

Particle Approximation for Optimal Forecasting in State-Space Models

Paul Bui Quang

2020

Abstract

State-space models are statistical models for time series where observations are generated from a latent state process. Particle filtering is a well established sequential Monte Carlo algorithm to estimate the state given the observations in state-space models. We propose an extension of particle filtering, called particle forecasting, to estimate future data points given available observations. Particle forecasting computes forecasts in the form of empirical distributions. We review convergence results (almost surely and in mean squared error) of particle filtering and we obtain similar results for particle forecasting. Simulation experiments illustrate the convergence analysis. In the simulations, we use probability integral transform to assess the performance of distributional forecasts.

1 Introduction

1.1 Context and Problem Statement

Forecasting is an important task in statistics and machine learning, consisting in predicting future values of an observed time series. Many forecasting techniques are based on statistical modelling, involving models such as ARIMA, GARCH, or regression models [1].

State-space models are flexible statistical models where observations are generated from a dynamic latent state. They arise in many scientific fields, especially econometrics [2] and engineering [3, 4]. Inference problems in state-space models studied in the literature concern estimating the hidden state over time (filtering or smoothing), or estimating the static parameter identifying the model [5, 6, 7]. Although they describe time series, forecasting in state-space models has not attracted much attention, outside the case where the observation process and the state process are linear with additive Gaussian noise [8].

We propose in this paper an algorithm to compute forecasts in general state-space models. This algorithm, called particle forecasting, is an extension of particle filtering (or sequential Monte Carlo method), which is a well established technique for state estimation in these models, based on importance sampling

[9, 10, 11]. In particular, particle forecasting yields forecasts in the form of distributions rather than point values, which allows to assess prediction uncertainty [12].

In Section 1.2 below, we describe linear Gaussian state-space models, also called dynamic linear models. In Section 2, we define the model setup and the two inference problems we focus on: Bayesian filtering and optimal forecasting. The particle filter for Bayesian filtering and the particle forecast for optimal forecasting are presented in Section 3. In Section 4, we review convergence results of particle filtering and show the convergence of particle forecasting (almost surely and in mean squared error), using the same methodology as Crisan and Doucet in [13], based on the description of the algorithms with measure-valued operators. Numerical experiments are presented in Section 5, where the quality of distribution forecasts is assessed with test-based procedures.

Throughout the article, the following notations are used:

- $a_{0:t}$: sequence $\{a_0, a_1, \dots, a_t\}$,
- δ_a : Dirac measure centered at a ,
- $\int \phi \nu$: integral of ϕ w.r.t. the measure ν ,
- $\mathcal{N}(m, \sigma^2)$: Gaussian distribution with mean m and variance σ^2 ,
- $\mathcal{U}(a, b)$: uniform distribution over the real segment $[a, b]$.

1.2 Dynamic Linear Models

A dynamic linear model is a state-space model where the hidden (unobserved) state obeys to the linear discrete-time dynamics

$$X_t = F_t X_{t-1} + V_t$$

for $t \geq 1$ and $X_0 \sim \mathcal{N}(m_0, Q_0)$, where $\{V_t\}_t$ is a Gaussian white noise, and the observation process is also linear as

$$Y_t = H_t X_t + W_t$$

for $t \geq 0$, where $\{W_t\}_t$ is a Gaussian white noise.

In a dynamic linear model, filtering, smoothing and forecasting are optimally solved by the Kalman filter and its extensions [14, 8]. We consider in this paper general state-space models, for example when the linear functions mapping X_{t-1} to X_t and X_t to Y_t above are replaced by nonlinear mappings. placed by nonlinear mappings.

2 Bayesian Filtering and Optimal Forecasting

2.1 Model Setup

General state-space models, also called hidden Markov models, are stochastic processes composed of an unobserved process $\{X_t\}_t$, called the state process, and an observed process $\{Y_t\}_t$, called the observation process, where $\{X_t, Y_t\}_t$ is a discrete-time time Markov process such that

$$P[X_t \in dx_t, Y_t \in dy_t \mid X_{0:t-1} = x_{0:t-1}, Y_{0:t} = y_{0:t}] = Q_t(x_{t-1}, dx_t) g_t(y_t \mid x_t) \lambda(dy_t)$$

for all $t \geq 1$ and $P[X_0 \in dx_0, Y_0 \in dy_0] = Q_0(dx_0) g_0(y_0 \mid x_0) \lambda(dy_0)$ [5, Chapter 1]. The state process takes values in \mathbb{R}^d equipped with its Borelian set $\mathcal{B}(\mathbb{R}^d)$; it is a Markov process with kernel Q_t . The observation process takes values in a measure space $(E, \mathcal{Y}, \lambda)$; $g_t(\cdot \mid x_t)$ is the conditional density of Y_t given $X_t = x_t$ w.r.t. the dominating measure λ .

We use throughout the article the term "state-space model" [8, 7] rather than the term "hidden Markov model" [5]. "State-space model" is more prevalent in the literature, especially on practical applications [4, 8]. Besides "hidden Markov model" often specifically refers to models where the state process is a discrete Markov chain [15], whereas we consider continuous value state processes in this article.

One typical example of state-space model is the following:

$$X_t = F_t(X_{t-1}) + V_t \tag{1}$$

$$Y_t = H_t(X_t) + W_t \tag{2}$$

for $t \geq 1$ with initial condition $X_0 \sim Q_0$, where $\{V_t\}_t$ and $\{W_t\}_t$ are independent noise processes (non necessarily Gaussian) and F_t and H_t are nonlinear mappings. This is the nonlinear non-Gaussian version of the dynamic linear model in Section 1.2.

Several statistical inference problems related to state-space models are addressed in the literature [5, 6, 7]:

- Parameter estimation: identifying the best data-generating process in a given parametric family. The parameter is static and is generally a vector. For example, in the model given by Equations (1) and (2), the parameter identifies the distributions of V_t and W_t and the mappings F_t and H_t .
- Filtering: estimating sequentially at each time step the current hidden state X_t given the available observations $Y_{0:t}$.
- Smoothing: estimating past hidden states X_s , where $s \leq t$, given observations $Y_{0:t}$.

In this paper, we address another inference problem, forecasting, defined as: estimating a future observation Y_{t+h} given the available observations $Y_{0:t}$.

In practical applications, the model structure is often known up to an unknown parameter and one must estimate the parameter before performing other

inference tasks (filtering, smoothing, forecasting). Parameter estimation can be done online (processing data points sequentially) or offline (processing the dataset as one batch), by likelihood maximization or Bayesian inference (see [6] for a comprehensive review). In this article, we address filtering and forecasting assuming that the model parameter, if any, is known or has been previously estimated.

2.2 Bayesian Filtering

Bayesian filtering consists in computing sequentially the conditional distribution of the current state X_t and of the next state X_{t+1} given the available observations $Y_{0:t}$ [5, 7].

Let the *predictor* and the *filter* respectively be the conditional distributions

$$\mu_{t|t-1}(dx_t) = P[X_t \in dx_t \mid Y_{0:t-1}]$$

and

$$\mu_t(dx_t) = P[X_t \in dx_t \mid Y_{0:t}]$$

at each time step $t \geq 0$ (with $\mu_{0|-1} = Q_0$ by convention).

The predictor and the filter obey to a recursive relation that is conveniently written with measure-valued operators. The prediction operator is defined as

$$\text{pred}_t(\nu) = \int Q_t(x, \cdot) \nu(dx) \quad (3)$$

and the update operator as

$$\text{upd}_t(\nu) = \frac{g_t(Y_t \mid \cdot) \nu(\cdot)}{\int g_t(Y_t \mid x) \nu(dx)} \quad (4)$$

at each time $t \geq 0$, where ν is a probability measure. One then has the prediction step

$$\mu_{t|t-1} = \text{pred}_t(\mu_{t-1})$$

from the Chapman-Kolmogorov equation and the update step

$$\mu_t = \text{upd}_t(\mu_{t|t-1})$$

from the Bayes formula.

2.3 Optimal Forecasting

Optimal forecasting consists in computing the conditional distribution of a future (non observed) data point Y_{t+h} , given the available observations $Y_{0:t}$, for some time horizon $h > 0$.

Let the *forecast* be the conditional distribution

$$f_{t+h|t}(dy_{t+h}) = P[Y_{t+h} \in dy_{t+h} \mid Y_{0:t}].$$

It is the "ideal forecast" in the terminology of Gneiting and Katzfuss [12].

