Advanced Simulation - Lecture 12

Patrick Rebeschini

February 21st, 2018

Outline

- Hidden Markov models, also called state space models.
- Various examples.
- Inference leads to high-dimensional integrals.

Time series

- Observations $(y_t)_{t\geq 1}$ assumed to be dependent, usually specified by an initial distribution: $Y_1 \sim \mu_\theta$, and a conditional distribution:

$$Y_t \mid Y_{1:t-1} = y_{1:t-1} \sim p_\theta(\cdot \mid y_1, \ldots, y_{t-1}),$$

  where we use the notation $y_{k:l} = (y_k, \ldots, y_l)$.

- The likelihood is given by

$$\forall \theta \in \Theta \quad \mathcal{L}(\theta; y_1, \ldots, y_t) = \mu_\theta(y_1) \prod_{s=2}^{t} p_\theta(y_s \mid y_1, \ldots, y_{s-1}).$$

- Put a prior on θ and consider the problem of sampling from the posterior given $y_{1:t}$.

Time series

- In simple cases, the likelihood can be computed point-wise.
- Example: Bayesian analysis of a Markov chain, in Chapter 3 of the lecture notes.
- ARCH(1) model: $y_1 \sim \mathcal{N}(0,1)$ and for all $t \geq 2$,

$$\begin{aligned} y_t &= \varepsilon_t \left(h_t\right)^{1/2}, \\ \varepsilon_t &\sim \mathcal{N}(0,1), \\ h_t &= \alpha_0 + \alpha_1 y_{t-1}^2. \end{aligned}$$

- In this case we can implement a Metropolis–Hastings algorithm to sample from $\pi(\theta \mid y_{1:t})$, for each t.
- Or importance sampling to obtain estimates at each intermediate time $1 \leq s \leq t$.

Hidden Markov Models

- We introduce $(X_t)_{t\geq 1}$ a latent/hidden/unobserved $\mathbb{X}$-valued Markov process defined by its initial density μ_θ

$$X_1 \sim \mu_\theta(\cdot),$$

and its homogeneous Markov transition kernel f_θ

$$X_t | X_{t-1} = x_{t-1} \sim f_\theta(\cdot | x_{t-1}).$$

- Sometimes we note $X_0 \sim \mu_\theta$.
- Hence the law of the path/trajectory $X_{1:t}$ is given by

$$\begin{aligned} p_{X_{1:t}}(x_{1:t}) &= p_{X_1}(x_1) \prod_{k=2}^t p_{X_k | X_{1:k-1}}(x_k | x_{1:k-1}) \quad \text{(chain rule)} \\ &= p_{X_1}(x_1) \prod_{k=2}^t p_{X_k | X_{k-1}}(x_k | x_{k-1}) \quad \text{(Markov)} \\ &= \mu_\theta(x_1) \prod_{k=2}^t f_\theta(x_k | x_{k-1}). \end{aligned}$$

Hidden Markov Models

- The $\mathbb{Y}$-valued observations $(Y_t)_{t\geq 1}$ are assumed to be independent conditional upon $(X_t)_{t\geq 1}$ and their conditional distribution satisfy

$$Y_t | \, X_t = x_t \sim g_\theta \left(\cdot | \, x_t \right),$$

  i.e. the distribution of Y_t is independent of $(X_k)_{k\neq t}$ conditional upon $X_t = x_t$.

- Hence we have the law of observations given the hidden process,

$$\begin{aligned} & p_{Y_{1:t}|(X_l)_{l\geq 1}} \left(y_{1:t} | \, (x_l)_{l\geq 1} \right) \\ = & \prod\nolimits_{k=1}^{t} p_{Y_k|X_k} \left(y_k | \, x_k \right) \text{ (cond. independent)} \\ = & \prod\nolimits_{k=1}^{t} g_\theta \left(y_k | \, x_k \right). \end{aligned}$$

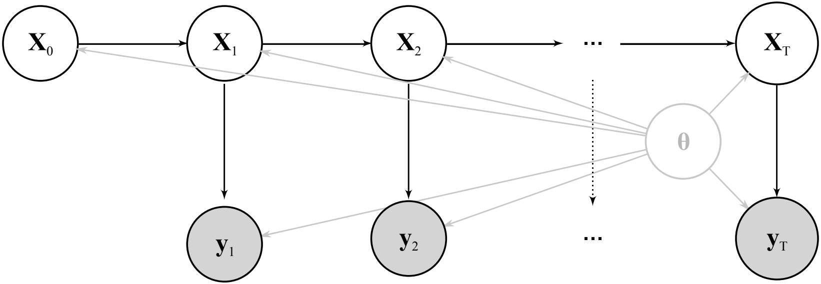


Figure: Graph representation of a general HMM.

