Bayesian Inference and Data Assimilation

Prof. Dr.-Ing. Sebastian Reich

Universität Potsdam

12 April 2021

4 Stochastic processes

In the previous two chapters, we discussed how probability measures can be used to express our uncertainty in estimating quantities.

Having obtained a probability measure expressing our uncertainty in the system state at a particular snapshot in time, we would like to know how our uncertainty evolves in time. This motivates a probabilistic view of dynamical systems. It is also often the case that the models that we use are themselves random, often because they approximately represent some complex dynamical process by a simple random process.

We can use the language of stochastic processes to describe such models and incorporate them into the probabilistic dynamical systems framework.

4.1 Dynamical systems and discrete-time Markov processes

Definition (Dynamical system)

Given a phase space $\mathcal{B} \subset \mathbb{R}^{N_z}$ embedded into a N_z-dimensional Euclidian space, an autonomous dynamical system is a map $\psi : \mathcal{B} \to \mathcal{B}$. Given an initial condition $z^0 \in \mathcal{B}$, a dynamical system defines a solution sequence or *trajectory* $\{z^n\}_{n \geq 0}$ via the iteration

$$z^{n+1} = \psi(z^n).$$

If the map ψ also depends on the iteration index n, then we call it a non-autonomous dynamical system.

Example (Lorenz-63 model)

Using this terminology, we find that

$$\psi(z) = z + \delta t f(z), \tag{1}$$

induces an autonomous dynamical system while

$$\psi_n(z) = z + \delta t \left(f(z) + g(t_n) \right), \qquad t_n = n\delta t, \tag{2}$$

constitutes a non-autonomous system for given $g(t_n)$. Here f denotes the right hand side of the Lorenz-63 model and $\delta t > 0$ is a step-size.

Recall from the Epiloge that the autonomous formulation (1) has been considered as a simplified model of the non-autonomous process (2) with $g(t_n)$ representing a model error.

Remark (Set-oriented dynamics)

From a forecasting perspective, if we do not know the precise value of the system state z, but instead only know that it lies in some subset $\mathcal{A}$ of phase space, we would like to know the minimum subset definitely containing the system state at later times.

More generally, this is the setting for many questions of interest about the properties of dynamical systems. The theory of dynamical systems is less concerned about the behaviour of individual trajectories and instead focuses on the collective or typical behaviour of many trajectories.

For example, we might ask whether trajectories remain bounded or whether they enter a compact subset of phase space after finitely many iterations.

Definition (Invariant sets)

Let us consider sequences of sets

$$\mathcal{A}^{n+1} = \psi(\mathcal{A}^n), \qquad \mathcal{A}^0 = \mathcal{A},$$

where $z^{n+1} \in \mathcal{A}^{n+1}$ if and only if there exists a $z^n \in \mathcal{A}^n$ such that $z^{n+1} = \psi(z^n)$.

A subset $\mathcal{A}$ of phase space is now called forward invariant if

$$\psi(\mathcal{A}) \subseteq \mathcal{A}.$$

In other words, if $z^{n_*} \in \mathcal{A}$ for some n_*, then $z^n \in \mathcal{A}$ for all $n \geq n_*$.

Example (Lorenz-63 model)

We recall the Lorenz-63 system and its discretisation in time by the forward Euler method, which gives rise to an autonomous dynamical system.

A first observation is that $f(0) = 0$ and hence $z^n = 0$ for all $n > 0$ provided $z^0 = 0$. Therefore, $\mathcal{A} = \{0\}$ is forward invariant.

However this is not an "interesting" forward invariant set since any initial condition $z^0 \neq 0$ with $\|z^0\|$ arbitrarily small will be repelled from the origin and will ultimately approach a much more interesting forward invariant set, which is called the Lorenz attractor.

A numerical demonstration of "attraction" is provided on the next slide, where one thousand solutions with initial conditions close to zero are displayed at $t = 10$ and $t = 100$.

The solutions approach the Lorenz attractor as time progresses.

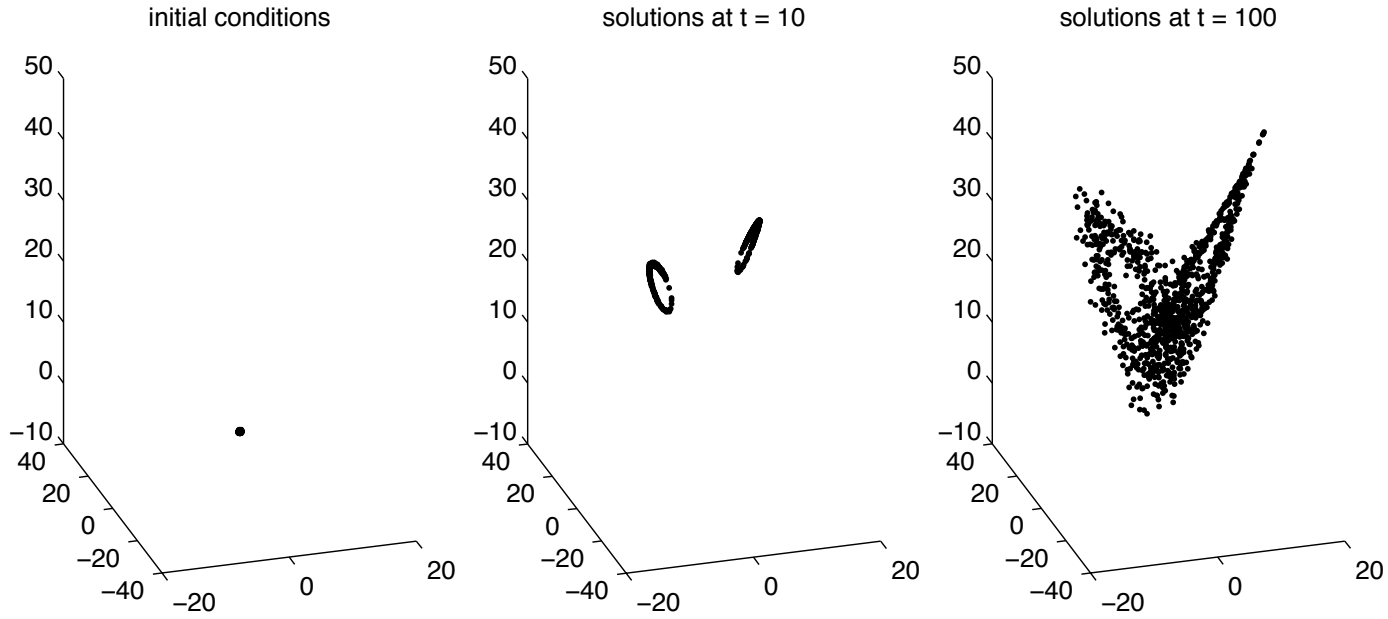


Figure: Snapshots of $M = 1000$ solutions starting near the origin at time zero (left panel) at time $t = 10$ (middle panel) and at $t = 100$. The solutions start approaching a set closely resembling the set computed in Section 1.

Definition (Distance from a point to a set)

For $z \in \mathcal{B}$, the distance from z to a subset $\mathcal{A} \subset \mathcal{B}$ is

$$d(z, \mathcal{A}) := \inf_{z' \in \mathcal{A}} \|z - z'\|.$$

Definition (Attractor)

A compact set $\mathcal{A} \subset \mathcal{B}$ is called an attractor of a dynamical system ψ if

(i) $\mathcal{A}$ is forward invariant;

(ii) $\mathcal{A}$ attracts a neighborhood of itself, *i.e.*, there exists $\delta > 0$ such that if $d(z^0, \mathcal{A}) < \delta$, then $d(z^n, \mathcal{A}) \to 0$, as $n \to \infty$; and

(iii) $\mathcal{A}$ is minimal, *i.e.*, there is no $\mathcal{A}' \subset \mathcal{A}$ such that both $\mathcal{A}$ and $\mathcal{A}'$ are forward invariant and attracting.