Let

$$\begin{aligned} Q_{t+1:t+h}(x_t, dx_{t+h}) &= P[X_{t+h} \in dx_{t+h} \mid X_t = x_t] \\ &= \int_{\{x_{t+h-1} \in \mathbb{R}^d\}} \cdots \int_{\{x_{t+1} \in \mathbb{R}^d\}} Q_{t+h}(x_{t+h-1}, dx_{t+h}) \times \cdots \times Q_{t+1}(x_t, dx_{t+1}) \\ &= \int_{\{x_{t+h-1} \in \mathbb{R}^d\}} Q_{t+h}(x_{t+h-1}, dx_{t+h}) \cdots \int_{\{x_{t+1} \in \mathbb{R}^d\}} Q_{t+1}(x_t, dx_{t+1}) \end{aligned}$$

and let

$$\text{pred}_{t+1:t+h} = \text{pred}_{t+h} \circ \cdots \circ \text{pred}_{t+1}.$$

be the composition product of the prediction operators from time $t+1$ to time $t+h$, so that

$$\text{pred}_{t+1:t+h}(\nu) = \int Q_{t+1:t+h}(x, \cdot) \nu(dx). \quad (5)$$

Let the *state forecast* be the conditional distribution

$$\mu_{t+h|t}(dx_{t+h}) = P[X_{t+h} \in dx_{t+h} \mid Y_{0:t}].$$

One has that

$$\mu_{t+h|t}(dx_{t+h}) = \int_{\{x_t \in \mathbb{R}^d\}} Q_{t+1:t+h}(x_t, dx_{t+h}) \mu_t(dx_t),$$

that is

$$\mu_{t+h|t} = \text{pred}_{t+1:t+h}(\mu_t).$$

Besides, let the forecast operator be the measure-valued mapping defined as

$$\text{for}_{t+h}(\nu) = \left[\int g_{t+h}(\cdot | x) \nu(dx) \right] \lambda(\cdot). \quad (6)$$

The forecast verifies

$$\begin{aligned} f_{t+h|t}(dy_{t+h}) &= P[Y_{t+h} \in dy_{t+h} \mid Y_{0:t}] \\ &= \int_{\{x_{t+h} \in \mathbb{R}^d\}} P[Y_{t+h} \in dy_{t+h} \mid X_{t+h} = x_{t+h}] P[X_{t+h} \in dx_{t+h} \mid Y_{0:t}] \\ &= \left[\int g_{t+h}(y_{t+h} \mid x_{t+h}) \mu_{t+h|t}(dx_{t+h}) \right] \lambda(dy_{t+h}), \end{aligned}$$

so that

$$f_{t+h|t} = \text{for}_{t+h}(\mu_{t+h|t}) = \text{for}_{t+h}(\text{pred}_{t+1:t+h}(\mu_t)).$$

3 Particle Filtering and Particle Forecasting

Particle filtering (for Bayesian filtering) and particle forecasting (for optimal forecasting) are described in this section. Particle filters, also called sequential Monte Carlo methods, are algorithms based on importance sampling that recursively approximate the filter and the predictor. Several tutorial articles on particle filtering are available [9, 16, 10, 11]. The first version of particle filter has been proposed by Gordon et al. in [17]. Particle forecasting is an extension of particle filtering, which we propose in this paper as a novel sequential Monte Carlo method to approximate the optimal forecast.

3.1 Particle Filtering

Particle filtering consists in approximating the predictor and the filter recursively by a weighted sum of Dirac measures.

At initial time $t = 0$, the predictor is $\mu_{0|-1} = Q_0$. One generates an i.i.d. sample ξ_0^1, \dots, ξ_0^N of random vectors (called particles) in \mathbb{R}^d from Q_0 . The particle approximation of the predictor is the empirical measure

$$\mu_{0|-1}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_0^i}.$$

The observation Y_0 is then used through the update operator defined in Equation (4). The particle approximation of the filter is

$$\mu_0^N = \text{upd}_0(\mu_{0|-1}^N) = \frac{\sum_{i=1}^N g_0(Y_0 | \xi_0^i) \delta_{\xi_0^i}}{\sum_{i=1}^N g_0(Y_0 | \xi_0^i)} = \sum_{i=1}^N w_0^i \delta_{\xi_0^i}$$

where the weights are $w_0^i = \frac{g_0(Y_0 | \xi_0^i)}{\sum_{i=1}^N g_0(Y_0 | \xi_0^i)}$ for all $i \in \{1, \dots, N\}$ (so they sum up to one). Thus μ_0^N is obtained by importance sampling, with a sampling step and a weighting step.

Suppose now one has an approximation μ_{t-1}^N of the filter at time $t - 1 \geq 0$ as

$$\mu_{t-1}^N = \sum_{i=1}^N w_{t-1}^i \delta_{\xi_{t-1}^i}.$$

By applying the prediction operator defined in Equation (3) to μ_{t-1}^N , one gets a mixture distribution

$$\text{pred}_t(\mu_{t-1}^N) = \int Q_t \mu_{t-1}^N = \sum_{i=1}^N w_{t-1}^i Q_t(\xi_{t-1}^i, \cdot). \quad (7)$$

The approximation $\mu_{t|t-1}^N$ of the predictor is obtained by drawing a sample of new particles from this mixture distribution, independently given $\xi_{t-1}^1, \dots, \xi_{t-1}^N$. In practice, the new particles ξ_t^1, \dots, ξ_t^N are generated by repeating the two following steps for all $i \in \{1, \dots, N\}$:

Selection Draw an index $j \in \{1, \dots, N\}$ where the drawing probabilities are the weights $w_{t-1}^1, \dots, w_{t-1}^N$, i.e. draw j from the Multinomial distribution over $\{1, \dots, N\}$ with parameters $w_{t-1}^1, \dots, w_{t-1}^N$.

Propagation Generate ξ_t^i from the distribution $Q_t(\xi_{t-1}^j, \cdot)$, i.e. the mixture component corresponding to the selected index.

The particle approximation of the predictor is then

$$\mu_{t|t-1}^N = \sum_{i=1}^N w_{t-1}^i \delta_{\xi_t^i}$$

where $\xi_t^1, \dots, \xi_t^N \mid (\xi_{t-1}^1, \dots, \xi_{t-1}^N) \stackrel{\text{i.i.d.}}{\sim} \text{pred}_t(\mu_{t-1}^N)$.

The particle approximation of the filter is obtained by updating the predictor approximation. Applying the update operator to $\mu_{t|t-1}^N$ yields

$$\mu_t^N = \text{upd}_t(\mu_{t|t-1}^N) = \sum_{i=1}^N w_t^i \delta_{\xi_t^i},$$

where the new weights are $w_t^i = \frac{g_t(Y_t \mid \xi_t^i)}{\sum_{i=1}^N g_t(Y_t \mid \xi_t^i)}$ for all $i \in \{1, \dots, N\}$.

The particle filter is summarized in Algorithm 1. This version corresponds to the seminal algorithm proposed by Gordon et al. in [17] and is also referred to as the bootstrap filter.

In the bootstrap filter, the particle sampling mechanism is given by the Markov kernel Q_t from the state process. It is possible to use a different sampling distribution, denoted \tilde{Q}_t , as soon as the derivative (in the Radon-Nikodym sense) $\tilde{g}_t(x_{t-1}, \cdot) = \frac{dQ_t(x_{t-1}, \cdot)}{d\tilde{Q}_t(x_{t-1}, \cdot)}$ exists for all $x_{t-1} \in \mathbb{R}^d$. In this case,

the particles are sampled as $\xi_t^i \sim \tilde{Q}_t(\xi_{t-1}^i, \cdot)$ and the weights are computed as $w_t^i = \frac{g_t(\xi_t^i) \tilde{g}_t(\xi_{t-1}^i, \xi_t^i)}{\sum_{i=1}^N g_t(\xi_t^i) \tilde{g}_t(\xi_{t-1}^i, \xi_t^i)}$ for all $i \in \{1, \dots, N\}$. Although choosing Q_t is

practically simpler, using a different sampling kernel \tilde{Q}_t in the propagation step can lead to lower weight variance and better algorithm performance [16, 10, 11].

The resampling step is necessary to avoid particle degeneracy. Without resampling, after some time most particles get a zero weight and only a few particles get a significant weight. This impoverishment of particle weights drastically degrades particle approximations. However it is not necessary in practice to perform resampling at each time step. Resampling is generally done according to a criterion monitoring the quality of the particle sample, for example when the effective sample size $N_{\text{eff}} = \frac{1}{\sum_{i=1}^N (w_t^i)^2}$ falls below a preset threshold.

Algorithm 1 Particle Filter

```
observe  $Y_0$ 
for  $i = 1 \dots N$  do
  sample  $\xi_0^i \sim Q_0$ 
  compute  $w_0^i \propto g_0(Y_0 \mid \xi_0^i)$ 
end for
normalize weights
set  $t$  to 1
loop
  observe  $Y_t$ 
  for  $i = 1 \dots N$  do
    sample  $j \sim \text{Multinomial}(\{1, \dots, N\}, \{w_{t-1}^1, \dots, w_{t-1}^N\})$   $\triangleright$  selection
    sample  $\xi_t^i \sim Q_t(\cdot \mid \xi_{t-1}^j)$   $\triangleright$  propagation
    compute  $w_t^i \propto g_t(Y_t \mid \xi_t^i)$   $\triangleright$  weighting
  end for
  normalize weights
  increment  $t$ 
end loop
```

3.2 Particle Forecasting

The algorithm proposed in this article approximates the forecast (conditional distribution of the future data point Y_{t+h} given the available observations $Y_{0:t}$) and the state forecast (conditional distribution of X_{t+h} given $Y_{0:t}$) defined in Section 2. The particle forecast is run after the particle filter and uses the current weights and particles at time t .

Suppose one has the approximation μ_t^N of the filter at time t given by the particle filter in Algorithm 1.

By applying the composition product of prediction operators defined in Equation (5), one gets

$$\text{pred}_{t+1:t+h}(\mu_t^N) = \sum_{i=1}^N w_t^i Q_{t+1:t+h}(\xi_t^i, \cdot). \quad (8)$$

It is a mixture distribution, as in Equation (7) for the particle filter. The approximation $\mu_{t+h|t}^N$ of the state forecast is obtained by sampling from this distribution independently given ξ_t^1, \dots, ξ_t^N . Like in particle filtering, this is done by repeating N times a selection step and a propagation step. Here, after an index $j \in \{1, \dots, N\}$ is selected according to the weights w_t^1, \dots, w_t^N , a new particle is generated by propagating h steps ahead the selected particle ξ_t^j . In practice, one draws from the distribution $Q_{t+1:t+h}(\xi_t^j, \cdot)$ by recursively generating new particles using the Markov kernel from time $t+1$ to time $t+h$. The particle approximation of the state forecast is then

$$\mu_{t+h|t}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{t+h|t}^i}$$

where $\xi_{t+h|t}^1, \dots, \xi_{t+h|t}^N \mid (\xi_t^1, \dots, \xi_t^N) \stackrel{\text{i.i.d.}}{\sim} \text{pred}_{t+1:t+h}(\mu_t^N)$.