(X_t): initial distribution μ_θ, transition f_θ.
(Y_t) given (X_t): measurement g_θ.
Prior on the parameter $\theta \in \Theta$.

Inference in HMMs, Cappé, Moulines, Ryden, 2005.

Example: S&P 500 index

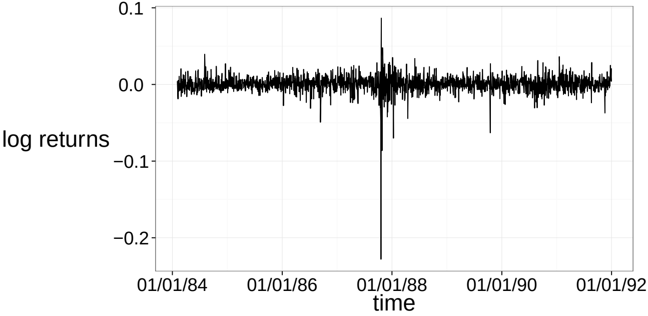


Figure: Daily returns $y_t = \log(p_t/p_{t-1})$ between 1984 and 1991.

Example: stochastic volatility model

- Latent stochastic volatility $(X_t)_{t\geq 1}$ of an asset is modeled through

$$X_t = \varphi X_{t-1} + \sigma V_t,\ Y_t = \beta \exp\left(X_t\right) W_t$$

where $V_t, W_t \sim \mathcal{N}\left(0,1\right).$

- Popular alternative to ARCH and GARCH models (Engle, 2003 Nobel Prize).

Example: battery voltage

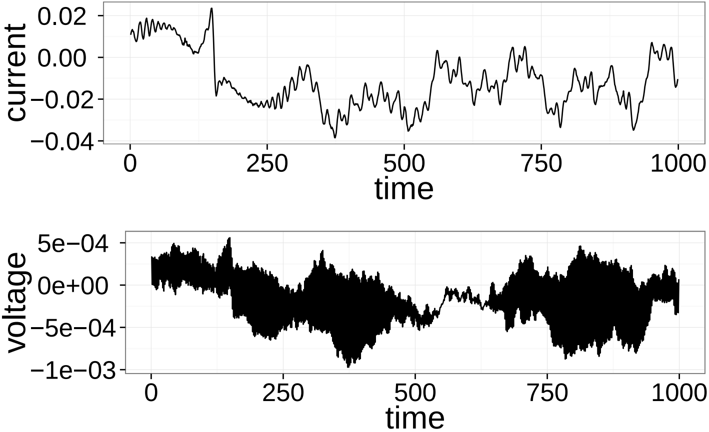


Figure: Current (input) and measured voltage (output) of a battery.

Example: phytoplankton – zooplankton

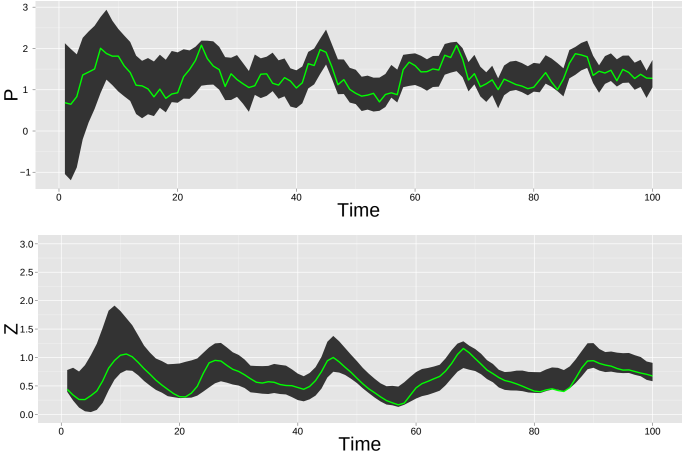


Figure: Filtering of the latent variables (top: P, bottom: Z).

Example: athletic records

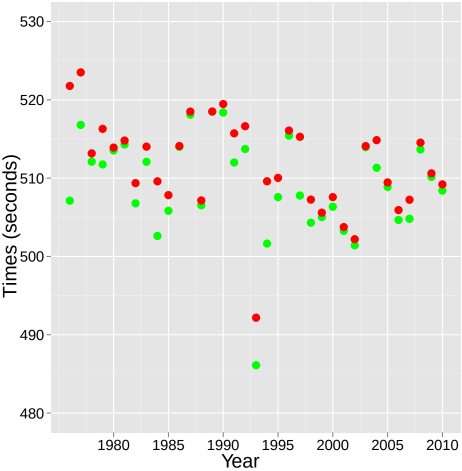
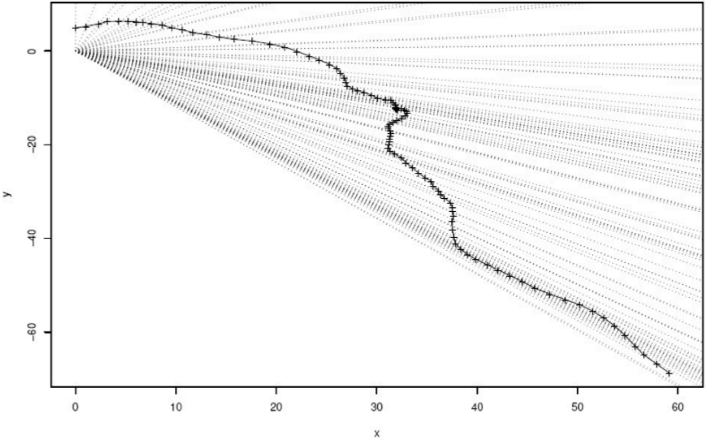


