Convergence and Central Limit Theorem related to Sequential Monte Carlo

Gareth Peters

Cambridge University Statistical Signal Processing Group

Overview

- Brief review of mathematical concepts
  - Types of Convergence and how they relate.
  - Convergence Theorems
- Fundamentals of CLT
- Brief review of SMC methodology
- Almost Sure Convergence for SMC

- Description of SMC Framework used to analyse CLT

- Conclusions & Questions

Types of Convergence : Overview

- Almost Sure convergence : if N is an event with $\nu(N) = 0$ and a statement holds for all $\varpi \in N^c$ then the statement is said to hold a.s.

  - sequence $\{X_n\}$ converges to X almost surely $\left[X_n \stackrel{a.s.}{\to} X\right]$ iff $\lim_{n\to\infty} X_n = X$ a.s. (that is for all sets/events $N \in \Omega$ which are not of zero measure)

- Convergence in Probability : $\{X_n\}$ converges to X in probability $\left[X_n \stackrel{p}{\to} X\right]$ iff for every fixed $\varepsilon > 0$ we have

$$\lim_{n\to\infty} P\left(\|X_n - X\| > \varepsilon\right) = 0$$

- Convergence in L_r or r^{th} moment $\left[X_n \xrightarrow{L_r} X\right]$: iff for $r > 0$

$$\lim_{n \to \infty} E \left\| X_n - X \right\|_r^r = 0$$

- Weak Convergence in distribution $\left[F_n \xrightarrow{w} F\right]$ or measure $\left[P_n \xrightarrow{w} P\right]$: iff for each continuity point x of F we have

$$\lim_{n \to \infty} F_n(x) = F(x)$$

- Convergence in Law / Distribution $\left[X_n \xrightarrow{d} X\right]$: $\{X_n\}$ converges to X in distribution or law iff $F_{X_n} \xrightarrow{w} F_X$

- **Convergence in probability, L_r or almost sure convergence** can be considered as techniques which analsyse, for large n, how well X_n and X approximate each other as functions on the original probability space.

- **Convergence in distribution or measure** depends only on the distributions F_{X_n} and F_X or probability measures P_{X_n} and P_X, which does not require that X_n and X are close in any sense. (X_n and X do not even have to be defined on same probability space)

Relationship between different forms of Convergence

$X, X_1, X_2, X_3, \ldots$ are random k-vectors

- If $X_n \xrightarrow{a.s.} X$ then $X_n \xrightarrow{p} X$
- If $X_n \xrightarrow{Lr} X$ for some $r > 0$ then $X_n \xrightarrow{p} X$
- If $X_n \xrightarrow{p} X$ then $X_n \xrightarrow{d} X$

- Skorohod's Theorem : If $X_n \xrightarrow{d} X$ then there are random vectors $Y, Y_1, Y_2, \ldots$ defined on a common probability space such that $P_Y = P_X$, $P_{Y_n} = P_{X_n}$ $n = 1, 2, \ldots$ and $Y_n \xrightarrow{a.s.} Y$

- If, for every $\epsilon > 0$, $\sum_{n=1}^{\infty} P\left(\|X_n - X\| > \varepsilon\right) < \infty$, then $X_n \xrightarrow{a.s.} X$

- If $X_n \xrightarrow{p} X$ then there is a subsequence $\left\{X_{n_j}, j = 1, 2, \ldots\right\}$ such that $X_{n_j} \xrightarrow{a.s.} X$ as $j \to \infty$

- If $X_n \xrightarrow{d} X$ and $P\left(X = c\right) = 1$, where $c \in \mathbb{R}^k$ is a constant vector, then $X_n \xrightarrow{p} c$

- Suppose $X_n \xrightarrow{d} X$, then for any $r > 0$,

$$\lim_{n\to\infty} E\,\|X_n\|_r^r = E\,\|X\|_r^r < \infty$$

iff $\{\|X_n\|_r^r\}$ is uniformly integrable :

$$\lim_{t\to\infty} \sup_n E\left(\|X_n\|_r^r\,\mathbb{I}_{\{\|X_n\|_r > t\}}\right) = 0$$