Then, by applying the forecast operator defined in Equation (6) to $\mu_{t+h|t}^N$, one gets

$$\text{for}_{t+h}(\mu_{t+h|t}^N) = \frac{1}{N} \sum_{i=1}^N g_{t+h}(\cdot \mid \xi_{t+h|t}^i) \lambda(\cdot). \quad (9)$$

It is again a mixture distribution. To generate from this mixture, an index $j \in \{1, \dots, N\}$ is randomly selected with probability $1/N$ and one draws from the corresponding mixture component $g_{t+h}(\cdot \mid \xi_{t+h}^j) \lambda(\cdot)$. The particle approximation of the forecast is then

$$f_{t+h|t}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\zeta_{t+h|t}^i}$$

where $\zeta_{t+h|t}^1, \dots, \zeta_{t+h|t}^N \mid (\xi_{t+h|t}^1, \dots, \xi_{t+h|t}^N) \stackrel{\text{i.i.d.}}{\sim} \text{for}_{t+h}(\mu_{t+h|t}^N)$.

Particle forecasting is thus based on the same algorithmic ideas as particle filtering. It is summarized in Algorithm 2.

Algorithm 2 Particle Forecast

```

for  $i = 1 \dots N$  do
  sample  $j \sim \text{Multinomial}(\{1, \dots, N\}, \{w_t^1, \dots, w_t^N\})$ 
  sample  $\xi_{t+1|t}^i \sim Q_{t+1}(\xi_t^j, \cdot)$ 
  for  $s = 2 \dots h$  do
    sample  $\xi_{t+s|t}^i \sim Q_{t+s}(\xi_{t+s-1|t}^i, \cdot)$ 
  end for
end for
for  $i = 1 \dots N$  do
  sample  $j \sim \text{Multinomial}(\{1, \dots, N\}, \{1/N, \dots, 1/N\})$ 
  sample  $\zeta_{t+h|t}^i \sim g_{t+h}(\cdot \mid \xi_{t+h|t}^j) \lambda(\cdot)$ 
end for

```

4 Convergence Analysis

First results on the convergence of particle filtering were obtained by Del Moral in [18, 19]. An analysis of the algorithm with randomly selected resampling times has been provided by Douc and Moulines in [20]. In [13], Crisan and Doucet present convergence results based on a description of particle filtering with continuous measure-valued operators. We follow a similar approach in this article to review results for particle filtering and extend them to particle forecasting. We address almost sure convergence and convergence in mean squared error in the weak topology of probability measures [21, Chapter 2]. In this topology, a sequence of probability measures $\{\lambda_n\}_n$ tends to λ if $\int \phi(x) \lambda_n(dx)$ tends to $\int \phi(x) \lambda(dx)$ for all continuous bounded function ϕ .

4.1 Sampling Operator

Particle filtering and particle forecasting involve sampling from random probability measures.

Let $\nu^N(V_{1:N})$ by a random measure determined by a random vector $V_{1:N} = (V_1, \dots, V_N)$, i.e.

$$\mathbb{E}[\int \phi \nu^N(V_{1:N}) \mid V_{1:N}] = \int \phi \nu^N(V_{1:N})$$

for all continuous bounded function ϕ . The sampling operator s^N maps $\nu^N(V_{1:N})$ to an empirical measure given by a sample U_1, \dots, U_N from $\nu^N(V_{1:N})$, such that U_i and U_j are conditionally independent given $V_{1:N}$ for all $i \neq j$. That is

$$s^N(\nu^N(V_{1:N})) = \frac{1}{N} \sum_{i=1}^N \delta_{U_i} \quad \text{where } U_1, \dots, U_N \mid V_{1:N} \stackrel{\text{i.i.d.}}{\sim} \nu^N(V_{1:N}).$$

In particle filtering, the distribution $\text{pred}_t(\mu_{t-1}^N)$ in Equation (7) is a random measure determined by $(\xi_{t-1}^1, \dots, \xi_{t-1}^N)$, hence it is in the form over which s^N applies. Drawing particles as

$$\xi_t^1, \dots, \xi_t^N \mid (\xi_{t-1}^1, \dots, \xi_{t-1}^N) \stackrel{\text{i.i.d.}}{\sim} \text{pred}_t(\mu_{t-1}^N)$$

to approximate the predictor $\mu_{t|t-1}^N$ is formalized as applying s^N to $\text{pred}_t(\mu_{t-1}^N)$, so that one has

$$\mu_{t|t-1}^N = s^N(\text{pred}_t(\mu_{t-1}^N)).$$

In particle forecasting, the distributions $\text{pred}_{t+1:t+h}(\mu_t^N)$ in Equation (8) and $\text{for}_{t+h}(\mu_{t+h|t}^N)$ in Equation (9) are also random measures respectively determined by the random vectors $(\xi_t^1, \dots, \xi_t^N)$ and $(\xi_{t+h|t}^1, \dots, \xi_{t+h|t}^N)$. Drawing

$$\xi_{t+h|t}^1, \dots, \xi_{t+h|t}^N \mid (\xi_t^1, \dots, \xi_t^N) \stackrel{\text{i.i.d.}}{\sim} \text{pred}_{t+1:t+h}(\mu_t^N),$$

to approximate the state forecast $\mu_{t+h|t}$, is formalized as

$$\mu_{t+h|t}^N = s^N(\text{pred}_{t+1:t+h}(\mu_t^N))$$

and drawing

$$\zeta_{t+h|t}^1, \dots, \zeta_{t+h|t}^N \mid (\xi_{t+h|t}^1, \dots, \xi_{t+h|t}^N) \stackrel{\text{i.i.d.}}{\sim} \text{for}_{t+h}(\mu_{t+h|t}^N),$$

to approximate the forecast $f_{t+h|t}$, as

$$f_{t+h|t}^N = s^N(\text{for}_{t+h}(\mu_{t+h|t}^N)).$$

Thus, Bayesian filtering and its particle approximation can be schematized as

$$\mu_{t-1} \xrightarrow{\text{pred}_t} \mu_{t|t-1} \xrightarrow{\text{upd}_t} \mu_t$$

$$\mu_{t-1}^N \xrightarrow{s^N \circ \text{pred}_t} \mu_{t|t-1}^N \xrightarrow{\text{upd}_t} \mu_t^N,$$

and optimal forecasting and its particle approximation can be schematized as

$$\begin{aligned} \mu_t &\xrightarrow{\text{pred}_{t+1:t+h}} \mu_{t+h|t} \xrightarrow{\text{for}_{t+h}} f_{t+h|t} \\ \mu_t^N &\xrightarrow{s^N \circ \text{pred}_{t+1:t+h}} \mu_{t+h|t}^N \xrightarrow{s^N \circ \text{for}_{t+h}} f_{t+h|t}^N. \end{aligned}$$

The two following lemmas hold for the random probability measure $\nu^N(V_{1:N})$ and the sampling operator s^N defined above. Their proofs are provided in Appendix 6.1.

Lemma 1. *For all continuous bounded function ϕ , if*

$$\int \phi \nu^N(V_{1:N}) \xrightarrow{N \rightarrow \infty} \int \phi \nu \text{ almost surely,}$$

then

$$\int \phi s^N(\nu^N(V_{1:N})) \xrightarrow{N \rightarrow \infty} \int \phi \nu \text{ almost surely.}$$

Lemma 2. *For all continuous bounded function ϕ and positive real number m , if*

$$\mathbb{E}[(\int \phi \nu^N(V_{1:N}) - \int \phi \nu)^2] \leq \frac{m}{N},$$

then

$$\mathbb{E}[(\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu)^2] \leq \frac{(\sqrt{m} + \sqrt{2}|\sup \phi|)^2}{N}.$$

4.2 Convergence of Particle Filtering

The almost sure convergence and the convergence in mean squared error, in the weak topology, of particle filtering can be established under Assumptions 1 and 2 below [13].

Assumption 1. *For all $t \geq 0$ and for all continuous bounded function ϕ , the mapping $x_{t-1} \mapsto \int \phi(x_t) Q_t(x_{t-1}, dx_t)$ is continuous and bounded.*

Assumption 2. *For all $t \geq 0$, $g_t(y | \cdot)$ is continuous, bounded and strictly positive for all y (λ almost surely).*

For example, consider the state process

$$X_t = F_t(X_{t-1}) + V_t$$

for $t \geq 1$ where $\{V_t\}_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_v^2)$ (and $X_0 \sim Q_0$). If the mapping F_t is continuous, then

$$\int \phi(x_t) Q_t(x_{t-1}, dx_t) = \frac{1}{\sqrt{2\pi}\sigma_v} \int \phi(x_t) e^{-\frac{1}{2\sigma_v^2}(x_t - F_t(x_{t-1}))^2} dx_t$$

is continuous w.r.t. x_{t-1} and Assumption 1 is verified.

Now consider the observation process

$$Y_t = H_t(X_t) + W_t \quad (10)$$

for $t \geq 0$ where $\{W_t\}_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2)$. If the mapping H_t is continuous, then

$$g_t(y | x) = \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{1}{2\sigma_w^2}(y-H_t(x))^2}$$

verifies Assumption 2.

Under Assumptions 1 and 2, the measure-valued operators for prediction and update in Bayesian filtering, defined in Equations (3) and (4), are continuous.

Proposition 1. *Under Assumption 1, the prediction operator pred_t is continuous in the weak topology.*

Proposition 2. *Under Assumption 2, the update operator upd_t is continuous in the weak topology.*

The proofs of Propositions 1 and 2 are provided in Appendix 6.2.

The convergence of particle filtering is stated in Theorems 1 and 2 below.