Figure: Best two times of each year in women's 3000m events.

Example: tracking



Example: tracking

- Markov model describing dynamic of the target

$$\begin{pmatrix} X_t^1 \\ \dot{X}_t^1 \\ X_t^2 \\ \dot{X}_t^2 \end{pmatrix} = \begin{pmatrix} 1 & \delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \delta \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_{t-1}^1 \\ \dot{X}_{t-1}^1 \\ X_{t-1}^2 \\ \dot{X}_{t-1}^2 \end{pmatrix} + V_t,\ V_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \Sigma_v\right),$$

- Measurements provided by the radar

$$Y_t = \tan^{-1}\left(\frac{X_t^1}{X_t^2}\right) + W_t,\ W_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right).$$

Finite State-Space HMM

- *Automatic speech recognition*: Y_t is the speech signal, X_t represents the word that is being spoken.
- *Activity recognition*: Y_t represents a video frame, X_t is the class of activity the person is engaged in (e.g., running, walking, sitting, etc.)
- *Part of speech tagging*: Y_t represents a word, X_t represents its part of speech (noun, verb, adjective, etc.)
- *Gene finding*: Y_t represents the DNA nucleotides (A,C,G,T), X_t represents whether we are inside a gene-coding region or not.

Specific algorithms allow to estimate $X_{1:t}$ given $y_{1:t}$ and to evaluate the likelihood of the parameters: Viterbi, forward-backward, Baum–Welch.

Linear Gaussian models

- Consider $\mathbb{X} = \mathbb{R}^{d_x}$ and $\mathbb{Y} = \mathbb{R}^{d_y}$. Let X_t be defined by

$$X_t = AX_{t-1} + \varepsilon_t$$

for $\varepsilon \sim \mathcal{N}(0, \Sigma_x)$, and some matrices A and Σ_x.

- Let the observations be defined by

$$Y_t = CX_t + \Sigma_y \eta_t$$

for $\eta \sim \mathcal{N}(0, \Sigma_y)$, and some matrices C and Σ_y.

- Then the distribution of $X_{1:t}$ given $Y_{1:t}$ can be retrieved by "Kalman recursions".
- The likelihood of the parameters $(A, C, \Sigma_x, \Sigma_y)$ can be evaluated exactly, which allows parameter estimation using standard techniques.

General inference in HMM

- Given $Y_{1:t} = y_{1:t}$ and θ, inference on $X_{1:t}$ relies on

$$p\left(x_{1:t} \middle| y_{1:t}, \theta\right) = \frac{p\left(x_{1:t}, y_{1:t} \mid \theta\right)}{p\left(y_{1:t} \mid \theta\right)}$$

where

$$p\left(x_{1:t}, y_{1:t} \mid \theta\right) = p\left(x_{1:t} \mid \theta\right) p\left(y_{1:t} \middle| x_{1:t}, \theta\right)$$

with

$$p\left(x_{1:t} \mid \theta\right) = \mu_\theta\left(x_1\right) \prod_{k=2}^{t} f_\theta\left(x_k \middle| x_{k-1}\right)$$

$$p\left(y_{1:t} \mid x_{1:t}, \theta\right) = \prod_{k=1}^{t} g_\theta\left(y_k \mid x_k\right).$$

- The (marginal) likelihood is given by

$$p\left(y_{1:t} \mid \theta\right) = \int_{\mathbb{X}^t} p\left(x_{1:t}, y_{1:t} \mid \theta\right) dx_{1:t}$$

General inference in HMM

- **Proposition**: The posterior $p\left(x_{1:t} \mid y_{1:t}, \theta\right)$ satisfies

$$p\left(x_{1:t} \mid y_{1:t}, \theta\right) = p\left(x_{1:t-1} \mid y_{1:t-1}, \theta\right) \frac{f_\theta\left(x_t \mid x_{t-1}\right) g_\theta\left(y_t \mid x_t\right)}{p\left(y_t \mid y_{1:t-1}, \theta\right)}$$

where

$$p\left(y_t \mid y_{1:t-1}, \theta\right) = \int p\left(x_{1:t-1} \mid y_{1:t-1}, \theta\right) f_\theta\left(x_t \mid x_{t-1}\right) g_\theta\left(y_t \mid x_t\right) dx_{1:t}.$$

