of Markov process $\{X_t\}$, conditional on $X_t \in A_t$ for each t.

problems with a structure similar to the sequential analysis of a state-space model

Consider the simulation of Markov process $\{X_t\}$, conditional on $X_t \in A_t$ for each t.
Take $Y_t = \mathbb{1}(X_t \in A_t)$, $y_t = 1$, then this tasks amounts to smoothing the corresponding state-space model.

A particular example: self-avoiding random walk

Consider a random walk in $\mathbb{Z}^2$, (i.e. at each time we may move north, south, east or west, wit probability $1/4$). We would to simulate $\{X_t\}$ conditional on the trajectory $X_{0:T}$ never visiting the same point more than once.

A particular example: self-avoiding random walk

Consider a random walk in $\mathbb{Z}^2$, (i.e. at each time we may move north, south, east or west, wit probability $1/4$). We would to simulate $\{X_t\}$ conditional on the trajectory $X_{0:T}$ never visiting the same point more than once.
How to define $\{X_t\}$ in this case?

Markov processes

nicolas.chopin@ensae.fr
(based on a previous PG course with O. Papaspiliopoulos)

Summary

- Introduce Markov processes via kernels
- Recursions of marginal distributions
- Conditional distributions
  - conditional independence
  - partially observed Markov processes & state-space models
- Graphical models

Definition

A probability kernel from $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ to $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, $P(x, \mathrm{d}y')$, is a function from $(\mathcal{X}, \mathcal{B}(\mathcal{Y}))$ to $[0, 1]$ such that

(a) for every x, $P(x, \cdot)$ is a probability measure on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$,

(b) for every $A \in \mathcal{B}(\mathcal{Y})$, $P(x, A)$ is a measurable function in $\mathcal{X}$.

Then, if

$$\mathbb{P}_1(\mathrm{d}x_{0:1}) = \mathbb{P}_0(\mathrm{d}x_0) P_1(x_0, \mathrm{d}x_1)$$

by construction, $\mathbb{P}_1(\mathrm{d}x_0) = \mathbb{P}_0(\mathrm{d}x_0)$ and

$$\mathbb{P}_1(X_1 \in \mathrm{d}x_1 | X_0 = x_0) = P_1(x_0, \mathrm{d}x_1) .$$

Backward kernel - Bayes

$$\mathbb{P}_1(\mathrm{d}x_0)P_1(x_0, \mathrm{d}x_1) = \mathbb{P}_1(\mathrm{d}x_1)\overleftarrow{P}_0(x_1, \mathrm{d}x_0),$$

Definition

A sequence of random variables $X_{0:T}$ with joint distribution given by

$$\mathbb{P}_T(X_{0:T} \in \mathrm{d}x_{0:T}) = \mathbb{P}_0(\mathrm{d}x_0) \prod_{s=1}^{T} P_s(x_{s-1}, \mathrm{d}x_s) \,,$$

is called a (discrete-time) Markov process with state-space $\mathcal{X}$, initial distribution $\mathbb{P}_0$ and transition kernel at time t, P_t. Likewise, a probability measure decomposed into a product of an initial distribution and transition kernels as in (2) will be called a Markov measure.

Conditional independence

$$\mathbb{P}_T(X_t \in \mathrm{d}x_t | X_{0:t-1} = x_{0:t-1}) = P_t(x_{t-1}, \mathrm{d}x_t) \,.$$

$$\mathbb{P}_T(X_t \in \mathrm{d}x_t | X_{0:s} = x_{0:s}) = P_{s+1:t}(x_s, \mathrm{d}x_t) \,, \quad \forall t \leq T, s < t \,,$$

where

$$P_{s+1:t}(x_s, A) = \int_{\mathcal{X}^{t-s}} P_{s+1}(x_s, \mathrm{d}x_{s+1}) P_{s+2}(x_{s+1}, \mathrm{d}x_{s+2}) \cdots P_t(x_{t-1}, A) \,.$$

A marginalisation property

Proposition

Consider a sequence of probability measures, index by t, defined as:

$$\mathbb{P}_t(X_{0:t} \in \mathrm{d}x_{0:t}) = \mathbb{P}_0(\mathrm{d}x_0) \prod_{s=1}^{t} P_s(x_{s-1}, \mathrm{d}x_s)\,,$$

where P_s are probability kernels. Then, for any $t \leq T$,

$$\mathbb{P}_T(\mathrm{d}x_{0:t}) = \mathbb{P}_t(\mathrm{d}x_{0:t})\,.$$

Some recursions

$$\mathbb{P}_t(X_t \in \mathrm{d}x_t) = \mathbb{E}_{\mathbb{P}_t}[\mathbb{P}_t(X_t \in \mathrm{d}x_t | X_{0:s})] = \mathbb{E}_{\mathbb{P}_t}[P_{s+1:t}(X_s, \mathrm{d}x_t)] \,.$$