Theorem 1. *Under Assumptions 1 and 2, for all continuous bounded function ϕ ,*

$$\int \phi \mu_{t|t-1}^N \xrightarrow{N \rightarrow \infty} \int \phi \mu_{t|t-1} \text{ almost surely}$$

and

$$\int \phi \mu_t^N \xrightarrow{N \rightarrow \infty} \int \phi \mu_t \text{ almost surely.}$$

Proof. At initial time, one has that $\mu_{0|-1}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_0^i}$ where $\xi_0^1, \dots, \xi_0^N \stackrel{\text{i.i.d.}}{\sim} \mu_{0|-1}$ ($\mu_{0|-1} = Q_0$ by convention). Then $\int \phi \mu_{0|-1}^N \xrightarrow{N \rightarrow \infty} \int \phi \mu_{0|-1}$ almost surely according to the strong law of large numbers. The operator upd_0 is continuous according to Proposition 2 and $\mu_0^N = \text{upd}_0(\mu_{0|-1}^N)$, hence $\int \phi \mu_0^N \xrightarrow{N \rightarrow \infty} \int \phi \mu_0$ almost surely.

Let $t \geq 1$. Suppose that $\int \phi \mu_{t-1}^N \xrightarrow{N \rightarrow \infty} \int \phi \mu_{t-1}$ almost surely. The operator pred_t is continuous according to Proposition 1, then $\int \phi \text{pred}_t(\mu_{t-1}^N) \xrightarrow{N \rightarrow \infty} \int \phi \text{pred}_t(\mu_{t-1})$ almost surely. Since $\mu_{t|t-1}^N = s^N(\text{pred}_t(\mu_{t-1}^N))$, then Lemma 1 yields $\int \phi \mu_{t|t-1}^N \xrightarrow{N \rightarrow \infty} \int \phi \mu_{t|t-1}$ almost surely.

The operator upd_t is continuous according to Proposition 2 and $\mu_t^N = \text{upd}_t(\mu_{t|t-1}^N)$, then $\int \phi \mu_t^N \xrightarrow{N \rightarrow \infty} \int \phi \mu_t$ almost surely. \square

Theorem 2. Under Assumptions 1 and 2, for all continuous bounded function ϕ , there exists positive real numbers $m_{t|t-1}(\phi)$ and $m_t(\phi)$ such that

$$\mathbb{E}[(\int \phi \mu_{t|t-1}^N - \int \phi \mu_{t|t-1})^2] \leq \frac{m_{t|t-1}(\phi)}{N}$$

and

$$\mathbb{E}[(\int \phi \mu_t^N - \int \phi \mu_t)^2] \leq \frac{m_t(\phi)}{N}.$$

Proof. At initial time, one has that $\mu_0 = \frac{g_0 Q_0}{\int g_0 Q_0}$ and $\mu_0^N = \frac{g_0 s^N(Q_0)}{\int g_0 s^N(Q_0)}$. Then

$$\begin{aligned} \frac{\int \phi g_0 s^N(Q_0)}{\int g_0 s^N(Q_0)} - \frac{\int \phi g_0 Q_0}{\int g_0 Q_0} &= \frac{\int \phi g_0 s^N(Q_0) - \int \phi g_0 Q_0}{\int g_0 Q_0} \\ &\quad - \frac{\int \phi g_0 s^N(Q_0)}{\int g_0 s^N(Q_0)} \times \frac{\int g_0 s^N(Q_0) - \int g_0 Q_0}{\int g_0 Q_0}. \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{E}[(\frac{\int \phi g_0 s^N(Q_0)}{\int g_0 s^N(Q_0)} - \frac{\int \phi g_0 Q_0}{\int g_0 Q_0})^2]^{1/2} &\leq \mathbb{E}[(\frac{\int \phi g_0 s^N(Q_0) - \int \phi g_0 Q_0}{\int g_0 Q_0})^2]^{1/2} \\ &\quad + \mathbb{E}[(\frac{\int \phi g_0 s^N(Q_0)}{\int g_0 s^N(Q_0)} \times \frac{\int g_0 s^N(Q_0) - \int g_0 Q_0}{\int g_0 Q_0})^2]^{1/2} \end{aligned}$$

according to the Minkowski inequality, where the first term of the right-hand side verifies

$$\mathbb{E}[(\frac{\int \phi g_0 s^N(Q_0) - \int \phi g_0 Q_0}{\int g_0 Q_0})^2] = \frac{1}{N} \frac{\int \phi^2 g_0^2 Q_0 - (\int \phi g_0 Q_0)^2}{(\int g_0 Q_0)^2}$$

and the second term verifies

$$\begin{aligned} \mathbb{E}[(\frac{\int \phi g_0 s^N(Q_0)}{\int g_0 s^N(Q_0)} \times \frac{\int g_0 s^N(Q_0) - \int g_0 Q_0}{\int g_0 Q_0})^2] &\leq (\sup \phi)^2 \mathbb{E}[(\frac{\int g_0 s^N(Q_0) - \int g_0 Q_0}{\int g_0 Q_0})^2] \\ &= \frac{(\sup \phi)^2}{N} \frac{\int g_0^2 Q_0 - (\int g_0 Q_0)^2}{(\int g_0 Q_0)^2} \end{aligned}$$

as $\int g_0 s^N(Q_0) = \frac{1}{N} \sum_{i=1}^N g_0(Y_0 | \xi_0^i)$ where $\xi_0^1, \dots, \xi_0^N \stackrel{\text{i.i.d.}}{\sim} Q_0$.

Let $t \geq 1$. Suppose that $\mathbb{E}[(\int \phi \mu_{t-1}^N - \int \phi \mu_{t-1})^2] \leq \frac{m_{t-1}(\phi)}{N}$ for all continuous bounded function ϕ . Then

$$\begin{aligned} &\mathbb{E}[(\int \phi [\int Q_t \mu_{t-1}^N] - \int \phi [\int Q_t \mu_{t-1}])^2] \\ &= \mathbb{E}[(\int [\int \phi Q_t] \mu_{t-1}^N - \int [\int \phi Q_t] \mu_{t-1})^2] \\ &\leq \frac{m_{t-1}(\int \phi Q_t)}{N} \end{aligned}$$

as $\int \phi(x_t) Q_t(\cdot, dx_t)$ is continuous and bounded from Assumption 1. Since $\mu_{t|t-1} = \int Q_t \mu_{t-1}$ and $\mu_{t|t-1}^N = s^N(\int Q_t \mu_{t-1}^N)$, then

$$\mathbb{E}[(\int \phi \mu_{t|t-1}^N - \int \phi \mu_{t|t-1})^2] \leq \frac{m_{t|t-1}(\phi)}{N} \quad (11)$$

according to Lemma 2, where $m_{t|t-1}(\phi) = (\sqrt{m_{t-1}(\int \phi Q_t)} + \sqrt{2}|\sup \phi|)^2$, for all continuous bounded function ϕ .

One has that $\mu_t = \frac{g_t \mu_{t|t-1}}{\int g_t \mu_{t|t-1}}$ and $\mu_t^N = \frac{g_t \mu_{t|t-1}^N}{\int g_t \mu_{t|t-1}^N}$. Then

$$\begin{aligned} \frac{\int \phi g_t \mu_{t|t-1}^N}{\int g_t \mu_{t|t-1}^N} - \frac{\int \phi g_t \mu_{t|t-1}}{\int g_t \mu_{t|t-1}} &= \frac{\int \phi g_t \mu_{t|t-1}^N - \int \phi g_t \mu_{t|t-1}}{\int g_t \mu_{t|t-1}} \\ &\quad - \frac{\int \phi g_t \mu_{t|t-1}^N}{\int g_t \mu_{t|t-1}^N} \times \frac{\int g_t \mu_{t|t-1}^N - \int g_t \mu_{t|t-1}}{\int g_t \mu_{t|t-1}}. \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{E}[(\frac{\int \phi g_t \mu_{t|t-1}^N}{\int g_t \mu_{t|t-1}^N} - \frac{\int \phi g_t \mu_{t|t-1}}{\int g_t \mu_{t|t-1}})^2]^{1/2} &\leq \mathbb{E}[(\frac{\int \phi g_t \mu_{t|t-1}^N - \int \phi g_t \mu_{t|t-1}}{\int g_t \mu_{t|t-1}})^2]^{1/2} \\ &\quad + \mathbb{E}[(\frac{\int \phi g_t \mu_{t|t-1}^N}{\int g_t \mu_{t|t-1}^N} \times \frac{\int g_t \mu_{t|t-1}^N - \int g_t \mu_{t|t-1}}{\int g_t \mu_{t|t-1}})^2]^{1/2} \end{aligned}$$

according to the Minkowski inequality, where the first term of the right-hand side verifies

$$\mathbb{E}[(\frac{\int \phi g_t \mu_{t|t-1}^N - \int \phi g_t \mu_{t|t-1}}{\int g_t \mu_{t|t-1}})^2] \leq \frac{m_{t|t-1}(\phi g_t)}{(\int g_t \mu_{t|t-1})^2 N}$$

and the second term verifies

$$\begin{aligned} &\mathbb{E}[(\frac{\int \phi g_t \mu_{t|t-1}^N}{\int g_t \mu_{t|t-1}^N} \times \frac{\int g_t \mu_{t|t-1}^N - \int g_t \mu_{t|t-1}}{\int g_t \mu_{t|t-1}})^2] \\ &\leq (\frac{\sup \phi}{\int g_t \mu_{t|t-1}})^2 \mathbb{E}[(\int g_t \mu_{t|t-1}^N - \int g_t \mu_{t|t-1})^2] \\ &\leq \frac{(\sup \phi)^2 m_{t|t-1}(g_t)}{(\int g_t \mu_{t|t-1})^2 N} \end{aligned}$$

according to Equation (11), as g_t is continuous and bounded from Assumption 2. \square

4.3 Convergence of Particle Forecasting

In this section, we address the convergence (almost surely and in mean squared error) of the particle approximation of the state forecast and of the forecast.