Example (Tent map)

Let us consider the dynamical system defined by $\mathcal{B} = [0,1] \subset \mathbb{R}$ and the tent map

$$\psi(z) := \begin{cases} 2z & \text{if } z \in [0, 1/2], \\ 2 - 2z & \text{otherwise.} \end{cases}$$

If we define $\mathcal{A}^0 = [0, 1/2]$, for example, then we find that $\mathcal{A}^1 = \mathcal{B} = [0,1]$ after a single iteration. Clearly $\mathcal{B}$ is forward invariant.

The choice $\mathcal{A}^0 = \{0, 1\}$ leads, on the other hand, after a single iteration to $\mathcal{A}^1 = \{0\}$ which itself is forward invariant. We find that $\mathcal{A} = \{0\}$ is attracting for certain initial conditions, since any $z^0 = 1/2^k$, $k = 0, 1, 2, \ldots$, will result in trajectories reaching $\{0\}$ after finitely many steps (in fact $k + 1$ iterations).

At the same time, there is no neighborhood $\mathcal{U}$ of $\{0\}$ in $[0,1]$ such that all $z^0 \in \mathcal{U}$ approach $\{0\}$ and, hence, $\{0\}$ is not an attractor as defined above.

Remark (From sets to probabilities and stochastic processes)

From the perspective of forecasting, rather than simply specifying a set that we think contains z, it is more useful to quantify our uncertainty about where z is by assigning probabilities to subsets $\mathcal{A} \subset \mathcal{B}$.

We can then compare subsets based on where z is most likely to be found. We will denote $\mathbb{P}^n(\mathcal{A})$ the probability of z being in set $\mathcal{A}$ at time n.

Once we have assigned probabilities $\mathbb{P}^0(\mathcal{A})$ to subsets at $n = 0$, their evolution is entirely determined by the dynamical system since

$$\mathbb{P}^{n+1}(\mathcal{A}) = \mathbb{P}^n(\mathcal{A}'), \quad n \geq 0,$$

with $\mathcal{A}' = \psi^{-1}(\mathcal{A})$ for all appropriate sets $\mathcal{A} \subset \mathcal{B}$.

We have already encountered this transformation rule (or conservation law) for probabilities in Chapter 2.

Using the language of probability theory, we define a PDF π_{Z^0} and a random variable $Z^0 : \Omega \to \mathcal{B}$ on phase space $\mathcal{B} \subset \mathbb{R}^{N_z}$ such that

$$\mathbb{P}^0(\mathcal{A}) = \int_{\mathcal{A}} \pi_{Z^0}(z^0) \mathrm{d}z^0.$$

After a single application of the dynamical system, we then obtain a random variable $Z^1 = \psi(Z^0)$, with marginal PDF π_{Z^1} defined by

$$\mathbb{P}^1(\mathcal{A}) = \int_{\mathcal{A}} \pi_{Z^1}(z^1) \mathrm{d}z^1 = \int_{\psi^{-1}(\mathcal{A})} \pi_{Z^0}(z^0) \mathrm{d}z^0,$$

for all subsets $\mathcal{A} \subseteq \mathcal{B}$.

The two variables Z^0 and Z^1 posses a natural coupling

$$\pi_{Z^1 Z^0}(z^0, z^1) = \delta(z^1 - \psi(z^0))\, \pi_{Z^0}(z^0)\,.$$

Example (Linear dynamical systems)

We consider the scalar, colorbluelinear iteration

$$z^{n+1} = dz^n + b,$$

with phase space $\mathcal{B} = \mathbb{R}$ and fixed parameters d and b. Assume that z^0 is obtained from a Gaussian random variable Z^0 with mean $\bar{z}^0$ and variance σ_0^2 on the initial conditions z^0. The PDF for Z^0 is

$$\pi_{Z^0}(z) = \mathrm{n}(z; \bar{z}^0, \sigma_0^2) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2}\left(\frac{z-\bar{z}^0}{\sigma_0}\right)^2}.$$

The random variable Z^1 is now defined by $Z^1 = dZ^0 + b$, which is again Gaussian with mean

$$\bar{z}^1 = \mathbb{E}[dZ^0 + b] = d\bar{z}^0 + b,$$

and variance

$$\sigma_1^2 = \mathbb{E}[(dZ^0 + b - \bar{z}^1)^2] = d^2\mathbb{E}[(Z^0 - \bar{z}^0)^2] = d^2\sigma_0^2.$$

Example (Continued)

By induction, Z^n is Gaussian for $n \geq 0$, and the mean and variance are obtained from the recursions

$$\bar{z}^{n+1} = d\bar{z}^n + b, \quad \sigma^2_{n+1} = d^2\sigma^2_n.$$

We denote the marginal distribution of Z^n by $\pi_{Z^n}(z) = \mathrm{n}(z; \bar{z}^n, \sigma^2_n)$.

Note that Z^n and Z^{n+1} are fully correlated since

$$\mathbb{E}[(Z^n - \bar{z}^n)(Z^{n+1} - \bar{z}^{n+1})] = \mathbb{E}[d(Z^n - \bar{z}^n)(Z^n - \bar{z}^n)] = d\sigma^2_n,$$

and the correlation becomes

$$\mathrm{corr}(Z^n, Z^{n+1}) = \frac{\mathbb{E}[(Z^n - \bar{z}^n)(Z^{n+1} - \bar{z}^{n+1})]}{\mathbb{E}[(Z^n - \bar{z}^n)^2]^{1/2}\,\mathbb{E}[(Z^{n+1} - \bar{z}^{n+1})^2]^{1/2}} = \begin{cases} 1 & \text{if } d > 0, \\ -1 & \text{if } d < 0. \end{cases}$$

This is, of course, not surprising since Z^n and Z^{n+1} are connected by a deterministic linear map.

Example (Continued)

These considerations can be generalised to multivariate iterations

$$z^{n+1} = Dz^n + b,$$

$z^n \in \mathbb{R}^{N_z}$, in which case the mean of the associated random variables Z^n is recursively given by

$$\bar{z}^{n+1} = D\bar{z}^n + b,$$

and the covariance matrices P^n follow the recursion

$$\begin{aligned} P^{n+1} &= \mathbb{E}[(Z^{n+1} - \bar{z}^{n+1})(Z^{n+1} - \bar{z}^{n+1})^T], \\ &= D\mathbb{E}[(Z^n - \bar{z}^n)(Z^n - \bar{z}^n)^T]D^T, \\ &= DP^nD^T. \end{aligned}$$

Furthermore, Gaussian PDFs remain Gaussian under any linear transformation and the marginal distribution of Z^n is given by the PDF

$$\pi_{Z^n}(z) = \mathrm{n}(z; \bar{z}^n, P^n) = \frac{1}{(2\pi)^{N_z/2}|P^n|^{1/2}} \exp\left(-\frac{1}{2}(z - \bar{z}^n)^T(P^n)^{-1}(z - \bar{z}^n)\right).$$

Example (Non-autonomous forcing)