Convergence Theorems : Overview

- Monotone Convergence Theorem
  - $\{X_n\}$ is sequence of Random Variables where $\{X_n\} \uparrow X$, then $\lim_{n\to\infty} E\left[X_n\right] = E\left[X\right]$

- Bounded Convergence Theorem
  - $\{X_n\}$ is sequence of RV's with $\lim_{n\to\infty} X_n = X$ and $|X_n| \le C \ \ \forall n$ with $C \in \mathbb{R}$ then $\lim_{n\to\infty} E[X_n] = E[X]$

- Fatou's Lemma
  - $E\left[\liminf_{n\to\infty} X_n\right] \le \liminf_{n\to\infty} E[X_n]$

- Dominated Convergence Theorem
  - $X, \{X_n\}$ are RV's and $\{X_n\} \to X$ w.p. 1 and if there is r.v. Y with $|X_n| \le Y \ \forall n$ and $E[Y] < \infty$ then $\lim_{n\to\infty} E[X_n] = E[X]$.

- Polya's Theorem : If $F_n \xrightarrow{w} F$ and F is continous on $\mathbb{R}^k$, then

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}^k} |F_n(x) - F(x)| = 0$$

Fundamentals of CLT

- The first CLT was proven for symmetric and Bernoulli independent distributions by De Moivre.

- Many extensions have been developed to deal with situations such as :

  - Multidimensional models

- Symmetric Statistics sequences
- Empirical Process
- General classes of nonidentically distributed and dependent random variables such as properly scaled random triangular arrays or Martingale sequences.

- Weak limit results can be determined from a given CLT using approximation techniques such as
  - Delta Method
  - Slutsky's technique

- Lindeberg's CLT (basic form) :
  - Sequence of r.v.'s $\{X_n\}$ which are $i.i.d.$ with mean μ and variance σ^2.
  - Consider normalised sum of random variables : $S_n = \frac{1}{n}(X_1 + X_2 + ... + X_n)$
  - Note : the pdf of S_n invloves the convolution of the pdf's of $X_1, X_2, ..., X_n$

    If one takes for example the pdf of unit box and convolves it many times then one sees the bell shape of the normal distribution and you obtain the Gaussian distribution after renormalising.
  - Basically the CLT states that as the sum grows $n \longrightarrow \infty$ then the cdf of a normalised version of S_n approaches the cdf of a Gaussian r.v. (convergence in distribution)

– Uses fact that Characteristic function specifies distributions uniquely :

Characteristic Function of r.v. X : $\varphi_X(w) = E\left[e^{iwX}\right] \quad -\infty < w < \infty$

Levy's convergence theorem : Shows that this implies that if a sequence of Characteristics functions converges to a limiting characteristic function, then the corresponding sequence of distributions converge to the limiting distribution.

Lindeberg Central Limit Theorem

$$\forall x \quad \lim_{n\to\infty} F_{Z_n}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$$

where

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{Var(S_n)}}$$

Proof :

Using two properties of Characteristic functions

- $\varphi_{X_1+X_2+...+X_n}(w) = E\left[e^{iw(X_1+X_2+...+X_n)}\right] = \left[\varphi_{X_1}(w)\right]^n$

- $\varphi_{aX}(w) = \varphi_{X_1}(aw)$

Therefore one gets :

$$\varphi_{Z_n}(w) = \varphi_{\sum\limits_{i=1}^{n}(X_i-\mu)}\left(\frac{w}{\sigma\sqrt{n}}\right) = \left[\varphi_{(X_1-\mu)}\left(\frac{w}{\sigma\sqrt{n}}\right)\right]^n$$

Now carry out a Taylor series expansion for $E\left[e^{iwX}\right]$

$$\begin{aligned} E\left[e^{iwX}\right] &= E\left[1 + (iwX) + \frac{(iwX)^2}{2!} + \ldots\right] \\ &= 1 + iwE[X] - \frac{w^2}{2!}E\left[X^2\right] + E\left[O\left(n^{-\frac{3}{2}}\right)\right] \end{aligned}$$