First, the results for the state forecast are stated in Theorems 3 and 4 below.

Theorem 3. *Under Assumptions 1 and 2, for all continuous bounded function ϕ ,*

$$\int \phi \mu_{t+h|t}^N \xrightarrow{N \rightarrow \infty} \int \phi \mu_{t+h|t} \text{ almost surely.}$$

Proof. One has that $\int \phi \mu_t^N \xrightarrow{N \rightarrow \infty} \int \phi \mu_t$ almost surely according to Theorem 1. The operator $\text{pred}_{t+1:t+h}$ is continuous according to Proposition 1, then $\int \phi \text{pred}_{t+1:t+h}(\mu_t^N) \xrightarrow{N \rightarrow \infty} \int \phi \text{pred}_{t+1:t+h}(\mu_t)$ almost surely. Since $\mu_{t+h|t}^N = s^N(\text{pred}_{t+1:t+h}(\mu_t^N))$, then Lemma 1 yields $\int \phi \mu_{t+h|t}^N \xrightarrow{N \rightarrow \infty} \int \phi \mu_{t+h|t}$ almost surely. \square

Theorem 4. *Under Assumptions 1 and 2, for all continuous bounded function ϕ , there exists a positive real number $m_{t+h|t}(\phi)$ such that*

$$\mathbb{E}[(\int \phi \mu_{t+h|t}^N - \int \phi \mu_{t+h|t})^2] \leq \frac{m_{t+h|t}(\phi)}{N^2}.$$

Proof. According to Theorem 2,

$$\begin{aligned} & \mathbb{E}[(\int \phi [\int Q_{t+h:t+1} \mu_t^N] - \int \phi [\int Q_{t+h:t+1} \mu_t])^2] \\ &= \mathbb{E}[(\int [\int \phi Q_{t+h:t+1}] \mu_t^N - \int [\int \phi Q_{t+h:t+1}] \mu_t)^2] \\ &\leq \frac{m_t(\int \phi Q_{t+h:t+1})}{N} \end{aligned}$$

as $\int \phi(x_{t+h}) Q_{t+h:t+1}(\cdot, x_{t+h})$ is continuous and bounded from Assumption 1.

Since $\mu_{t+h|t} = \int Q_{t+h:t+1} \mu_t$ and $\mu_{t+h|t}^N = s^N(\int Q_{t+h:t+1} \mu_t^N)$, then

$$\mathbb{E}[(\int \phi \mu_{t+h|t}^N - \int \phi \mu_{t+h|t})^2] \leq \frac{m_{t+h|t}(\phi)}{N^2}$$

according to Lemma 2, where $m_{t+h|t}(\phi) = (\sqrt{m_t(\int \phi Q_{t+h:t+1})} + \sqrt{2}|\sup \phi|)^2$. \square

We then additionally require Assumption 3 below.

Assumption 3. For all $t \geq 0$ and continuous bounded function ϕ , the mapping $x \mapsto \int \phi(y) g_t(y | x) \lambda(dy)$ is continuous.

For example, if H_t is continuous in the observation process in Equation (10) above, then Assumption 3 is also verified.

Under Assumption 3, the measure-valued operator for optimal forecast, defined in Equation (6), is continuous.

Proposition 3. Under Assumption 3, the forecast operator for_t is continuous in the weak topology.

The proof of Proposition 3 is provided in Appendix 6.2.

The convergence of the particle approximation of the forecast is stated in Theorems 5 and 6 below.

Theorem 5. Under Assumptions 1, 2 and 3, for all continuous bounded function ϕ ,

$$\int \phi f_{t+h|t}^N \xrightarrow{N \rightarrow \infty} \int \phi f_{t+h|t} \text{ almost surely.}$$

Proof. One has that $\int \phi \mu_{t+h|t}^N \xrightarrow{N \rightarrow \infty} \int \phi \mu_{t+h|t}$ almost surely according to Theorem 3. The operator for_{t+h} is continuous according to Proposition 3, then $\int \phi \text{for}_{t+h}(\mu_{t+h|t}^N) \xrightarrow{N \rightarrow \infty} \int \phi \text{for}_{t+h}(\mu_{t+h|t})$ almost surely. Since $f_{t+h|t}^N = s^N(\text{for}_{t+h}(\mu_{t+h|t}^N))$, then Lemma 1 yields $\int \phi f_{t+h|t}^N \xrightarrow{N \rightarrow \infty} \int \phi f_{t+h|t}$ almost surely. \square

Theorem 6. Under Assumptions 1, 2 and 3, for all continuous bounded function ϕ , there exists a positive real number $m'_{t+h}(\phi)$ such that

$$\mathbb{E}[(\int \phi f_{t+h|t}^N - \int \phi f_{t+h|t})^2] \leq \frac{m'_{t+h}(\phi)}{N^2}.$$

Proof. According to Theorem 4,

$$\begin{aligned} & \mathbb{E}[(\int \phi(y_{t+h}) [\int g_{t+h}(y_{t+h} | \cdot) \mu_{t+h|t}^N(\cdot)] \lambda(dy_{t+h}) - \int \phi(y_{t+h}) [\int g_{t+h}(y_{t+h} | \cdot) \mu_{t+h|t}(\cdot)] \lambda(dy_{t+h}))^2] \\ &= \mathbb{E}[(\int [\int \phi(y_{t+h}) g_{t+h}(y_{t+h} | \cdot) \lambda(dy_{t+h})] \mu_{t+h|t}^N(\cdot) - \int [\int \phi(y_{t+h}) g_{t+h}(y_{t+h} | \cdot) \lambda(dy_{t+h})] \mu_{t+h|t}^N(\cdot))^2] \\ &\leq \frac{m_{t+h|t}(\int \phi(y_{t+h}) g_{t+h}(y_{t+h} | \cdot) \lambda(dy_{t+h}))}{N} \end{aligned}$$

as $\int \phi(y_{t+h}) g_{t+h}(y_{t+h} | \cdot) \lambda(dy_{t+h})$ is bounded and continuous from Assumption

3. Since $f_{t+h|t} = (\int g_{t+h} \mu_{t+h|t}) \lambda$ and $f_{t+h|t}^N = s^N((\int g_{t+h} \mu_{t+h|t}^N) \lambda)$, then

$$\mathbb{E}[(\int \phi f_{t+h|t}^N - \int \phi f_{t+h|t})^2] \leq \frac{m'_{t+h}(\phi)}{N}$$

according to Lemma 2, where $m'_{t+h|t}(\phi) = (\sqrt{m_{t+h|t}(\int \phi(y_{t+h}) g_{t+h}(y_{t+h} | \cdot) \lambda(dy_{t+h})) + \sqrt{2}|\sup \phi|)^2$. \square

5 Numerical Experiments

In this section we perform a simulation experiment to empirically check that particle forecast correctly approximates the optimal forecast as the number of particles increases.

5.1 Model and Performance Evaluation

We consider the following reference nonlinear state-space model [16, 10, 17, 22, 23]:

$$X_t = \frac{X_{t-1}}{2} + \frac{25X_{t-1}}{1 + X_{t-1}^2} + 8 \cos(1.2 t) + V_t \quad (12)$$

$$Y_t = \frac{X_t^2}{20} + W_t \quad (13)$$

for $t \geq 1$ with initial condition $X_0 \sim \mathcal{N}(0, 10)$, where $\{V_t\}_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 10)$ and $\{W_t\}_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ are independent Gaussian white noise processes.

The quality of forecasts is assessed with the goodness-of-fit procedure of Diebold et al. [24, 12]. This procedure applies to one-dimensional distribution forecasts and is based on the probability integral transform (PIT) of forecasts. Given observed data $Y_{0:t}$, when a forecast for Y_{t+h} is available in the form of a cumulative distribution function (CDF) F , then its PIT is $F(Y_{t+h})$ (Y_{t+h} is the data point actually observed at time $t+h$). When the forecast is optimal, i.e. when F is the true conditional cdf of Y_{t+h} given $Y_{0:t}$, then the PIT is independent of $Y_{0:t}$ and follows the uniform distribution over $[0, 1]$ (see Appendix 6.3).

Thus, once the data points corresponding to the forecasts are actually observed, one can perform statistical tests to verify that the sample of PITs at times $t+h$, $t+2h$, ... is an i.i.d. sample from $\mathcal{U}(0, 1)$, which must be the case if the forecasting method produces optimal forecasts.

In particle forecasting, the conditional CDF of Y_{t+h} given $Y_{0:t}$ can readily be computed as $F_{t+h|t}^N(x) = \int_{-\infty}^x f_{t+h|t}^N = \frac{1}{N} \sum_{i=1}^N 1(\zeta_{t+h|t}^i < x)$ (where $1(a < b) = 1$ if $a < b$ and 0 otherwise).

In our experiments, we compute sequences of PITs with time lag h . We then perform the Ljung-Box test [25] and the Kolmogorov-Smirnov test [26], which respectively verify the two following null hypotheses:

- the PITs are independently distributed,
- the PITs are i.i.d. from a given distribution (here $\mathcal{U}(0, 1)$).

5.2 Results

We simulate a time series from the model given in Equations (12) and (13) from time 0 to 1000. Data points are observed sequentially and forecasting starts as soon as the first data point is observed. After each new observation, new particles and their weights are computed by one particle filter step, and h -step ahead particle forecasting is run. The PIT of each forecast is computed when the data point corresponding to the forecast is observed. The algorithm is described in Algorithm 3.