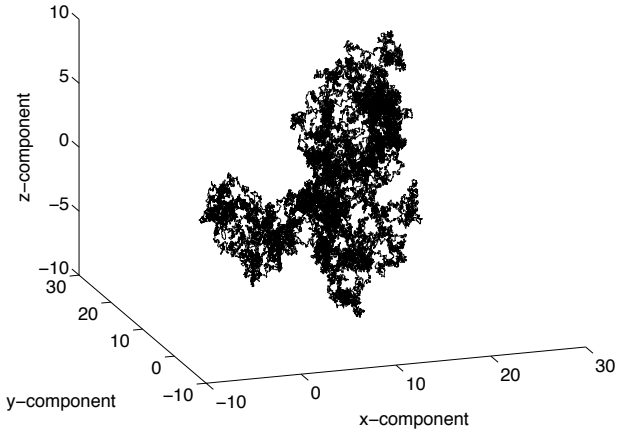
We consider the iteration

$$\mathrm{x}^{n+1} = \mathrm{x}^n + \delta t g_1^n,\ \mathrm{y}^{n+1} = \mathrm{y}^n + \delta t g_2^n,\ \mathrm{z}^{n+1} = \mathrm{z}^n + \delta t g_3^n,$$

in $\mathbb{R}^3$ with step-size $\delta t = 0.001$, initial condition $\mathrm{x}^0 = \mathrm{y}^0 = \mathrm{z}^0 = 0$, and with the non-autonomous forcing $g(t_n) = g^n = (g_1^n, g_2^n, g_3^n)^T \in \mathbb{R}^3$ generated as follows: Set $a = 1/\sqrt{\delta t}$ and, for $n \geq 0$ define recursively

$$g_i^{n+1} = \begin{cases} 1.99999 g_i^n + a/2 & \text{if } g_i^n \in [-a/2, 0), \\ -1.99999 g_i^n + a/2 & \text{otherwise.} \end{cases}$$

random motion generated by modified tent map

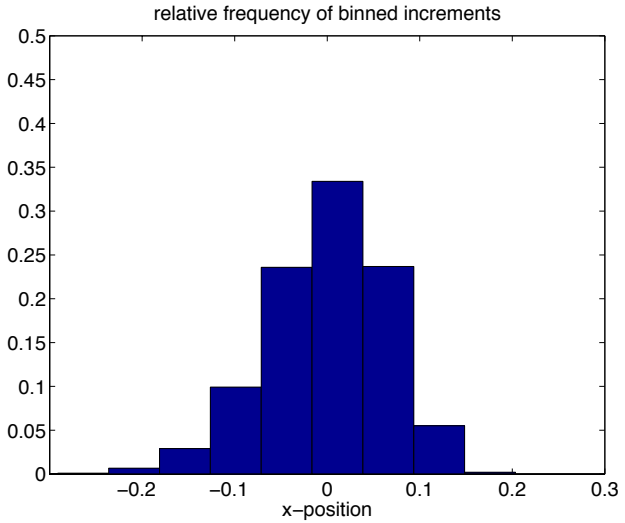


Example (Continued)

The trajectory clearly displays "random walk" behaviour in $\mathbb{R}^3$. In order to analyse this phenomenon in more detail we define solution increments

$$\Delta \mathrm{x}^k = \mathrm{x}(\mathrm{t}_k) - \mathrm{x}(\mathrm{t}_{k-1}),$$

for $k \geq 1$ with corresponding expressions for $\Delta \mathrm{y}^k$ and $\Delta \mathrm{z}^k$. Relative frequencies of the $\Delta \mathrm{x}^k$ values are displayed in the figure below. We approximate this distribution by a Gaussian with mean zero and variance $\sigma_x^2 \approx 0.0838 \Delta t_{\mathrm{out}}$.



Example (Continued)

We also compute the (normalised) autocorrelation coefficients

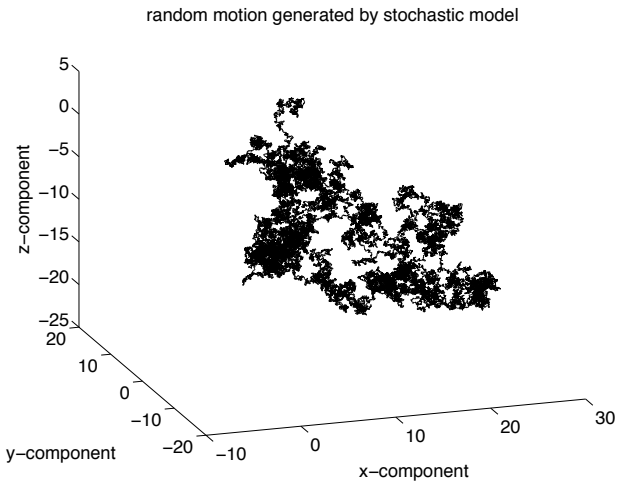
$$C(\tau) = \frac{\sum_{k\geq 1} \Delta \mathrm{x}^k \Delta \mathrm{x}^{k+\tau}}{\sum_{k\geq 1} \Delta \mathrm{x}^k \Delta \mathrm{x}^k},$$

for $\tau = 0, 1, \ldots, 10$. By definition, $C(0) = 1$ and all other coefficients are found to be smaller than 10^{-2} in absolute value. In other words, increments can be treated as mutually uncorrelated.

These findings suggest that we could approximate our iteration with a stochastic difference equation

$$\mathrm{x}(\mathrm{t}_{k+1}) = \mathrm{x}(\mathrm{t}_k) + \xi^k,$$

where the increments $\{\xi^k\}$ are realisations of independent and Gaussian distributed random variables $\{\Xi^k\}$ with mean zero and variance $\sigma_x^2 \approx 0.0838 \Delta t_{\mathrm{out}}$.



The previous example, leads us from deterministic dynamical systems

$$Z^{n+1} = \psi(Z^n)$$

to dynamic processes where the future value Z^{n+1} is not uniquely determined by the current Z^n.

Example (Random walk)

We consider a sequence of univariate random variables $\{Z^n\}_{n\geq 0}$ defined by the recursion

$$Z^{n+1} = Z^n + \Xi^n, \quad n = 0, 1, 2, \ldots \tag{3}$$

Here $\Xi^n \sim \mathrm{N}(0, 1)$ are assumed to be independent with respect to each other and to $\{Z^k\}_{k=0}^n$ (but not Z^k for $k > n$, of course). We will discuss this recursion probabilistically, which means that we want to understand how the probability measures for the sequence of random variables $\{Z^n\}$ are related.

See Example 4.9 in the textbook for details.

Definition (Stochastic process)

Let T be a set of indices. A stochastic process is a family $\{Z^t\}_{t\in T}$ of random variables $Z^t : \Omega \to \mathcal{Z}$ with joint probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Remark (Continuous vs. discrete time)

The variable t typically corresponds to time; we distinguish between continuous time $t \in [0, t_{\rm end}] \subset \mathbb{R}$ and discrete time $t_n = n\delta t$, $n \in \{0, 1, 2, \ldots\} = T$, with time-increment $\delta t > 0$.

In the discrete time setting we generally prefer Z^n to Z^{t_n}. We will also use the notation $Z(t)$ or $Z(t_n)$, respectively, whenever superscript indices would be more confusing.

Remark (Stochastic processes, random variables & trajectories)

A stochastic process can be seen as a function of two arguments: t and ω.

For fixed $\omega \in \Omega$, $Z^t(\omega)$ becomes a function of $t \in T$, which we call a realisation or trajectory of the stochastic process. In the case of continuous time processes, we will only consider processes for which $Z^t(\omega)$ is continuous in t.

Alternatively, we can fix a time $t \in T$ and consider the random variable $Z^t(\cdot)$ and its distribution (as we did when we obtained the marginal distribution for Z^2 above).

More generally, we can consider l-tuples $(t_1, t_2, \ldots, t_l)$ and associated l-tuples of random variables $(Z^{t_1}(\cdot), Z^{t_2}(\cdot), \ldots, Z^{t_l}(\cdot))$ with their joint distributions. This leads to concepts such as autocorrelation, which we shall discuss later in this chapter.

Definition (Discrete-time Markov processes)

The discrete time stochastic process $\{Z^n\}_{n\in T}$ with $\mathcal{Z} = \mathbb{R}^{N_z}$ and $T = \{0, 1, 2, \ldots)$ is called a (time-independent) Markov process if its joint PDFs can be written as

$$\pi_{Z^0 \ldots Z^n}(z^0, z^1, \ldots, z^n) = \pi(z^n|z^{n-1})\pi(z^{n-1}|z^{n-2}) \cdots \pi(z^1|z^0)\pi_{Z^0}(z^0)$$

for all $n \in \{1, 2, \ldots\} = T$.