With the marginalisation it yields the Chapman-Kolmogorov equation

$$\mathbb{P}_t(X_t \in \mathrm{d}x_t) = \mathbb{E}_{\mathbb{P}_s}[P_{s+1:t}(X_s, \mathrm{d}x_t)] \,, \quad \forall s \leq t-1 \,.$$

Backward process

$$\mathbb{P}_T(X_{0:T} \in \mathrm{d}x_{0:T}) = \mathbb{P}_T(\mathrm{d}x_T) \prod_{s=1}^{T} \overleftarrow{P}_{T-s}(x_{T-s+1}, \mathrm{d}x_{T-s}),$$

POMP & SSM

$$\mathbb{P}_T(X_{0:T} \in \mathrm{d}x_{0:T}, Y_{0:T} \in \mathrm{d}y_{0:T}) = \mathbb{P}_T(\mathrm{d}x_{0:T}) \prod_{t=0}^{T} f_t(y_t|x_t) \prod_{t=0}^{T} \nu(\mathrm{d}y_t)$$

$$= \mathbb{P}_0(\mathrm{d}x_0) \prod_{t=1}^{T} P_t(x_{t-1}, \mathrm{d}x_t) \prod_{t=0}^{T} f_t(y_t|x_t) \prod_{t=0}^{T} \nu(\mathrm{d}y_t)$$

When relevant, $f_t^\theta(y_t|x_t)$ and $P_t^\theta(x_{t-1}, \mathrm{d}x_t)$

Components of a SSM

Likelihood

$$p_t(y_{0:t}) = \mathbb{E}_{\mathbb{P}_t}\left[\prod_{s=0}^{t} f_s(y_s|x_s)\right]$$

is the density (likelihood/partition function) of the law of $Y_{0:T}$; Likelihood factors

$$p_t(y_{0:t}) = p_0(y_0)\prod_{s=1}^{t} p_s(y_s|y_{0:s-1})\,.$$

and

$$p_{t+k}(y_{t:t+k}|y_{0:t-1}) = p_{t+k}(y_{0:t+k})/p_{t-1}(y_{0:t-1})\,, \quad k \geq 0\,,\ t \geq 1\,.$$

Law of states given observations

$$\mathbb{P}_t(X_{0:t} \in \mathrm{d}x_{0:t} | Y_{0:t} = y_{0:t}) = \frac{1}{p_t(y_{0:t})} \left\{ \prod_{s=0}^{t} f_s(y_s|x_s) \right\} \mathbb{P}_t(\mathrm{d}x_{0:t}) \, .$$

(To see this, multiply both sides by $p_t(y_{0:t}) \prod_{s=0}^{t} \nu(\mathrm{d}y_s)$)
Another SSM function that will be is likelihood of future observations given the current value of the state.

$$p_T(y_{t+1:T}|x_t) = \frac{\mathbb{P}_T(Y_{t+1:T} \in \mathrm{d}y_{t+1:T}|X_{0:t} = x_{0:t}, Y_t = y_t)}{\nu^{T-t}(\mathrm{d}y_{t+1:T})} \, , t < T \, ,$$

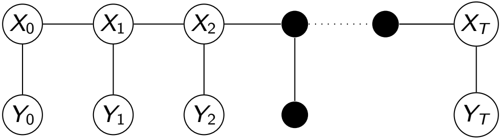
where by conditional independence it does not depend on $x_{0:t-1}, y_t$

Restating SSM aims

- **state prediction**: deriving $\mathbb{P}_t(X_{t+1:t+h} \in \mathrm{d}x_{t+1:t+h} | Y_{0:t} = y_{0:t})$, for $h \geq 1$;
- **filtering**: deriving $\mathbb{P}_t(X_t \in \mathrm{d}x_t | Y_{0:t} = y_{0:t})$;
- **fixed-lag smoothing**: deriving $\mathbb{P}_t(X_{t-l:t} \in \mathrm{d}x_{t-l:t} | Y_{0:t} = y_{0:t})$ for some $l \geq 1$;
- **(complete) smoothing**: deriving $\mathbb{P}_t(X_{0:t} \in \mathrm{d}x_{0:t} | Y_{0:t} = y_{0:t})$;
- **likelihood computation**: deriving $p_t(y_{0:t})$.

Graphical models

Variables as nodes; when any two are linked by a kernel draw an edge:



Path; conditional independence; Markov property of $(X_{0;T}, Y_{0:T})$, $X_{0:T}$, $X_{0:T}$ conditionally on $Y_{0:T}$ but not of $Y_{0:T}$.

Further reading

- Conditional independence, Chapter 5 of *Foundations of modern Probability* (Kallenberg, Springer)
- Intro to graphical models: Chapter 8 of *Pattern recognition and machine learning* (Bishop, Springer)