Algorithm 3 Sequential Particle Filtering and Forecasting

```

observe  $Y_0$ 
for  $i = 1 \dots N$  do                                      $\triangleright$  particle filter initialization
    sample  $\xi_0^i \sim Q_0$ 
    compute  $w_0^i \propto g_0(Y_0 \mid \xi_0^i)$ 
end for
normalize weights
set  $t$  to 0
loop
    for  $i = 1 \dots N$  do                                      $\triangleright$  particle forecast ( $h$ -step ahead)
        sample  $j \sim \text{Multinomial}(\{1, \dots, N\}, \{w_t^1, \dots, w_t^N\})$ 
        sample  $\xi_{t+1|t}^i \sim Q_{t+1}(\xi_t^j, \cdot)$ 
        for  $s = 2 \dots h$  do
            sample  $\xi_{t+s|t}^i \sim Q_{t+s}(\xi_{t+s-1|t}^i, \cdot)$ 
        end for
    end for
    for  $i = 1 \dots N$  do
        sample  $j \sim \text{Multinomial}(\{1, \dots, N\}, \{1/N, \dots, 1/N\})$ 
        sample  $\zeta_{t+h|t}^i \sim g_{t+h}(\cdot \mid \xi_{t+h|t}^j) \lambda(\cdot)$ 
    end for
    increment  $t$ 
    observe  $Y_t$ 
    for  $i = 1 \dots N$  do                                      $\triangleright$  particle filter step
        sample  $j \sim \text{Multinomial}(\{1, \dots, N\}, \{w_{t-1}^1, \dots, w_{t-1}^N\})$ 
        sample  $\xi_t^i \sim Q_t(\cdot \mid \xi_{t-1}^j)$ 
        compute  $w_t^i \propto g_t(Y_t \mid \xi_t^i)$ 
    end for
    normalize weights
end loop

```

Different particle sample sizes N are used. For each N , the simulation is run 100 times. For each run, the Ljung-Box test of independence and the Kolmogorov-Smirnov test of uniformity are performed.

In Figures 1 to 4, the empirical CDF (ECDF) of their test statistics over the 100 runs is displayed for different values of N . The true CDF under the null hypothesis is also plotted. Results are for 1-step ahead and 5-step ahead

forecasting. When the number of particle increases, the empirical distributions of the test statistics are seen to converge towards the distributions under the null hypothesis.

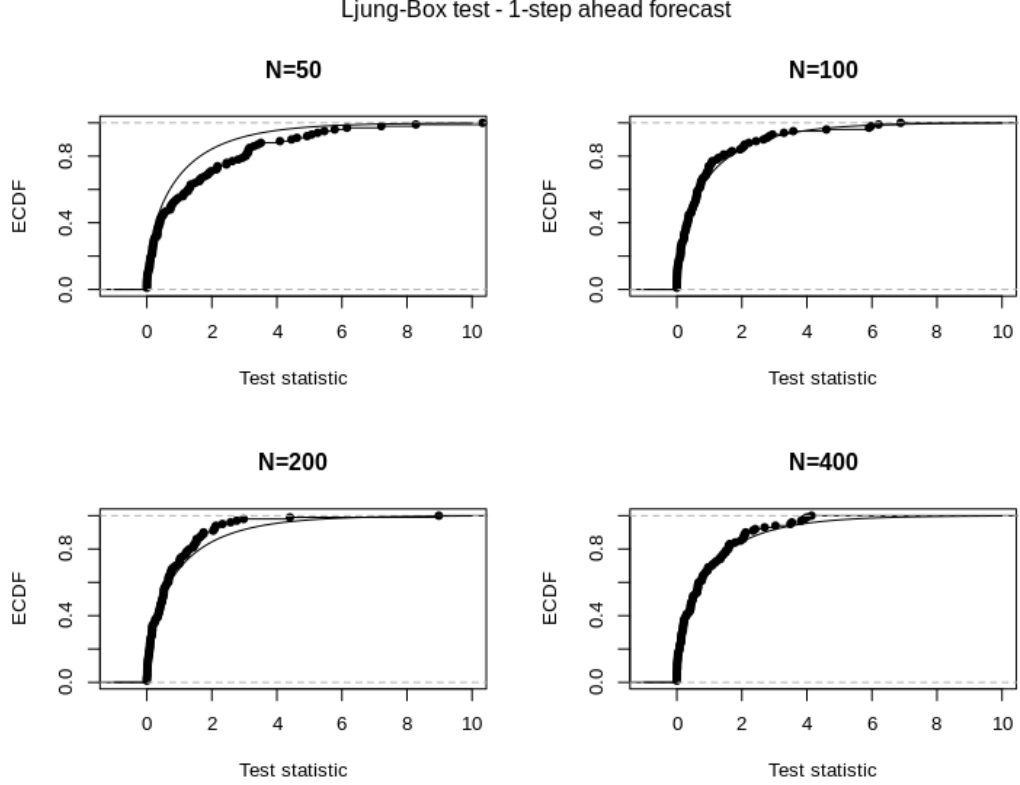


Figure 1: ECDF of the test statistic of the Ljung-Box test of independence, for 1-step ahead forecasting. The continuous line represents the true CDF under the null hypothesis.

Tables 1 and 2 show the average p-value of both tests over the 100 runs for different particle sample sizes N and for 1-step ahead and 5-step ahead forecasting. The average p-values are observed to converge to 0.5 when N increases, which is the expected value under the null hypothesis.

Thus, the null hypotheses of the Ljung-Box test of independence and the Kolmogorov-Smirnov test of uniformity tend to be more statistically likely when the number of particles increases. This empirically validates the converge of particle forecasting to the optimal forecast.

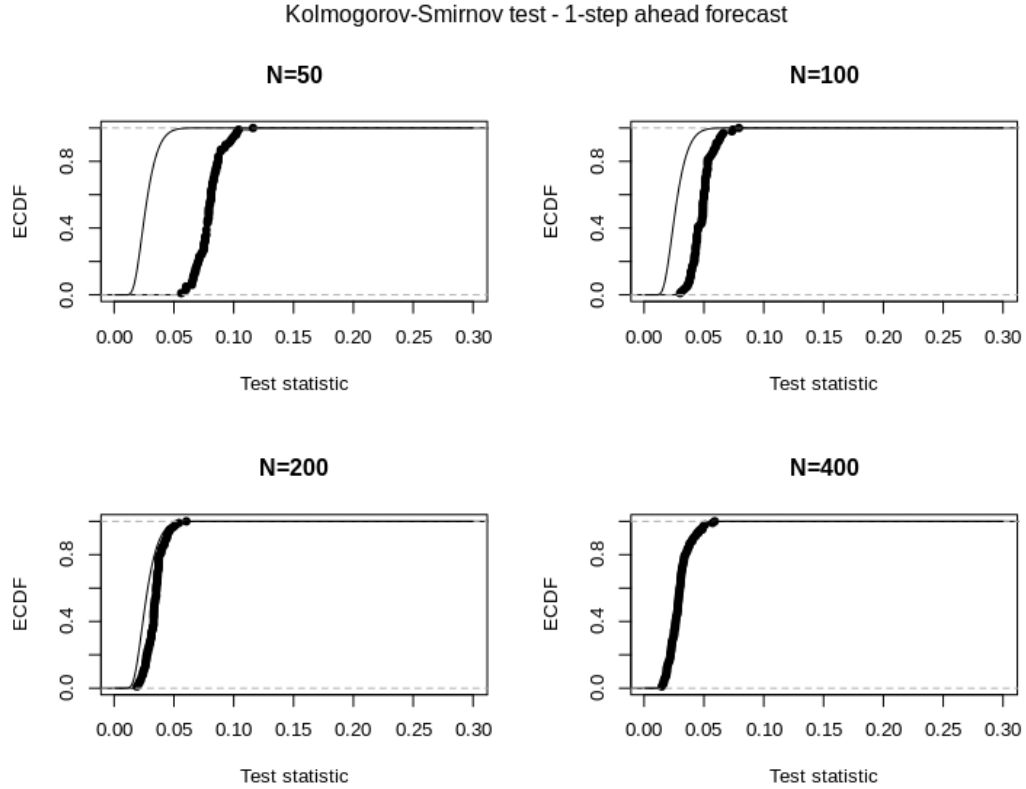


Figure 2: ECDF of the test statistic of the Kolmogorov-Smirnov test of uniformity, for 1-step ahead forecasting. The continuous line represents the true CDF under the null hypothesis.

6 Conclusion

In this article, we propose a novel forecasting algorithm for state-space models, particle forecasting, which is an extension of particle filtering. We obtain convergence results, almost surely and in mean squared error, for particle forecasting and we review results for particle filtering. Simulations are run to illustrate the convergence of particle filtering, using performance assessment techniques based on probability integral transform and adapted to distributional forecasts.

An application example for particle forecasting is value-at-risk (VaR) forecasting in stochastic volatility (SV) models for financial risk management [27]. SV models are state-space models describing time series of financial asset returns. VaR is a measure of risk based on quantiles of forecasts, hence the importance of having distributional forecasts, from which quantiles can be extracted.