Example (Random walk)

Recall the sequence of univariate random variables $\{Z^n\}_{n\geq 0}$ defined by the recursion

$$Z^{n+1} = Z^n + \Xi^n, \quad n = 0, 1, 2, \ldots, \tag{4}$$

where $\Xi^n \sim \mathrm{N}(0, 1)$ are assumed to be independent with respect to each other. Then

$$\pi_{Z^0 \ldots Z^n}(z^0, z^1, \ldots, z^n) = \mathrm{n}(z^n; z^{n-1}, 1) \cdots \mathrm{n}(z^1; z^0, 1)\, \mathrm{n}(z_0; \bar{z}_0, \sigma_0)\,.$$

Definition (Discrete state space random processes)

A finite state space random process $\{Z^n\}_{n\in T}$ is one in which Z^n can only take a finite set of values for each n, taken from the same finite state space $\mathcal{B} \subset \mathbb{R}^{N_z}$. We write $\mathcal{B} = \{a_1, a_2, \ldots, a_M\}$.

The probability distribution for the random variable $Z^n : \Omega \to \mathcal{B}$ is then entirely characterised by an M-tuple of non-negative numbers $p^n(a_i) = \mathbb{P}(Z^n = a_i)$, $i = 1, \ldots, M$, which satisfy $\sum_i p^n(a_i) = 1$.

A discrete-time Markov process is then defined through transition probabilities $p_{ij} = \mathbb{P}(Z^{n+1} = a_i | Z^n = a_j)$, which we collect in an $M \times M$ transition matrix P, which represents the conditional PDF $\pi(z'|z)$.

We obtain the linear recursion

$$p^{n+1}(a_i) = \sum_{j=1}^{M} p_{ij}\, p^n(a_j).$$

A matrix P with entries $p_{ij} = (\mathrm{P})_{ij}$ such that (i) $p_{ij} \geq 0$ and (ii) $\sum_i p_{ij} = 1$ is called a stochastic matrix. By definition, a transition matrix is always a stochastic matrix.

Remark (Chapman-Kolmogorov equation)

The marginal distributions associated with a Markov process,

$$\pi_{Z^n}(z^n) = \int_{\mathcal{Z}} \cdots \int_{\mathcal{Z}} \pi_{Z^0 \ldots Z^n}(z^0, z^1, \ldots, z^n) \mathrm{d}z^0 \cdots \mathrm{d}z^{n-1},$$

i.e., the probability density functions for Z^n, $n = 1, 2, \ldots$, satisfy the Chapman-Kolmogorov equation

$$\pi_{Z^{n+1}}(z') = \int_{\mathcal{Z}} \pi(z'|z) \pi_{Z^n}(z) \mathrm{d}z,$$

from which the marginal distributions can be computed recursively for given initial π_{Z^0} in order to yield the family $\{\pi_{Z^n}\}_{n \in \mathcal{T}}$.

A Markov process is called time-dependent if the conditional PDF $\pi(z'|z)$ depends on the time level n.

Remark (Conditional distributions)

Our definition is equivalent to the more traditional definition which states that a process is Markov if the conditional distributions $\pi_{Z^n}(z^n|z^0, z^1, \ldots, z^{n-1})$, $n \geq 1$, *satisfy*

$$\pi_{Z^n}(z^n|z^0, z^1, \ldots, z^{n-1}) = \pi(z^n|z^{n-1}).$$

We also find that the Chapman-Kolmogorov equation is obtained from the joint PDF

$$\pi_{Z^n Z^{n+1}}(z^n, z^{n+1}) = \pi(z^{n+1}|z^n)\pi_{Z^n}(z^n),$$

followed by marginalisation,

$$\pi_{Z^{n+1}}(z^{n+1}) = \int_{\mathcal{Z}} \pi_{Z^n Z^{n+1}}(z, z^{n+1})\mathrm{d}z,$$

to obtain $\pi_{Z^{n+1}}(z^{n+1})$.

Example (Dynamical systems and Markov processes)

We consider a diffeomorphism $\psi : \mathcal{B} \to \mathcal{B}$ with $\mathcal{B} \subset \mathbb{R}^{N_z}$ and the associated dynamical system

$$z^{n+1} = \psi(z^n)\,.$$

If we view the initial condition as the realisation of a random variable $Z^0 \sim \pi_0$, then the dynamical system gives rise to a Markov process with (formal) transition probability density

$$\pi(z'|z) = \delta(z' - \psi(z)).$$

Indeed

$$\begin{aligned}
\pi_{Z^{n+1}}(z') &= \int_{\mathcal{B}} \pi(z'|z)\,\pi_{Z^n}(z)\mathrm{d}z = \int_{\mathcal{B}} \delta(z' - \psi(z))\,\pi_{Z^n}(z)\mathrm{d}z, \\
&= \int_{\psi^{-1}(\mathcal{B})} \delta(z' - \hat{z})\pi_{Z^n}(\psi^{-1}(\hat{z}))|D\psi^{-1}(\hat{z})|\mathrm{d}\hat{z}, \\
&= \pi_{Z^n}(\psi^{-1}(z'))|D\psi^{-1}(z')|
\end{aligned}$$

which is the formula we had derived previously.

Example (Lorenz-63 model)

We return to our mechanistic model from the Prolog and combine it with our findings from a previous example in order to find a better mechanistic model for our surrogate physical process, in terms of a stochastic Markov process. See Example 4.13 from the textbook for details.

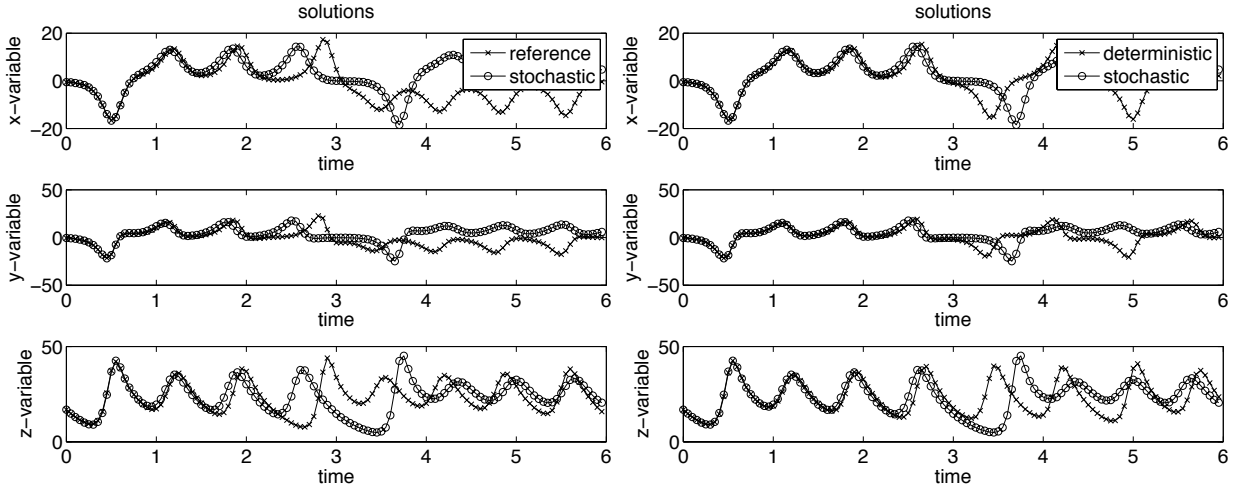


Figure: Comparison between the behaviour of our modified stochastic model and the reference trajectory $z_{\mathrm{ref}}(t)$, under the assumption that both start from the same initial condition at time $t = 0$. We also compare the results from the reduced deterministic and the stochastic model (right panel).

4.2 Stochastic difference and differential equations

We will examine the continuous time limit of stochastic difference models.

We start from the stochastic difference equation

$$Z^{n+1} = Z^n + \delta t\, f(Z^n) + \sqrt{2\delta t}\,\Xi^n, \quad t_{n+1} = t_n + \delta t, \tag{5}$$

where $\delta t > 0$ is a small parameter (the step-size), f is a given (Lipschitz continuous) function, and $\Xi^n \sim \mathrm{N}(0, Q)$ are independent and identically distributed random variables with covariance matrix Q.

Here we have written the stochastic difference equations directly in terms of the involved random variables $\{Z^n\}_{n\geq 0}$.