Hence one gets

$$\begin{aligned}
\varphi_{(X_1-\mu)}\left(\frac{w}{\sigma\sqrt{n}}\right) &= 1 + \frac{iw}{\sigma\sqrt{n}}E\left[X-\mu\right] - \frac{w^2}{2n\sigma^2}E\left(X-\mu\right)^2 + \frac{1}{n}E\left[R_n\right] \\
&= 1 - \frac{w^2}{2n} + \frac{1}{n}E\left[R_n\right]
\end{aligned}$$

Therefore

$$\varphi_{Z_n}\left(w\right) = \left[1 - \frac{w^2}{2n} + \frac{1}{n}E\left[R_n\right]\right]^n$$

and

$$\ln\left(\varphi_{Z_n}\left(w\right)\right) = n\ln\left[1 - \frac{w^2}{2n} + \frac{1}{n}E\left[R_n\right]\right]$$

note :

$$\ln(1-z) = -\left[z + \frac{z^2}{2} + \frac{z^3}{3} + ...\right] \qquad |z| < 1$$

now we make the following identity :

$$z = \frac{w^2}{2n} - \frac{1}{n}E\left[R_n\right]$$

hence

$$\begin{aligned} \ln\left(\varphi_{Z_n}(w)\right) &= n\left[-\frac{w^2}{2n} + \frac{1}{n}E\left[R_n\right] - \frac{n}{2}\left[\frac{w^2}{2n} - \frac{E\left[R_n\right]}{n}\right]^2 + ...\right] \\ &= -\frac{w^2}{2} + E\left[R_n\right] - \frac{n}{2}\left(\frac{w^2}{2n} - \frac{E\left[R_n\right]}{n}\right)^2 + ... \end{aligned}$$

now taking the limit as the sum of random variables grows :

$$\lim_{n\to\infty} \ln\left(\varphi_{Z_n}(w)\right) = \ln\left(\lim_{n\to\infty} \varphi_{Z_n}(w)\right) = -\frac{w^2}{2}$$

$$\Rightarrow \quad \lim_{n\to\infty} \varphi_{Z_n}(w) = e^{-\frac{w^2}{2}}$$

this is the Characteristic function of $N(0,1)$.

- This CLT holds when r.v.'s are $i.i.d.$, when this is not the case on must use more sophisticated CLT's such as Lyapunov's CLT

Lyaponov's CLT :

- – Sequence of independent r.v.'s $\{X_n\}$ must satisfy :

  – mean $\mu_n < \infty$ and variance $\sigma_n^2 < \infty$

  – $\beta_n = E\,|X_n - \mu_n|^3 < \infty$

- Define :

$$B_n = \left(\sum_{i=1}^{n} \beta_i\right)^{\frac{1}{3}}$$

$$C_n = \left(\sum_{i=1}^{n} \sigma_i^2\right)^{\frac{1}{2}}$$

$$Z_n = \frac{\sum_{i=1}^{n} (X_i - \mu_i)}{C_n}$$

If $\lim_{n \to \infty} \frac{B_n}{C_n} = 0$ then $\forall x$:

$$\lim_{n \to \infty} F_{Z_n}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$$

Martingale CLT

- Let $\{X_n\}$ be a sequence of integrable random variables on probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{F}_1 \subset \mathcal{F}_2 ... \subset \mathcal{F}$ is a sequence of $\sigma - fields$ such that $\sigma(X_n) \subset \mathcal{F}_n$, $n = 1, 2, ...$

- Sequence $\{X_n, \mathcal{F}_n : n = 1, 2, ...\}$ is a Martingale iff :

$$E\left[X_{n+1}|\mathcal{F}_n\right] = X_n$$

Sigma field $\mathcal{F}_n$ has the interpretation of information available at time n and X_n denotes a random quantity whose value $X_n(w)$ is revealed at time n.