- *Proof*. Dropping the parameter θ, we have

$$\begin{aligned} p\left(x_{1:t}, y_{1:t}\right) &= p\left(x_{1:t-1}, y_{1:t-1}\right) f\left(x_t \mid x_{t-1}\right) g\left(y_t \mid x_t\right) \\ p\left(y_{1:t}\right) &= p\left(y_{1:t-1}\right) p\left(y_t \mid y_{1:t-1}\right) \end{aligned}$$

so

$$p\left(x_{1:t} \mid y_{1:t}\right) = \frac{p\left(x_{1:t}, y_{1:t}\right)}{p\left(y_{1:t}\right)} = \underbrace{\frac{p\left(x_{1:t-1}, y_{1:t-1}\right)}{p\left(y_{1:t-1}\right)}}_{p\left(x_{1:t-1} \mid y_{1:t-1}\right)} \frac{f\left(x_t \mid x_{t-1}\right) g\left(y_t \mid x_t\right)}{p\left(y_t \mid y_{1:t-1}\right)}$$

and the expression for $p\left(y_t \mid y_{1:t-1}\right)$ follows.

General inference in HMM

- **Proposition**: The marginal posterior $p\left(x_t \middle| y_{1:t}\right)$ satisfies the following recursion

$$p\left(x_t \middle| y_{1:t-1}\right) = \int f\left(x_t \middle| x_{t-1}\right) p\left(x_{t-1} \middle| y_{1:t-1}\right) dx_{t-1}$$

$$p\left(x_t \middle| y_{1:t}\right) = \frac{g\left(y_t \middle| x_t\right) p\left(x_t \middle| y_{1:t-1}\right)}{p\left(y_t \middle| y_{1:t-1}\right)}$$

  where

$$p\left(y_t \middle| y_{1:t-1}\right) = \int g\left(y_t \middle| x_t\right) p\left(x_t \middle| y_{1:t-1}\right) dx_t.$$

- This recursion can be implemented exactly for finite state-space HMM and linear Gaussian models.
- Otherwise these integrals are intractable.

General inference in HMM

- In general, the filtering problem is thus intractable:

$$\int \varphi(x_t) p(x_t \mid y_{1:t}, \theta) dx_t = \int \varphi(x_t) p(x_{1:t}, y_{1:t} \mid \theta) dx_{1:t}$$

$$= \int \varphi(x_t) \mu_\theta(x_1) \prod_{s=1}^{t} f_\theta(x_s \mid x_{s-1}) \prod_{s=1}^{t} g_\theta(y_s \mid x_s) dx_{1:t}$$

- It is a $t \times \dim(\mathbb{X})$ dimensional integral.
- The likelihood is also intractable:

$$p(y_{1:t} \mid \theta) = \int p(x_{1:t}, y_{1:t} \mid \theta) dx_{1:t}$$

$$= \int \mu_\theta(x_1) \prod_{s=1}^{t} f_\theta(x_s \mid x_{s-1}) \prod_{s=1}^{t} g_\theta(y_s \mid x_s) dx_{1:t}$$

- Thus we cannot compute it pointwise, e.g. to perform Metropolis–Hastings algorithm on the parameter space.

General inference in HMM

- The historical approach consists in performing Gibbs sampling on the joint space of θ and $X_{1:t}$.

- Alternate between sampling from $\theta \mid x_{1:t}, y_{1:t}$, with conditional distribution

$$\begin{aligned} p(\theta \mid x_{1:t}, y_{1:t}) &\propto p(\theta) p(x_{1:t}, y_{1:t} \mid \theta) \\ &= p(\theta) \mu_\theta(x_1) \prod_{s=1}^{t} f_\theta(x_s \mid x_{s-1}) \prod_{s=1}^{t} g_\theta(y_s \mid x_s) \end{aligned}$$

  which can (perhaps) be evaluated pointwise.

- And sampling from $p(x_{1:t} \mid y_{1:t}, \theta)$. How?