Equation (5) is a discrete time Markov process, and so the time evolution of the associated marginal densities π_{Z^n} is governed by the Chapman-Kolmogorov equation

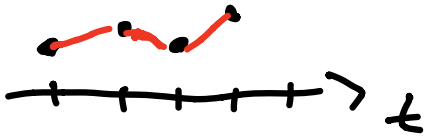
$$\pi_{Z^{n+1}}(z') = \int_{\mathcal{Z}} \pi(z'|z)\pi_{Z^n}(z)\mathrm{d}z,$$

$f(z) = Az + b$

with conditional PDF

$$\pi(z'|z) = \frac{1}{(4\pi\delta t)^{N_z/2}|Q|^{1/2}} \times \exp\left(-\frac{1}{4\delta t}(z' - z - \delta t f(z))^{\mathrm{T}} Q^{-1}(z' - z - \delta t f(z))\right). \quad (6)$$

We wish to investigate the limit $\delta t \to 0$ for fixed final time $T > 0$. This can be achieved by defining $\delta t = T/N$ with $N \to \infty$ and iteration indices $n \in \{0, 1, \ldots, N\}$ in (5).



Example (Discrete-time Brownian motion)

Let us first consider the special case $N_z = 1$, $f(z) = 0$, and $Q = 1$ (univariate Gaussian random variable). We also drop the factor $\sqrt{2}$ in (5) and consider linear interpolation in between time-steps, *i.e.*,

$$Z(t) = Z^n + \sqrt{\delta t}\frac{t - t_n}{t_{n+1} - t_n}\Xi^n, \qquad t \in (t_n, t_{n+1}], \tag{7}$$

setting $Z^0(\omega) = 0$ with probability one.

For each family of realisations $\{\xi^n = \Xi^n(\omega)\}_{n\geq 0}$ we obtain a piecewise linear and continuous function $Z(t, \omega)$.

The limit $\delta t \to 0$ is well defined and the stochastic process $Z(t)$, $t \geq 0$, can be shown to converge weakly to a stochastic process called Brownian motion.

Definition (Brownian motion)

Standard univariate Brownian motion is a stochastic process $W(t)$, $t \geq 0$, with the following properties:

(i) $W(0) = 0$,

(ii) Realisations $W(t, \omega)$ are continuous in t,

(iii) $W(t_2) - W(t_1) \sim \mathrm{N}(0, t_2 - t_1)$ for $t_2 > t_1 \geq 0$, and

(iv) Increments $W(t_4) - W(t_3)$ and $W(t_2) - W(t_1)$ are mutually independent for $t_4 > t_3 \geq t_2 > t_1 \geq 0$.

Remark (Differentiability of Brownian motion)

It follows from our construction (7) that with probability one the Brownian motion is not differentiable, since

$$\frac{Z^{n+1} - Z^n}{\delta t} = \delta t^{-1/2} \Xi^n \sim \mathrm{N}(0, \delta t^{-1})$$

and so the limit $\delta t \to 0$ is not well defined.

Using the concept of Brownian motion, we can rewrite the stochastic difference equations (5) in the form

$$Z^{n+1} = Z^n + \delta t\, f(Z^n) + \sqrt{2} Q^{1/2} \left(W(t_{n+1}) - W(t_n) \right). \tag{8}$$

Z^n, $n \to \infty$ ↑ Chapman-Kolm. $\pi(Z^{n+1} \mid Z^n)$ ↗ weak formal

Theorem (Stochastic differential equation and Fokker-Planck equation)

After taking the limit $\delta t \to 0$ in (8), we obtain the SDE

$$\mathrm{d}Z = f(Z)\mathrm{d}t + \sqrt{2}Q^{1/2}\mathrm{d}W, \tag{9}$$

where $W(t)$ denotes standard N_z-dimensional Brownian motion. The Fokker-Planck equation,

$$\frac{\partial \pi_Z}{\partial t} = -\nabla_z \cdot (\pi_Z f) + \nabla_z \cdot (Q \nabla_z \pi_Z), \tag{10}$$

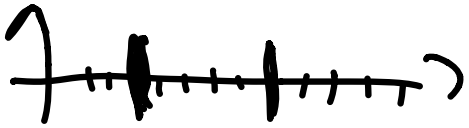
describes the time evolution of the marginal densities $\pi_Z(z,t)$, that is,

$$Z(t) \sim \pi_Z(z,t)\,,$$

and (10) is a continuous-time version of the Chapman-Kolmogorov equation.

Proof.

See proof of Proposition 4.16 in the textbook (pages 112-115). See also explain video. □



Remark

In data assimilation, practical computations will rely on a chosen, fixed step-size $\delta t > 0$ for the dynamical model, which may be different from observation intervals $\Delta t_{\rm out} = N_{\rm out}\delta t$.

In this context it is helpful to generalise the Chapman-Kolmogorov equation to its $N_{\rm out}$-fold recursive application which propagates a PDF π_{Z^n} to $\pi_{Z^{n+N_{\rm out}}}$ according to

$$\begin{aligned}\pi_{Z^{n+N_{\rm out}}}(z) &= \int_{\mathbb{R}^{N_z}} \cdots \int_{\mathbb{R}^{N_z}} \pi(z|z')\,\pi(z'|z'')\cdots \\ &\qquad \pi\big(z^{(N_{\rm out}-1)}\big|z^{(N_{\rm out})}\big)\,\pi_{Z^n}\big(z^{(N_{\rm out})}\big)\,{\rm d}z'\cdots{\rm d}z^{(N_{\rm out})} \\ &= \int_{\mathbb{R}^{N_z}} \pi_{N_{\rm out}}(z|\tilde z)\,\pi_{Z^n}(\tilde z)\,{\rm d}\tilde z, \end{aligned} \tag{11}$$

where the one-step transition probability $\pi(z|z')$ is given by (6) and the implied $N_{\rm out}$-step transition probability is denoted by $\pi_{N_{\rm out}}(z|z')$.

$\rho_n \to \rho^*$ as $n \to \infty$ $\qquad \rho_{n+1} = T\rho_n$, $\rho^* = T\rho^*$

Definition (Invariant measure)

Let π_Z^* be a steady state solution to (10) (or the time-discrete Chapman-Kolmogorov equation). Then π_Z^* is called an invariant measure for the underlying Markov process.

$\int \bar{\pi}(z'|z)\pi^*(z)\,dz = \pi^*(z')$

Definition (Stationary processes)

If the random variable Z^0 has probability distribution with an invariant measure as its marginal PDF, *i.e.*, $\pi_{Z^0} = \pi_Z^*$, then all subsequent random variables Z^n with $n > 0$ will have the same marginal PDF. This means that expectation values $\mathbb{E}[f(Z^n)]$ are independent of n, with value

$$\mathbb{E}[f(Z^n)] = \int_{\mathcal{Z}} f(z)\pi_Z^*(z)\mathrm{d}z.$$

Such processes are called stationary.

Remark (Equilibration)

More generally, let us assume that the Chapman-Kolmogorov equation or (10) *has a unique invariant measure* π_Z^* *and that any initial PDF* π_{Z^0} *converges to* π_Z^* *in a weak sense, i.e.,*

$$\mathbb{E}[g(Z^n)] \to \int_{\mathcal{Z}} g(z)\pi_Z^*(z)\mathrm{d}z$$

for all bounded and continuous functions g *as* $n \to \infty$.

This process of convergence is called equilibration, with Z^n *said to be "close to equilibrium" if* $\pi_{Z^n} \approx \pi_Z^*(z)$.