- If $X_n = Z_0 + ... + Z_n$, where $\{Z_i\}$ are $i.i.d.$ with mean 0 and variance 1

- Classical CLT states : $\frac{X_n}{\sqrt{n}} \Rightarrow N(0, 1)$

General Martingales :

- A similar result also holds for more general martingales $\{X_n\}$.
- Let $\mathcal{F}_n = \sigma(X_0, X_1, ..., X_n)$ $\quad \sigma_0^2 = Var(X_0)$ and for $n \geq 1$ set $\sigma_n^2 = Var(X_n | \mathcal{F}_{n-1}) \equiv E\left[(X_n - X_{n-1})^2 | \mathcal{F}_{n-1}\right] = E\left[X_n^2 - X_{n-1}^2 | \mathcal{F}_{n-1}\right]$
- Now by induction : $E\left(X_n^2\right) = \sum_{k=0}^{n} E\left[\sigma_k^2\right]$
- $E(X_n) = E(X_0)$ does not grow with n

- If set the following stopping time random variable $v_t = \text{min}\left\{n \geq 0; \sum_{k=0}^{n} \sigma_k^2 \geq t\right\}$ then for large t one has the following Martingale difference CLT

$$t \longrightarrow \infty \text{ implies that } \frac{X_{v_t}}{\sqrt{t}} \Rightarrow N\left(0,1\right)$$

- Del Moral and others tend to use a Martingale difference array to establish a CLT

- let $\{X_n : n \in \mathbb{N}\}$ be a Martingale written as a sum of increments

$$X_n := X_0 + \xi_1 + \xi_2 + ... + \xi_n$$

- $\{\xi_n\}$ are Martingale differences if they satisfy integrability condition and $E\left[\xi_n | \mathcal{F}_{n-1}\right] \overset{a.s.}{=} 0$

– For each n in $\mathbb{N}$ let $\left\{\xi_{nj} : j = 0, ..., k_n\right\}$ be a Martingale difference array, with respect to filtration $\left\{\mathcal{F}_{nj}\right\}$ for which the following hold

1. $\sum_j \xi_{nj}^2 \to 1$ in probability

2. $\max_j \left|\xi_{nj}\right| \to 0$ in probability

3. $\sup_n \mathbb{P} \max_j \xi_{nj}^2 < \infty$

then one has the following Martingale triangular array CLT

$$\sum_j \xi_{nj} \xrightarrow{d} N(0,1) \text{ as } n \to \infty$$

- So the above provided tools to prove the existence of a CLT for a given sequence of r.v.'s. However what is more useful to practitioners is a method of obtaining an expression for the variance of the assymptotic Gaussian as a function of the number of terms in the sum n.

Delta Method

- Used as a method of obtaining a recursive relationship for the assymptotic variance in a CLT
  - Kunsch, Chopin and Del Moral use this technique in the papers mentioned.

1-D Case

Have some function denoted :

$$\phi : \mathbb{R}^1 \rightarrow \mathbb{R}^1$$

want distribution of

$$\sqrt{N}\left(\phi\left(S_N\right) - \phi\left(\mu_{\psi(x)}\right)\right)$$

Delta method states that if :

$$\sqrt{N}\left(S_N - \mu_{\psi(x)}\right) \overset{N \rightarrow \infty}{\rightarrow} Z \sim N\left(0, \sigma^2_{\psi(x)}\right)$$

where $S_N = \frac{1}{N}\sum_{i=1}^{N} \psi\left(X_i\right)$ and $E\left[\psi\left(X_i\right)\right] = \mu_{\psi(X_i)}$

then

$$\begin{aligned}\sqrt{N}\left(\phi\left(S_N\right)-\phi\left(\mu_{\psi(x)}\right)\right) &\approx \phi'\left(\mu_{\psi(x)}\right)\sqrt{N}\left(S_N-\mu_{\psi(x)}\right) \\ &\rightarrow Z_* \sim N\left(0, \phi'\left(\mu_{\psi(x)}\right)^2 \sigma^2_{\psi(x)}\right)\end{aligned}$$