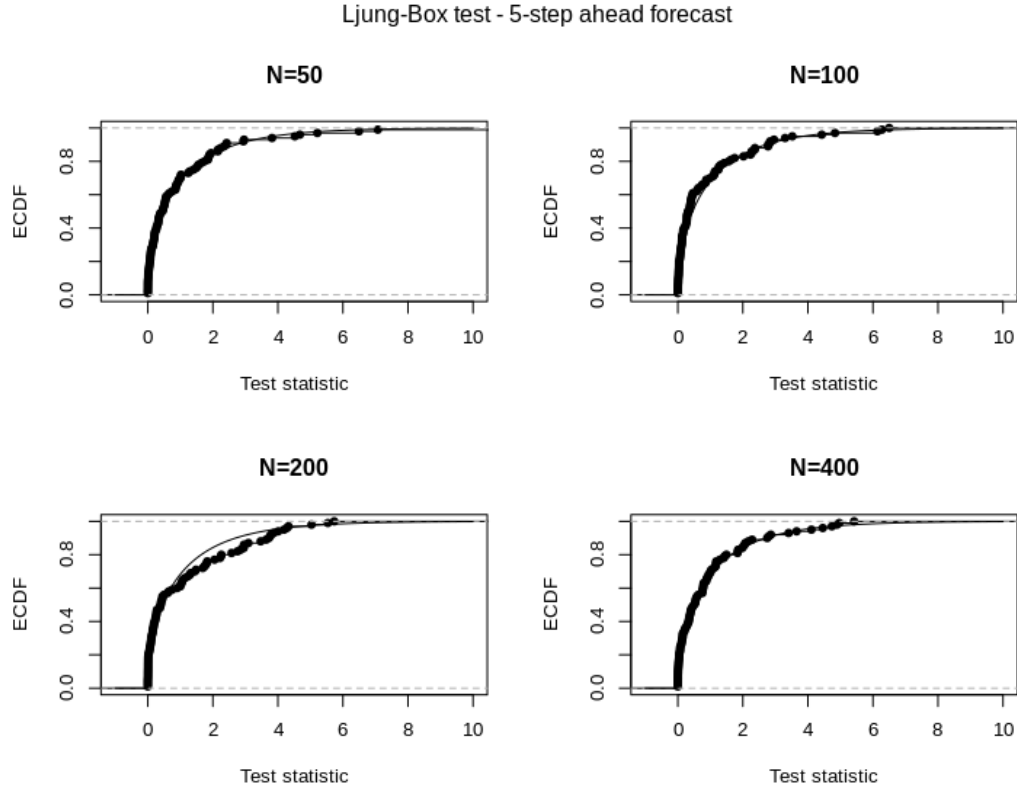


Figure 3: ECDF of the test statistic of the Ljung-Box test of independence, for 5-step ahead forecasting. The continuous line represents the true CDF under the null hypothesis.

Particle forecasting produces converging Monte Carlo approximations of the optimal forecast when the model is known. In practice, when working with real datasets, model specification strongly impacts forecasting performance. Particle forecasting can yield arbitrarily small approximation error, but it does not address the modelling error between the true data-generating process and the model.

Besides, particle forecasting relies on the possibility to generate samples from the observation model. In some settings, the conditional distribution of observations given the hidden state is intractable, which has motivated the development of approximate Bayesian computation [28]. In this context, particle forecasting cannot be straightforwardly applied and requires some adjustments.

Issues to investigate include the application of particle forecasting to real problems, especially those necessitating forecasts in the form of distributions [12], like VaR forecasting. On the theoretical side, convergence in distribution

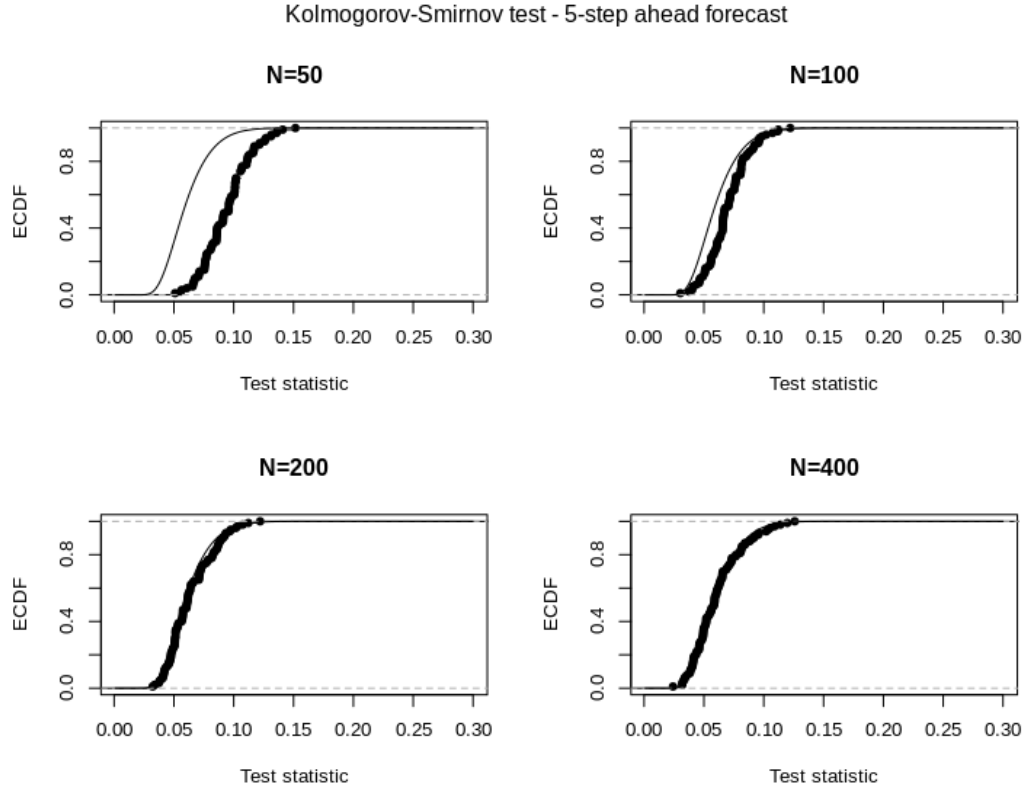


Figure 4: ECDF of the test statistic of the Kolmogorov-Smirnov test of uniformity, for 5-step ahead forecasting. The continuous line represents the true CDF under the null hypothesis.

of particle forecasting remains to be assessed, typically with a central limit theorem with closed-formed asymptotic variance, as it is already available for particle filtering [29].

References

References

- [1] R. J. Hyndman, G. Athanasopoulos, Forecasting: principles and practice, OTexts, 2018.
URL <https://otexts.org/fpp2/>
- [2] S. J. Taylor, Modelling Financial Time Series, John Wiley, 1986.

| | | h | |
|-----|-----|------|------|
| | | 1 | 5 |
| N | 50 | 0.41 | 0.50 |
| | 100 | 0.49 | 0.53 |
| | 200 | 0.51 | 0.50 |
| | 400 | 0.50 | 0.50 |

Table 1: Average p-value of the Ljung-Box test of independence.

| | | h | |
|-----|-----|------|------|
| | | 1 | 5 |
| N | 50 | 0.00 | 0.11 |
| | 100 | 0.04 | 0.35 |
| | 200 | 0.26 | 0.46 |
| | 400 | 0.44 | 0.52 |

Table 2: Average p-value of the Kolmogorov-Smirnov test of uniformity.

- [3] P. Bui Quang, C. Musso, F. Le Gland, Particle filtering and the Laplace method for target tracking, *IEEE Transactions on Aerospace and Electronic Systems* 52 (2016).
- [4] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, P.-J. Nordlund, Particle filters for positioning, navigation and tracking, *IEEE Transactions on Signal Processing* 50 (2002).
- [5] O. Cappé, E. Moulines, T. Rydén, *Inference in Hidden Markov Models*, Springer, 2005.
- [6] N. Kantas, A. Doucet, S. S. Singh, J. Maciejowski, N. Chopin, On particle methods for parameter estimation in state-space models, *Statistical Science* 30 (2015).
- [7] S. Särkkä, *Bayesian Filtering and Smoothing*, Cambridge University Press, 2013.
- [8] G. Petris, S. Petrone, P. Campagnoli, *Dynamic Linear Models with R*, Springer, 2009.
- [9] M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Transactions on Signal Processing* 50 (2002).
- [10] A. Doucet, S. Godsill, C. Andrieu, On sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing* 10 (2000).
- [11] A. Doucet, A. M. Johansen, A tutorial on particle filtering and smoothing: Fifteen years later, in: D. Crisan, B. Rozovskii (Eds.), *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press, 2011, pp. 656–704.

- [12] T. Gneiting, M. Katzfuss, Probabilistic forecasting, *Annual Review of Statistics and Its Application* 1 (2014).
- [13] D. Crisan, A. Doucet, A survey of convergence results on particle filtering methods for practitioners, *IEEE Transactions on Signal Processing* 50 (2002).
- [14] R. E. Kalman, A new approach to linear filtering and prediction problems, *Journal of Basic Engineering* 82 (1960).
- [15] L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE* 77 (1989).
- [16] O. Cappé, S. J. Godsill, E. Moulines, An overview of existing methods and recent advances in sequential Monte Carlo, *Proceedings of the IEEE* 95 (2007).
- [17] N. J. Gordon, D. J. Salmond, A. F. M. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE-Proceedings-F* 140 (1993).
- [18] P. Del Moral, Nonlinear filtering: Interacting particle resolution, *Markov Processes and Related Fields* 2 (1996).
- [19] P. Del Moral, Measure-valued processes and interacting particle systems. Application to nonlinear filtering problems, *Annals of Applied Probability* 8 (1998).
- [20] R. Douc, E. Moulines, Limit theorems for weighted samples with applications to sequential monte carlo methods, *The Annals of Statistics* 36 (2008).
- [21] P. J. Huber, *Robust Statistics*, John Wiley & Sons, 1981.
- [22] G. Kitagawa, Non-Gaussian state-space modeling of nonstationary time series, *Journal of the American Statistical Association* 82 (1987).
- [23] G. Kitagawa, Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics* 5 (1996).
- [24] F. X. Diebold, T. A. Gunther, A. S. Tay, Evaluating density forecasts with applications to financial risk management, *International Economic Review* 39 (1998).
- [25] G. M. Ljung, G. E. P. Box, On a measure of lack of fit in time series models, *Biometrika* 65 (1978).
- [26] G. Marsaglia, W. W. Tsang, J. Wang, Evaluating Kolmogorov's distribution, *Journal of Statistical Software* 8 (2003).