In other words, such a stochastic process becomes stationary for n *sufficiently large, which motivates the following definition of ergodicity.*

Definition (Ergodicity)

Let $\{Z^n\}$ be a discrete Markov process with unique invariant measure π_Z^*. The process is said to be ergodic if time averages along specific process realisations $\{Z^n(\omega)\}$ converge almost surely to expectations with respect to π_Z^*, *i.e.*,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(Z^n(\omega)) = \int_{\mathcal{Z}} f(z)\pi_Z^*(z)\mathrm{d}z,$$

for all functions f with

$$\left| \int_{\mathcal{Z}} f(z)\pi_Z^*(z)\mathrm{d}z \right| < \infty.$$

Remark

Note that $Z_n \sim \pi_Z^*$ for all $n \gg 0$ does not imply that the Z_n's are independent! This is in contrast to the standard independent and identically distributed assumption of classical statistics.

4.3 Probabilistic forecasting and ensemble prediction

In probabilistic forecasting, our goal is calculate marginal PDFs $\pi_Z(z,t)$ for a Markov process, given the initial PDF $\pi_Z(z,0)$. For high dimensional models, the PDF $\pi_Z(z,t)$ is very unwieldy, and so, following Chapter 3, we concentrate on computing expectations.

This motivates a Monte Carlo approximation of the marginal PDFs $\pi_Z(z,t)$, $t \geq 0$, evolving under an SDE model

$$\mathrm{d}z = f(z)\mathrm{d}t + \sqrt{2}Q^{1/2}\mathrm{d}W(t)$$

or its difference approximation

$$z_i^{n+1} = z_i^n + \Delta t f(z_i^n) + \sqrt{2\Delta t}Q^{1/2}\eta_i^n \qquad i = 1,\ldots,M \quad (12)$$

$z_i^0 \sim \pi_0 \quad i = 1, \ldots, M$

Definition (Ensemble Prediction)

A Monte Carlo approximation to the marginal PDFs π_{Z^n} can be obtained from solving iteration

$$z_i^{n+1} = z_i^n + \delta t\, f(z_i^n) + \sqrt{2\delta t} Q^{1/2}\, \eta_i^n, \quad i = 1, \ldots, M,$$

where the $\eta_i^n \sim \mathrm{N}(0, I)$ and the initial conditions $z_i^0 \sim \pi_{Z^0}$ are sampled independently.

Approximations of expectations at later times $t > 0$ are obtained from

$$\bar{g}_M(t_n) := \frac{1}{M} \sum_{i=1}^{M} g(z_i^n) \approx \mathbb{E}[g(Z(t_n))]\,,$$

for $t_n = n\,\delta t$.

Definition (random probability measures)

Recall that Monte Carlo methods lead to empirical measures of the form

$$\hat{\pi}_Z(z) = \frac{1}{M}\sum_{i=1}^{M} \delta(z - z_i) \qquad z_i \sim \pi_Z$$

Since the z_i' are realisations of i.i.d. random variables Z_i with PDF π_Z, the associated measure $\hat{\pi}_Z$ is called a random probability measure.

We already introduced the operator $\mathcal{S}^M$ to denote the transition from π_Z to $\hat{\pi}_Z$, i.e.

$$\hat{\pi}_Z = \mathcal{S}^M \pi_Z \,.$$

A single step of the ensemble prediction method can now be written in the abstract form

$$\hat{\pi}_{Z^1} = \mathcal{S}^M \mathcal{P}_{\delta t} \mathcal{S}^M \pi_{Z^0}$$

$d(\hat{\pi}_{2'}, \pi_{2'})$

$\pi_{2'} = \mathcal{P}_{\delta t}\, \pi_{20}$

Remark (Random probability measures)

We like to characterise the error between π_{Z^n} and its Monte Carlo approximation $\hat{\pi}_{Z^n}$ using a distance between random probability measures.

Expectations over the randomness in an empirical measure will be denoted by $\mathbb{E}_\Omega$.

We note, for example, that

$$\mathbb{E}_\Omega \mathbb{E}_{\hat{Z}}[f] = \mathbb{E}_\Omega \left[\frac{1}{M} \sum_{i=1}^{M} f(Z_i) \right] = \mathbb{E}_Z[f].$$

We also introduce the shorthand

$$\mathbb{E}_Z[f] = \pi_Z[f], \qquad \mathbb{E}_{\hat{Z}}[f] = \hat{\pi}_Z[f],$$

etc. to make the dependence of the expectation value on the measure more explicit.

Definition (root mean square distance between random probability measures)

Given two random probability measures π_Z and $\hat{\pi}_Z$, we define their root mean square distance by

$$d(\pi_Z, \hat{\pi}_Z) := \sup_{|f|_\infty \leq 1} \sqrt{\mathbb{E}_\Omega |\pi_Z[f] - \hat{\pi}_Z[f]|^2}.$$

Lemma

Let $\mathcal{P}$ denote the operator associated with any Markov transition kernel and π_Z and $\hat{\pi}_Z$ any two random probability measures, then

$$d(\mathcal{P}\pi_Z, \mathcal{P}\hat{\pi}_Z) \leq d(\pi_Z, \hat{\pi}_Z).$$

Let $\mathcal{S}^M$ denote the MC resampling operator that draws realisations from M i.i.d. random variables $Z_i \sim \pi_Z$, then

$$d(\pi_Z, \mathcal{S}^M \pi_Z) \leq \frac{1}{\sqrt{M}}$$

for the associated empirical measure $\hat{\pi}_Z = \mathcal{S}^M \pi_Z$.

Theorem (convergence of ensemble prediction)

Let

$$\pi_{Z^1} = [\mathcal{P}_{\delta t}]\, \pi_{Z^0}$$

denote the exact measure of the stochastic model (12) after a single time step and

$$\hat{\pi}_{Z^n} = \mathcal{S}^M \mathcal{P}_{\delta t} \mathcal{S}^M \pi_{Z^0}$$

its Monte Carlo approximation.

Then

$$d(\pi_{Z^1}, \hat{\pi}_{Z^1}) \leq \frac{2}{\sqrt{M}}.$$

Proof.

$$d(\pi_{Z^1}, \hat{\pi}_{Z^1}) \leq d(\mathcal{P}_{\delta t}\pi_{Z^0}, \mathcal{P}_{\delta t}\mathcal{S}^M\pi_{Z^0}) + d(\mathcal{P}_{\delta t}\mathcal{S}^M\pi_{Z^0}, \mathcal{S}^M\mathcal{P}_{\delta t}\mathcal{S}^M\pi_{Z^0})$$
$$\leq d(\pi_{Z^0}, \mathcal{S}^M\pi_{Z^0}) + \frac{1}{\sqrt{M}} \leq \frac{2}{\sqrt{M}}$$

Example (Lorenz-63)

We return to the Lorenz-63 model with and without stochastic perturbations. We perform an ensemble prediction and compute the empirical expectation value of the absolute value $|\mathrm{x}(t)|$ of the x solution component, *i.e.*,

$$\overline{|\mathrm{x}|}_M(t) = \frac{1}{M}\sum_{i=1}^{M} |\mathrm{x}_i(t)|,$$

at initial time $t = 0$ and final time $t = 5$. We also estimate the variance

$$\mathrm{var}_M(t) = \frac{1}{M}\sum_{i=1}^{M} (|\mathrm{x}_i(t)| - \overline{|\mathrm{x}|}_M(t))^2.$$

Under the simplifying assumption of a Gaussian distribution, the 95% confidence interval for the estimated expectation values is itself estimated by

$$\left[\overline{|\mathrm{x}|}_M(t) - 1.96\sqrt{\frac{\mathrm{var}_M(t)}{M}}, \overline{|\mathrm{x}|}_M(t) + 1.96\sqrt{\frac{\mathrm{var}_M(t)}{M}}\right].$$

Example (Continued)