General inference in HMM

- Sampling from $p(x_{1:t} \mid y_{1:t}, \theta)$ can be done by iteratively sampling x_k given x_{k-1}, y_k, x_{k+1} and θ.
- Indeed

$$\begin{aligned} p(x_k \mid x_{-k}, y_{1:t}, \theta) &= p(x_k \mid x_{k-1}, y_k, x_{k+1}, \theta) \\ &\propto p(x_k \mid x_{k-1}, \theta) p(y_k, x_{k+1} \mid x_k, \theta) \\ &= f_\theta(x_k \mid x_{k-1}) f_\theta(x_{k+1} \mid x_k) g_\theta(y_k \mid x_k) \end{aligned}$$

  and (perhaps) we can evaluate this density point-wise.
- In which case, we could use Metropolis–Hastings to update each component of $X_{1:t}$ given the others.
- By definition, the components of $X_{1:t}$ are highly correlated, thus this Gibbs sampling approach will fail (remember the bivariate normal!).

Hidden Markov Models

- Usually, batch estimation of $p(\theta, x_{1:T} \mid y_{1:T})$ using MCMC performs poorly because of high correlations between components.
- Filtering given a fixed θ, i.e. approximating $p(x_{1:t} \mid y_{1:t}, \theta)$, can be efficiently performed using particle filters.
- We'll see that particle filters also provide estimators of the likelihood, which can be used to estimate θ.
- Particle Markov chain Monte Carlo for batch estimation of $p(\theta, x_{1:T} \mid y_{1:T})$ (Andrieu, Doucet, Holenstein, 2010).

Objects of practical interest

Various by-products of the joint posterior $p(\theta, x_{0:t} \mid y_{1:t})$:

- prediction under parameter uncertainty through $p(x_{t+1} \mid y_{1:t})$,

$$p(x_{t+1} \mid y_{1:t}) = \int_{\Theta} \int_{\mathcal{X}^{t+1}} \underbrace{f_\theta(x_{t+1} \mid x_t)}_{\text{transition}} \underbrace{p(d\theta, dx_{0:t} \mid y_{1:t})}_{\text{joint posterior}}.$$

- predictive checking through $\mathbb{P}(Y_{t+1} \leq y_{t+1} \mid y_{1:t})$,

$$\mathbb{P}(Y_{t+1} < y_{t+1} \mid y_{1:t}) = \int_{\Theta} \int_{\mathcal{X}} \int_{-\infty}^{y_{t+1}} g_\theta(dy \mid x_{t+1}) p(dx_{t+1}, d\theta \mid y_{1:t}).$$

- sequential model comparison through $p(y_{1:t})$.

$$p(y_{1:t}) = \int_{\Theta} \int_{\mathcal{X}^{t+1}} p(d\theta, dx_{0:t}, y_{1:t})$$

Sequential Importance Sampling

- We now consider the parameter θ to be fixed. We want to infer $X_{1:t}$ given $y_{1:t}$.

- Let us consider the problem of approximating the first filtering distribution $p(x_1 \mid y_1)$:

$$p(x_1 \mid y_1) = \frac{\mu(x_1)g(y_1 \mid x_1)}{\int_{\mathbb{X}} \mu(x_1)g(y_1 \mid x_1)dx_1} \propto \mu(x_1)g(y_1 \mid x_1).$$

- Drawing $X_1^{1:N}$ from q_1,

$$\forall i \in \{1, \ldots, N\} \quad w_1^i = \frac{\mu(X_1^i)g(y_1 \mid X_1^i)}{q_1\left(X_1^i\right)}.$$

Sequential Importance Sampling

- Empirical distribution $\pi_1^N(x_1)$ approximates $p(x_1 \mid y_1)$:

$$\pi_1^N(x_1) = \frac{\sum_{i=1}^N w_1^i \delta_{X_1^i}(x_1)}{\sum_{j=1}^N w_1^j} = \sum_{i=1}^N W_1^i \delta_{X_1^i}(x_1),$$

  in the sense

$$I^N(\varphi_1) = \int \varphi_1(x)\pi_1^N(x_1)dx_1 = \sum_{i=1}^N W_1^i \varphi_1(X_1^i)$$

$$\xrightarrow[N\to\infty]{a.s.} \int \varphi_1(x) p(x_1 \mid y_1) dx.$$

- Marginal likelihood estimator:

$$p^N(y_1) = \frac{1}{N}\sum_{i=1}^N w_1^i \xrightarrow[N\to\infty]{a.s.} \int \frac{\mu(x_1) g(y_1 \mid x_1)}{q_1(x_1)} q_1(x_1)\, dx_1 = p(y_1).$$