- [27] P. Bui Quang, T. Klein, N. H. Nguyen, T. Walther, Value-at-Risk for South-East Asian stock markets: Stochastic volatility vs. GARCH, Journal of Risk and Financial Management 11 (2018).
- [28] A. Jasra, S. S. Singh, J. S. Martin, E. McCoy, Filtering via approximate Bayesian computation, Statistics and Computing 22 (2012).
- [29] N. Chopin, Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference, The Annals of Statistics 32 (2004).

Appendix

6.1 Proofs of Lemmas 1 and 2

Proof of Lemma 1.

$$\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu = \int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N}) + \int \phi \nu^N(V_{1:N}) - \int \phi \nu,$$

where it is assumed that $\int \phi \nu^N(V_{1:N}) - \int \phi \nu \xrightarrow[N \rightarrow \infty]{} 0$ almost surely. Then one must prove that $\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N}) \xrightarrow[N \rightarrow \infty]{} 0$ almost surely.

One has that

$$\mathbb{E}[(\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N}))^4] = \mathbb{E}[(\frac{1}{N} \sum_{i=1}^N \phi(U^i) - \int \phi \nu^N(V_{1:N}))^4 \mid V_{1:N}]$$

where $U_1, \dots, U_N \mid V_{1:N} \stackrel{\text{i.i.d.}}{\sim} \nu^N(V_{1:N})$, so that

$$\begin{aligned} & \mathbb{E}[(\frac{1}{N} \sum_{i=1}^N \phi(U_i) - \int \phi \nu^N(V_{1:N}))^4 \mid V_{1:N}] \\ &= \frac{1}{N^4} \sum_i \mathbb{E}[(\phi(U_i) - \int \phi \nu^N(V_{1:N}))^4 \mid V_{1:N}] \\ & \quad + \frac{3}{N^4} \sum_{i \neq j} \mathbb{E}[(\phi(U_i) - \int \phi \nu^N(V_{1:N}))^2 \mid V_{1:N}] \mathbb{E}[(\phi(U_j) - \int \phi \nu^N(V_{1:N}))^2 \mid V_{1:N}] \\ &\leq \frac{N(2|\sup \phi|)^4}{N^4} + \frac{3N(N-1)(2|\sup \phi|)^4}{N^4} \\ &\leq \frac{48|\sup \phi|^4}{N^2}. \end{aligned}$$

Thus

$$\mathbb{E}[(\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N}))^4] \leq \frac{48|\sup \phi|^4}{N^2}.$$

For all $\varepsilon > 0$,

$$\begin{aligned}
& \sum_{N=1}^{\infty} \mathbb{P}[|\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N})| \geq \varepsilon] \\
& \leq \sum_{N=1}^{\infty} \frac{\mathbb{E}[(\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N}))^4]}{\varepsilon^4} \\
& \leq \frac{48|\sup \phi|^4}{\varepsilon^4} \sum_{N=1}^{\infty} \frac{1}{N^2} < \infty,
\end{aligned}$$

by the Markov inequality. From the Borel-Cantelli lemma, one can conclude that $\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N}) \xrightarrow[N \rightarrow \infty]{} 0$ almost surely. \square

Proof of Lemma 2. One has that

$$\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu = \int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N}) + \int \phi \nu^N(V_{1:N}) - \int \phi \nu.$$

According to the Minkowski inequality,

$$\begin{aligned}
\mathbb{E}[(\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu)^2]^{1/2} & \leq \mathbb{E}[(\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N}))^2]^{1/2} \\
& \quad + \mathbb{E}[(\int \phi \nu^N(V_{1:N}) - \int \phi \nu)^2]^{1/2}
\end{aligned}$$

where it is assumed that $\mathbb{E}[(\int \phi \nu^N(V_{1:N}) - \int \phi \nu)^2]^{1/2} \leq \sqrt{\frac{m}{N}}$.

Then

$$\mathbb{E}[(\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N}))^2] = \mathbb{E}[\mathbb{E}[(\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N}))^2 \mid V_{1:N}]]$$

and

$$\begin{aligned}
& \mathbb{E}[(\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N}))^2 \mid V_{1:N}] \\
& = \mathbb{E}[(\frac{1}{N} \sum_{i=1}^N \phi(U_i) - \int \phi \nu^N(V_{1:N}))^2 \mid V_{1:N}] \\
& = \frac{1}{N} (\int \phi^2 \nu^N(V_{1:N}) - (\int \phi \nu^N(V_{1:N}))^2) \\
& \leq \frac{2|\sup \phi|^2}{N}
\end{aligned}$$

where $U_1, \dots, U_N \mid V_{1:N} \stackrel{\text{i.i.d.}}{\sim} \nu^N(V_{1:N})$, thus

$$\mathbb{E}[(\int \phi s^N(\nu^N(V_{1:N})) - \int \phi \nu^N(V_{1:N}))^2]^{1/2} \leq \frac{\sqrt{2}|\sup \phi|}{\sqrt{N}}.$$

\square

6.2 Proofs of Propositions 1, 2 and 3

Proof of Proposition 1. Let ϕ be a continuous bounded function. Let $\{\nu^N\}_N$ be a converging sequence of probability measures with limit ν in the weak topology. One has that

$$\begin{aligned}\int \phi \text{ pred}_t(\nu^N) &= \int \phi(x') \left[\int Q_t(x, dx') \nu^N(dx) \right] \\ &= \int \left[\int \phi(x') Q_t(x, dx') \right] \nu^N(dx) \\ &\xrightarrow[N \rightarrow \infty]{} \int \left[\int \phi(x') Q_t(x, dx') \right] \nu(dx)\end{aligned}$$

as $\int \phi(x_t) Q_t(\cdot, dx_t)$ is continuous and bounded from Assumption 1 and $\nu^N \xrightarrow[N \rightarrow \infty]{} \nu$

$$\begin{aligned}&= \int \phi(x') \left[\int Q_t(x, dx') \nu(dx) \right] \\ &= \int \phi \text{ pred}_t(\nu).\end{aligned}$$

□

Proof of Proposition 2. Let ϕ be a continuous bounded function. Let $\{\nu^N\}_N$ be a converging sequence of probability measures with limit ν in the weak topology. One has that

$$\int \phi \text{ upd}_t(\nu^N) = \frac{\int \phi(x) g_t(Y_t | x) \nu^N(dx)}{\int g_t(Y_t | x) \nu^N(dx)}.$$

In this ratio, from Assumption 2, the numerator verifies

$$\int \phi(x) g_t(Y_t | x) \nu^N(dx) \xrightarrow[N \rightarrow \infty]{} \int \phi(x) g_t(Y_t | x) \nu(dx)$$

because ϕg is continuous and bounded and the denominator verifies

$$\int g_t(Y_t | x) \nu^N(dx) \xrightarrow[N \rightarrow \infty]{} \int g_t(Y_t | x) \nu(dx) > 0$$

because g is continuous, bounded and strictly positive. As the limit of the denominator is nonzero, one has that

$$\int \phi \text{ upd}_t(\nu^N) \xrightarrow[N \rightarrow \infty]{} \frac{\int \phi(x) g_t(Y_t | x) \nu(dx)}{\int g_t(Y_t | x) \nu(dx)} = \int \phi \text{ upd}_t(\nu).$$

□

Proof of Proposition 3. Let ϕ be a continuous bounded function. Let $\{\nu^N\}_N$ be a converging sequence of probability measures with limit ν in the weak topology. One has that

$$\int \phi \text{ for}_t(\nu^N) = \int \phi(y) \left[\int g_t(y | x) \nu^N(dx) \right] \lambda(dy) = \int \left[\int \phi(y) g_t(y | x) \lambda(dy) \right] \nu^N(dx)$$

where $\int \phi(y) g_t(y | \cdot) \lambda(dy)$ is continuous from Assumption 3 and bounded by $\sup |\phi|$. Thus

$$\int \phi \text{ for}_t(\nu^N) \xrightarrow{N \rightarrow \infty} \int \left[\int \phi(y) g_t(y | x) \lambda(dy) \right] \nu(dx)$$

where the limit equals $\int \phi(y) \left[\int g_t(y | x) \nu(dx) \right] \lambda(dy) = \int \phi \text{ for}_t(\nu)$. \square

6.3 Probability Integral Transform

Let $F_{t+h|t}$ be the cdf of the optimal forecast $f_{t+h|t}$, i.e. the conditional cdf of Y_{t+h} given $Y_{0:t}$.

The PIT $F_{t+h|t}(Y_{t+h})$ is a random variable taking values in $[0, 1]$ equipped with its Borelian set. One has that

$$\mathbb{E}[\phi(F_{t+h|t}(Y_{t+h})) | Y_{0:t}] = \int_0^1 (\phi \circ F_{t+h|t}) f_{t+h|t} = \int_0^1 \phi(f_{t+h|t} \circ F_{t+h|t}^{-1})$$

where the pushforward measure $f_{t+h|t} \circ F_{t+h|t}^{-1}$ is the Lebesgue measure. Indeed, let $[a, b[\subset [0, 1]$ ($a < b$) and $a', b' \in \mathbb{R}$ such that $F_{t+h|t}(a') = a$ and $F_{t+h|t}(b') = b$, then $f_{t+h|t}(F_{t+h|t}^{-1}([a, b[)) = f_{t+h|t}([a', b'[) = F_{t+h|t}(b') - F_{t+h|t}(a') = b - a$.

Thus $F_{t+h|t}(Y_{t+h})$ and $Y_{0:t}$ are independent and $F_{t+h|t}(Y_{t+h}) \sim \mathcal{U}(0, 1)$.