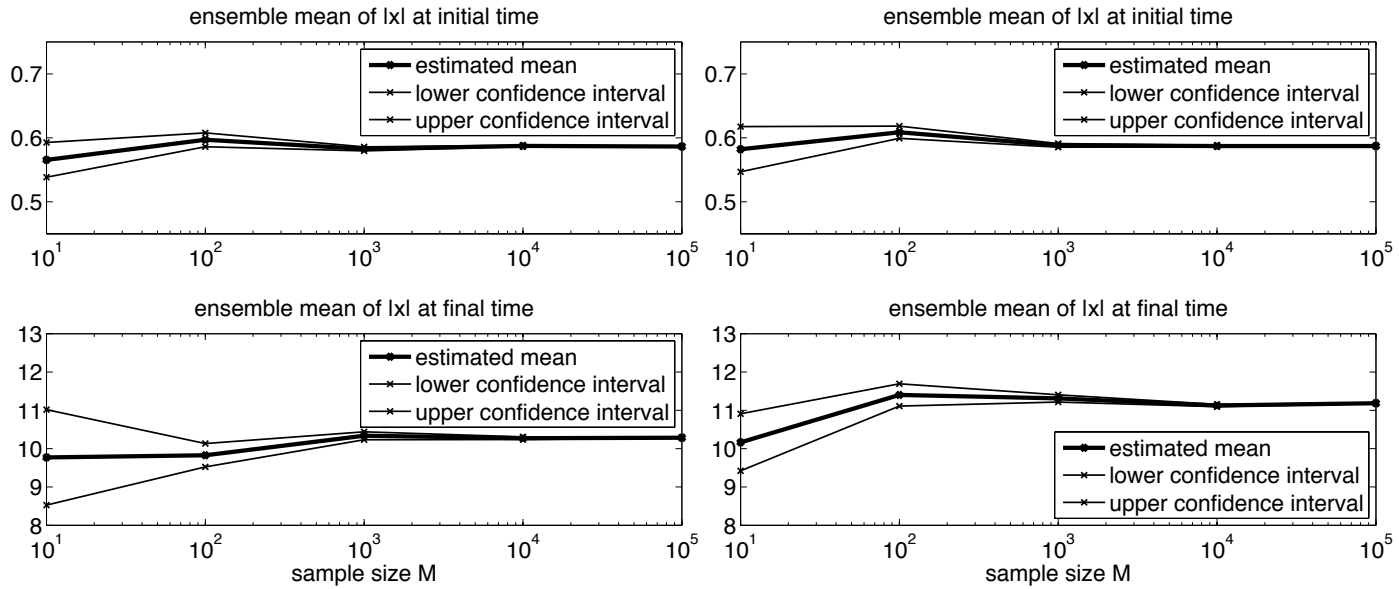


Figure: Ensemble estimates for $\overline{|\mathrm{x}|}$ and their confidence intervals for the deterministic Lorenz-63 model in the left panel and the stochastically perturbed variant for increasing ensemble size M in the right panel. Both models display a similar behaviour. However the asymptotic values for $\overline{|\mathrm{x}|}$ differ by about 10%.

4.4 Scoring rules for probabilistic forecasting

In probabilistic forecasting, we need to assess the impact of model and numerical errors on statistics, which are the main "product" that we are interested in.

A difficulty in analysing the impact of model errors on statistics is that the model is of some physical system where we cannot measure the complete system state or construct exact solutions.

One approach is to collect a time series of observations from the physical system, and to compare it with simulated observations from the model; this will determine how well calibrated the model is to the physical system.

We will assume that partial observations $y = h(z)$ of a physical process are available at equally spaced time instances $\mathrm{t}_k = k\Delta t_{\rm out}$, $k = 1, \ldots, N_{\rm obs}$, with $\Delta t_{\rm out} > 0$. We denote the observed values by $y_{\rm obs}(\mathrm{t}_k)$, and will ignore measurement errors throughout this section.

For simplicity, we will also assume that the observed quantity is a scalar. Then, given a model predicted PDF $\pi_Z(z, \mathrm{t}_k)$, we can deduce predicted probability measures $\mu_Y(\mathrm{d}y, \mathrm{t}_k)$ in the observable *via*

$$\mu_Y(A, \mathrm{t}_k) = \int_{h^{-1}(A)} \pi_Z(z, \mathrm{t}_k)\, \mathrm{d}z$$

for all Borel sets $A \subset \mathbb{R}$.

We will assume that there is a PDF $\pi_Y(y, \mathrm{t}_k)$ such that $\mu_Y(\mathrm{d}y, \mathrm{t}_k) = \pi_Y(y, \mathrm{t}_k)\mathrm{d}y$, *i.e.*, the measures $\mu_Y(\mathrm{d}y, \mathrm{t}_k)$ are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$.

Definition (Probability integral transform)

Given a scalar random variable Y with cumulative probability density function F_Y, the probability integral transform (PIT) defines a new random variable P with

$$P = F_Y(Y). \tag{13}$$

Lemma (Uniform distribution for PIT)

Let P be defined as the PIT of Y, as above, and assume that F_Y is strictly monotonic increasing on the range of Y. Then $P \sim U[0,1]$.

Proof.

See Lemma 4.24 in the textbook and its proof. □

Remark (Calibration of forecasts)

We can apply this to the calibration of dynamic models, if we assume that the observations $\{y_{\rm obs}(\mathrm{t}_k)\}$ *are ergodic with respect to some stationary invariant measure. We also assume that the forecast PDFs* $\pi_Y(y,\mathrm{t}_k)$ *become stationary as* $\mathrm{t}_k \to \infty$.

Under these assumption, we can combine the sequence of forecast PDFs $\pi_Y(y,\mathrm{t}_k)$ *together with observations* $y_{\rm obs}(\mathrm{t}_k)$ *for* $\mathrm{t}_k = k\Delta t_{\rm out}$, $k = 1,\ldots,N_{\rm obs}$ *with* $N_{\rm obs}$ *sufficiently large, to obtain PIT values*

$$p_{\mathrm{t}_k} = F_Y(y_{\rm obs}(\mathrm{t}_k),\mathrm{t}_k).$$

Here F_Y *denotes the induced cumulative distribution function given by*

$$F_Y(y,\mathrm{t}_k) = \int_{-\infty}^{y} \pi_Y(y',\mathrm{t}_k)\,\mathrm{d}y'.$$

Definition (Rank histogram)

Set $M+1$ counters ξ_i equal to zero initially. For $k = 1, \ldots, N_{\rm obs}$ do:

(i) Sort the (scalar) predicted observations $y_i = h(z_i(\mathrm{t}_k))$, $i = 1, \ldots, M$, such that $y_1 \leq y_2 \leq \cdots \leq y_M$.

(ii) Define the $M+1$ intervals $I_1 = (-\infty, y_1]$, $I_{i+1} = (y_i, y_{i+1}]$ for $i = 1, \ldots, M-1$, and $I_{M+1} = (y_M, +\infty)$.

(iii) Find the interval I_{i^*} containing the actual observation $y_{\rm obs}(\mathrm{t}_k)$, and increase the associated counter ξ_{i^*} by one.

Definition (Calibrated ensemble forecast)

A Monte Carlo ensemble forecast $\{y_i(\mathrm{t}_k)\}$, $k = 1, \ldots, N_{\rm obs}$, is called calibrated if the histogram of ranks is flat, *i.e.*,

$$\xi_i \approx \frac{N_{\rm obs}}{M+1}$$

for $N_{\rm obs}$ sufficiently large.

Definition (Forecast score)

Let F_Y be the cumulative distribution function produced by the forecast, and let $Y_{\rm obs}$ be a random variable which is being forecast by F_Y. $Y_{\rm obs}$ has PDF $\pi_{Y_{\rm obs}}$. A scoring rule is a random variable $S(F_Y, Y_{\rm obs})$ that assigns a scalar score to F_Y based on the actually observed value of $y_{\rm obs} = Y_{\rm obs}(\omega)$.