Sequential Importance Sampling

- How to approximate $p(x_{1:2} \mid y_{1:2})$, $p(x_2 \mid y_{1:2})$ and $p(y_{1:2})$?
- At step $t-1$, assume N trajectories $X^i_{1:t-1}$ sampled from q_{t-1} and with weights

$$w^i_{t-1} \propto p(X^i_{1:t-1} \mid y_{1:t-1})/q_{t-1}(X^i_{1:t-1}).$$

- Introduce proposal $q_{t|t-1}(x_t \mid x_{t-1})$. For each i,

$$X^i_t \sim q_{t|t-1}(x_t \mid X^i_{t-1}),$$

then $X^i_{1:t} = (X^i_{1:t-1}, X^i_t)$ follows

$$q_t(x_{1:t}) := q_{t-1}(x_{1:t-1})q_{t|t-1}(x_t \mid x_{t-1})$$

- The correct importance weight is

$$w(x_{1:t}) \propto \frac{p(x_{1:t} \mid y_{1:t})}{q_t(x_{1:t})} \propto \frac{p\left(x_{1:t-1} \middle| y_{1:t-1}\right) f\left(x_t \middle| x_{t-1}\right) g\left(y_t \middle| x_t\right)}{q_{t-1}(x_{1:t-1})q_{t|t-1}(x_t \mid x_{t-1})}$$

$$\propto w(x_{1:t-1}) \frac{f\left(x_t \middle| x_{t-1}\right) g\left(y_t \middle| x_t\right)}{q_{t|t-1}(x_t \mid x_{t-1})}.$$

Sequential Importance Sampling

- Thus the incremental weights are

$$\omega_t^i := \omega_t\left(X_{t-1}^i, X_t^i\right) := \frac{f\left(X_t^i \middle| X_{t-1}^i\right) g\left(y_t \middle| X_t^i\right)}{q_{t|t-1}(X_t^i \mid X_{t-1}^i)}.$$

- The new weights are obtained by the update

$$w_t^i = w_{t-1}^i \times \omega_t^i.$$

- The new "particles" are thus $(w_t^i, X_{1:t}^i)$ for $i \in \{1, \dots, N\}$.

Sequential Importance Sampling

- Then the "particle approximation" π_t^N of $p(x_{1:t} \mid y_{1:t})$ is consistent, in the sense that for any test function φ_t on $\mathbb{X}^t$,

$$I^N(\varphi_t) = \int \varphi_t(x_{1:t})\pi_t^N(x_{1:t})dx_{1:t} = \frac{\sum_{i=1}^N w_t^i \varphi_t(X_{1:t}^i)}{\sum_{i=1}^N w_t^i}$$
$$\xrightarrow[N\to\infty]{a.s.} \int \varphi_t(x_{1:t})p(x_{1:t} \mid y_{1:t})dx_{1:t}.$$

- The incremental likelihood can be approximated too:

$$p^N(y_t \mid y_{1:t-1}) = \frac{\sum_{i=1}^N w_{t-1}^i \omega_t^i}{\sum_{i=1}^N w_{t-1}^i} \xrightarrow[N\to\infty]{a.s.} p(y_t \mid y_{1:t-1}),$$

by a standard importance sampling argument.

Sequential Importance Sampling: algorithm

- *At time* $t = 1$
  - Sample $X_1^i \sim q_1(\cdot)$.
  - Compute the weights

$$w_1^i = \frac{\mu(X_1^i) g(y_1 \mid X_1^i)}{q_1\left(X_1^i\right)}.$$

- *At time* $t \geq 2$
  - Sample $X_t^i \sim q_{t|t-1}(\cdot \mid X_{t-1}^i)$.
  - Compute the weights

$$\begin{aligned} w_t^i &= w_{t-1}^i \times \omega_t^i \\ &= w_{t-1}^i \times \frac{f\left(X_t^i \middle| X_{t-1}^i\right) g\left(y_t \middle| X_t^i\right)}{q_{t|t-1}(X_t^i \mid X_{t-1}^i)}. \end{aligned}$$

Sequential Importance Sampling: diagnostics

- As already seen for IS, we can compute the effective sample size

$$\mathrm{ESS}_t = \frac{\left(\sum_{i=1}^{N} w_t^i\right)^2}{\left(\sum_{i=1}^{N} (w_t^i)^2\right)} = \frac{1}{\sum_{i=1}^{n} \left(W_t^i\right)^2}.$$

- $\mathrm{ESS}_t = N$ if $W_t^i = N^{-1}$ for all i.

- If there exists i such that $W_t^i \approx 1$, and for $j \neq i$, $W_t^j \approx 0$, then $\mathrm{ESS}_t \approx 1$.

- As a rule of thumb, the higher the ESS the better our approximation.

Sequential Importance Sampling: prior proposal

- Default choice of proposal:

$$q_1(x_1) = \mu(x_1),$$
$$q_{t|t-1}(x_t \mid x_{t-1}) = f(x_t \mid x_{t-1}).$$

- Then the incremental weight takes the form

$$\omega(x_{t-1}, x_t) = g(y_t \mid x_t).$$

- This proposal blindly propagates x_{t-1} to x_t without taking y_t into account.

- We can implement SIS as soon as we can sample from the hidden process $(X_t)_{t\geq 1}$ and evaluate $g(y \mid x)$ pointwise.