The expected value of the scoring rule is

$$\mathbb{E}\left[S(F_Y, Y_{\rm obs})\right] = \int_{\mathcal{Y}} S(F_Y, y_{\rm obs})\pi_{Y_{\rm obs}}(y_{\rm obs})\mathrm{d}y_{\rm obs}.$$

Definition (Proper scoring rules)

A scoring rule is:

(i) proper if

$$\mathbb{E}\left[S(F_Y, Y_{\rm obs})\right] \geq \mathbb{E}\left[S(F_{Y_{\rm obs}}, Y_{\rm obs})\right],$$

for all cumulative distribution functions F_Y, where $F_{Y_{\rm obs}}$ is the cumulative distribution function for $Y_{\rm obs}$, and

(ii) *strictly proper* if

$$\mathbb{E}\left[S(F_Y, Y_{\rm obs})\right] = \mathbb{E}\left[S(F_{Y_{\rm obs}}, Y_{\rm obs})\right] \implies F_Y = F_{Y_{\rm obs}}.$$

If $S(F_Y, y_{\rm obs})$ is strictly proper then $d(F_Y, F_{Y_{\rm obs}})$, defined by

$$d(F_Y, F_{Y_{\rm obs}}) = \mathbb{E}\left[S(F_Y, Y_{\rm obs})\right] - \mathbb{E}\left[S(F_{Y_{\rm obs}}, Y_{\rm obs})\right], \geq 0$$

satisfies the properties of a divergence, *i.e.*, $d(F_1, F_2) \geq 0$ for all F_1, F_2, and $d(F_1, F_2) = 0$ implies that $F_1 = F_2$.

Definition (Prediction skill)

Given a proper scoring rule $S(F_Y, Y_{\rm obs})$, we say that a prediction F_1 is more skillful than another prediction F_2 if

$$\mathbb{E}\left[S(F_1, Y_{\rm obs})\right] < \mathbb{E}\left[S(F_2, Y_{\rm obs})\right].$$

Remark (Empirical averaged scoring rule)

We do not have access to the probability distribution for the observations $\{y_{\rm obs}(t_k)\}$. Instead we must again assume that the statistics of the observations are ergodic and stationary and that the distributions $F_Y(\cdot, t_k)$ become stationary as $k \to \infty$. We can then calculate the empirical averaged scoring rule

$$\bar{S}(\{F_Y(\cdot, t_k)\}, \{y_{\rm obs}(t_k)\}) = \frac{1}{N_{\rm obs}} \sum_{k=1}^{N_{\rm obs}} S(F_Y(\cdot, t_k), y_{\rm obs}(t_k)).$$

In addition, we do not have access to $F_Y(\cdot, t_k)$ either, and must use the empirical CDF to approximate the scoring rule.

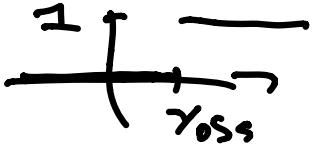
Definition (Continuous ranked probability score)

Given a cumulative distribution function $F_Y(y)$ and an observed value $y_{\rm obs}$ without observation errors, the continuous ranked probability score is defined by

$$S_{\rm crps}(F_Y, y_{\rm obs}) = \int_{-\infty}^{\infty} (F_Y(y) - F_{y_{\rm obs}}(y))^2 \, {\rm d}y, \tag{14}$$

where $F_{y_{\rm obs}}$ is the empirical cumulative probability distribution function

$$F_{y_{\rm obs}} = \begin{cases} 0 & \text{for } y < y_{\rm obs} \\ 1 & \text{for } y \geq y_{\rm obs} \end{cases} .$$



Lemma

The continuous ranked probability score is strictly proper.

Proof.

See Lemma 4.29 in the textbook and its proof. □

$M \to \infty$

Remark (Evaluation of CRPS)

Using the empirical cumulative distribution function instead of F_Y in the definition of the continuous ranked probability score, $S_{\mathrm{crps}}(F_{\hat{Y}_M}, y_{\mathrm{obs}})$ can be explicitly evaluated and we obtain

$$S_{\mathrm{crps}}(F_{\hat{Y}_M}, y_{\mathrm{obs}}) = \frac{2}{M}\sum_{i=1}^{M} \frac{i-1/2}{M}(y_{\mathrm{obs}} - y_i)_+ + \frac{2}{M}\sum_{i=1}^{M}\left(1 - \frac{i-1/2}{M}\right)(y_i - y_{\mathrm{obs}})_+.$$

Here $(y)_+ = 0$ if $y \leq 0$ and $(y)_+ = y$ for $y > 0$. For a single member forecast $y(\mathrm{t}_k) = h(z(\mathrm{t}_k))$, i.e., $M = 1$, this formula reduces to

$$S_{\mathrm{crps}}(F_{\hat{Y}_M}(\cdot, \mathrm{t}_k), y_{\mathrm{obs}}(\mathrm{t}_k)) = |y(\mathrm{t}_k) - y_{\mathrm{obs}}(\mathrm{t}_k)|.$$

Definition (Root mean square error)

The root mean square error scoring rule is

$$S(F_Y, y_{\rm obs}) = (\bar{y} - y_{\rm obs})^2,$$

where $\bar{y} = \mathbb{E}[Y]$. Application of this scoring rule in our framework gives rise to the root mean square error (RMSE)

$$\mathrm{RMSE}(F_Y) = \sqrt{\frac{1}{N_{\rm obs}} \sum_{k=1}^{N_{\rm obs}} (\bar{y}_M(\mathrm{t}_k) - y_{\rm obs}(\mathrm{t}_k))^2}, \quad \bar{y}_M(\mathrm{t}_k) = \frac{1}{M} \sum_{i=1}^{M} y_i(\mathrm{t}_k).$$

Lemma

The root mean square error scoring rule is proper but not strictly proper.

Proof.

It holds that

$$d(F_Y, F_{Y_{\rm obs}}) = (\overline{y} - \overline{y}_{\rm obs})^2 \geq 0$$

while there are many distributions with identical mean. □

Example (Lorenz-63)

We apply the rank histogram and the empirical averaged continuous ranked probability score (ACRPS),

$$\bar{S}_{\mathrm{crps}}(\{F_{\hat{Y}_M}(\cdot, \mathrm{t}_k)\}, \{y_{\mathrm{obs}}(\mathrm{t}_k)\}) = \frac{1}{N_{\mathrm{obs}}} \sum_{k=1}^{N_{\mathrm{obs}}} S_{\mathrm{crps}}(F_{\hat{Y}_M}(\cdot, \mathrm{t}_k), y_{\mathrm{obs}}(\mathrm{t}_k)),$$

to our deterministic and stochastically perturbed Lorenz-63 models. The resulting score for the deterministic model is 9.0049 while the stochastically perturbed model leads to 8.9786. We also run the reference model and obtain a score of 8.9572.

Furthermore, the rank histograms for the deterministic and the stochastically perturbed model reveal that both imperfect models can be considered as calibrated.

Example (Continued)

We now investigate the impact of numerical time-stepping errors.

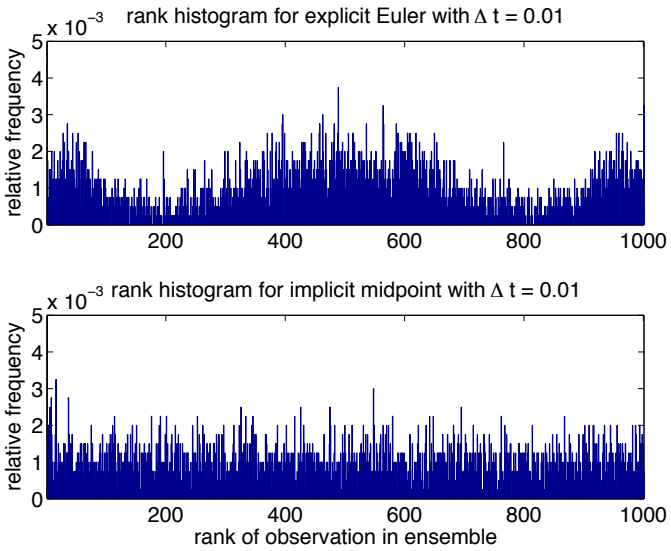


Figure: Effect of different time-stepping methods on calibration. While the Euler method with step-size $\Delta t = 0.01 = 10\delta t$ leads to non-calibrated forecasts, the implicit midpoint is still calibrated even though the step-size has increased by a factor of ten.

The computed ACRPS for the explicit Euler is 9.2786 while the implicit midpoint leads to 8.9823.