Sequential Importance Sampling: optimal proposals

- Proposal $q_{t|t-1}\left(x_t|\, x_{t-1}\right)$ that minimizes the variance of $(\omega_t^i)_{i=1}^N$.

- Turns out to be

$$q_{t|t-1}^{\text{opt}}\left(x_t|\, x_{t-1}\right) = \frac{f\left(x_t|\, x_{t-1}\right) g\left(y_t|\, x_t\right)}{p\left(y_t|\, x_{t-1}\right)}.$$

- This uses the observation y_t to guide the propagation of x_t.

- Associated incremental weight:

$$\omega_t^{\text{opt}}\left(x_{t-1}, x_t\right) = p\left(y_t|\, x_{t-1}\right),$$

does not depend on x_t.

Sequential Importance Sampling: example

- Model equations:

$$\forall t \geq 1 \quad X_t = \varphi X_{t-1} + \sigma_V V_t,$$
$$\forall t \geq 1 \quad Y_t = X_t + \sigma_V W_t,$$

  with $X_0 \sim \mathcal{N}\left(0, \sigma_V^2\right)$, $V_t, W_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $\varphi = 0.95$, $\sigma_V = 1$, $\sigma_W = 1$.

- Synthetic dataset of size $T = 100$.

- We can compute the filtering quantities using Kalman filters.

- We want to estimate them using SIS, with $N = 1000$ particles, using the prior proposal or the optimal proposal.

Sequential Importance Sampling: example

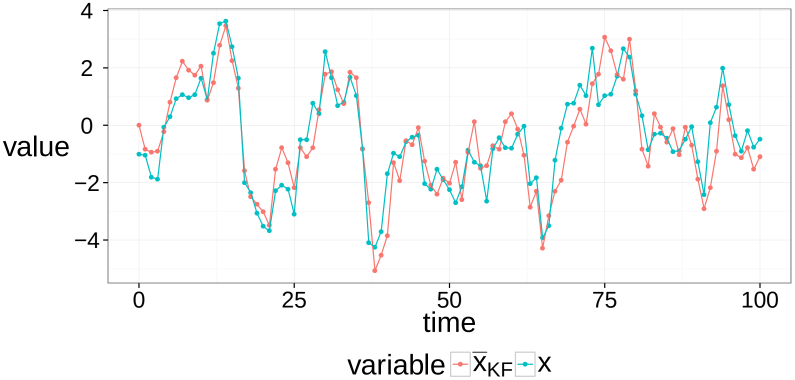


Figure: Generated hidden process $(X_t)_{t\geq 1}$, along with the filtering mean calculated using the Kalman filter, $(\bar{X}_{KF})$.

Sequential Importance Sampling: example

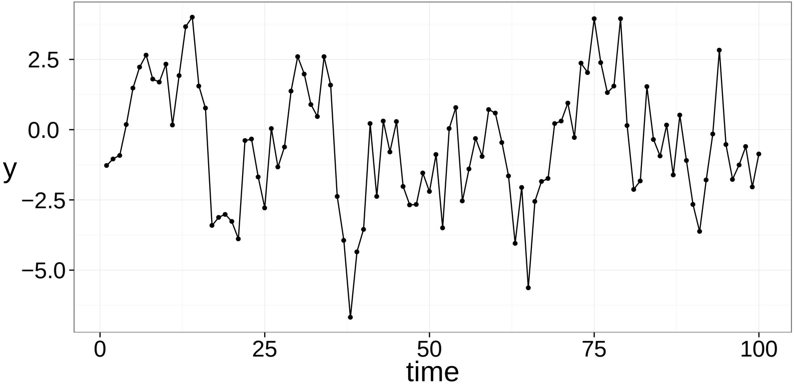


Figure: Generated observations $(Y_t)_{t \geq 1}$.

Sequential Importance Sampling: example

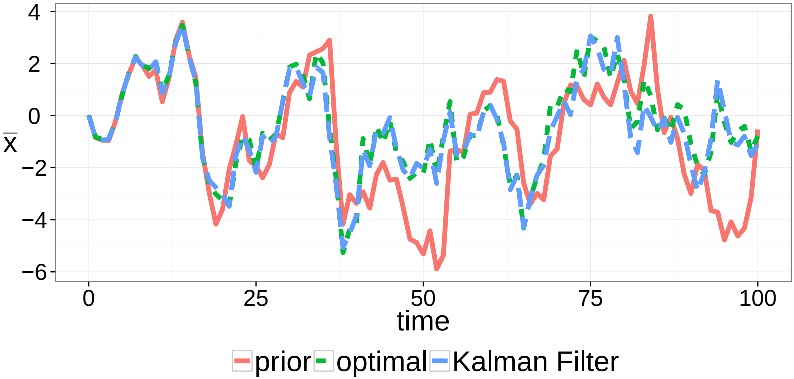


Figure: Estimation of filtering means $\mathbb{E}\left(x_t \mid y_{1:t}\right)$.

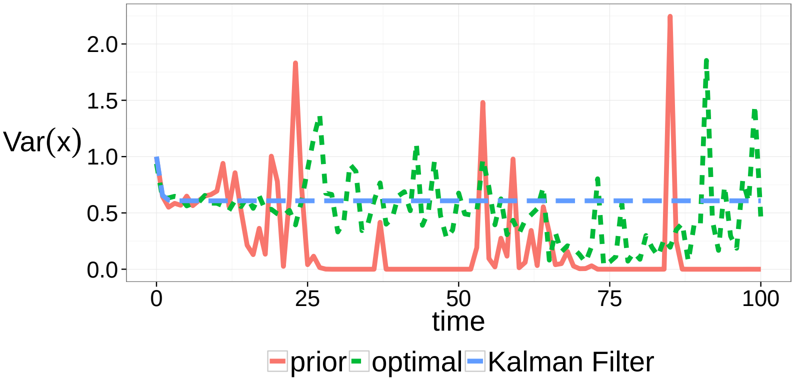


Figure: Estimation of filtering variances $\mathbb{V}\left(x_t \mid y_{1:t}\right)$.

Sequential Importance Sampling: example

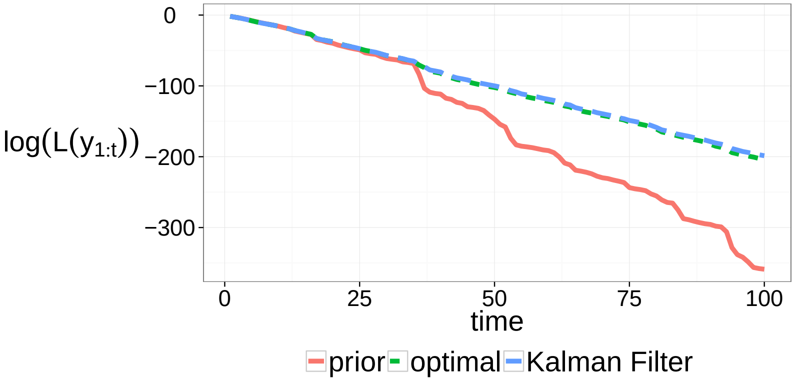


Figure: Estimation of marginal log likelihoods $\log p(y_{1:t})$.

Sequential Importance Sampling: example

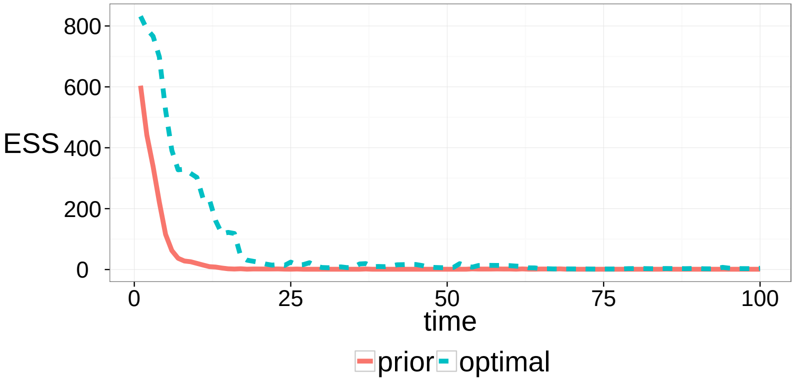


Figure: Effective sample size over time.

Sequential Importance Sampling: example

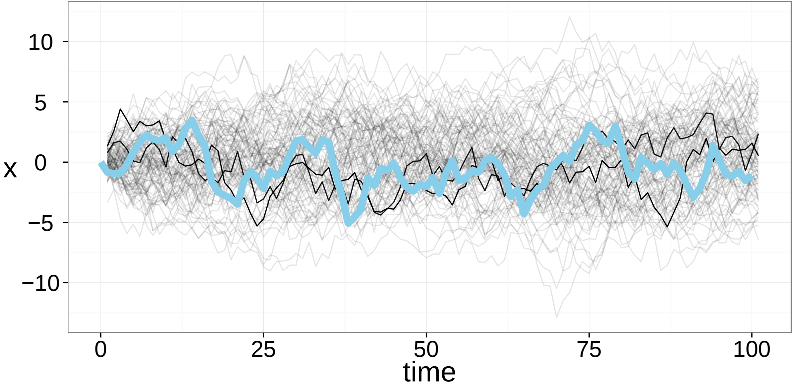


Figure: Spread of 100 paths drawn from the prior proposal, and KF means in blue. Darker lines indicate higher weights.