Multi-dimensional Case

Have a function denoted

$$\begin{aligned}\phi &: \mathbb{R}^k \rightarrow \mathbb{R}^m \\ \phi\left(X_1, X_2, ..., X_k\right) &= \left(Y_1, Y_2, ..., Y_m\right)\end{aligned}$$

have

$$\begin{aligned}S_N &\in \mathbb{R}^k \quad S_N = \left(S_{N,1}, S_{N,2}, ..., S_{N,k}\right) \\ S_{N,i} &= \frac{1}{N}\sum_{j=1}^{N}\psi_i\left(X_j\right)\end{aligned}$$

and

$$\mu_{\psi(x)} = \left(\mu_{\psi_1(x)}, \mu_{\psi_2(x)}, ..., \mu_{\psi_k(x)}\right)$$

then as before get

$$\begin{aligned}
\sqrt{N}\left(\phi\left(S_N\right) - \phi\left(\mu_{\psi(x)}\right)\right) &\approx \phi'\left(\mu_{\psi(x)}\right)\sqrt{N}\left(S_N - \mu_{\psi(x)}\right) \\
&\rightarrow Z_* \sim N_m\left(0, \phi'\left(\mu_{\psi(x)}\right)\sum \phi'\left(\mu_{\psi(x)}\right)^T\right)
\end{aligned}$$

where

$$\begin{aligned}
\sum &= \begin{bmatrix} c_{11} & \cdots & c_{1k} \\ \cdot & & \cdot \\ c_{k1} & \cdots & c_{kk} \end{bmatrix} \\
c_{ij} &= E\left[\left(S_{N,i} - \mu_{\psi_i(x)}\right)\left(S_{N,j} - \mu_{\psi_j(x)}\right)\right] = \begin{cases} \sigma^2_{\psi_i(x)} & i = j \\ c_{\psi_i(x)\psi_j(x)} & i \neq j \end{cases}
\end{aligned}$$

and

$$\phi'\left(\mu_{\psi(x)}\right) = \begin{bmatrix} \frac{\delta\phi_1}{\delta X_1}\left(\mu_{\psi(x)}\right) & \cdots & \frac{\delta\phi_1}{\delta X_k}\left(\mu_{\psi(x)}\right) \\ . & & . \\ \frac{\delta\phi_m}{\delta X_1}\left(\mu_{\psi(x)}\right) & \cdots & \frac{\delta\phi_m}{\delta X_k}\left(\mu_{\psi(x)}\right) \end{bmatrix}$$

Brief Review of Aspects of SMC

- Sequential Monte Carlo SMC methods are a general class of iterative algorithms provideing empirical Monte Carlo estimates of a sequence of distributions $\{\pi_n\}_{n\in\mathcal{N}}$.

- Brief mention of algorithmic aspects of SMC developed :
  - Sequential Importance Sampling (SIS)
  - Resampling (SIR) (*Gordon, Kitigawa*)
  - Resample and Move (*Gilks and Berzuinni*)
  - multinomial, residual and stratefied resampling
  - auxiliary particle filtering (*Pitt and Shephard, Davy and Doucet*)
  - Rao Blackwellised particle filtering (*Doucet and Liu*)
  - Smoothing (*Godsill, Doucet and West*)

- PHD filters (*Vo, Singh and Doucet*)
- SMC samplers (*DelMoral, Doucet and Peters*)
- More general formulations often known as Interacting Particle Systems (*Del-Moral*) which provide a mean field particle approximation of a general class of Feynman-Kac path measures.

- Most people are familiar with basic aspects of Particle Filters or Sequential Monte Carlo :
  - Posterior distribution of a state is approximated by a large set of Dirac-delta masses ($samples/particles$) that evolve randomly in time according to the dynamics of the models and the observations.

- An interacting particle method is a sequential simulation method, where particles explore the state space by mimicking the evolution of an underlying random process.

- They learn the environment by evaluating a fitness function, and interact so only the most successful particles, survive and get offspring.

- This is termed as Mutation and Selection stages which have the effect of concentrateing particles in regions of interest in the state space.When thinking of particle filtering each particle represents a possible hidden state, and is multiplied or discarded at the next generation on the basis of its consistency with the current observation, as measured by the likelihood function.

Convergence Results Obtained for SMC Methodology

1. *A survey of Convergence Results on Particle Filtering Methods for Practitioners*, Dan Crisan and A. Doucet

2. *Recursive Monte Carlo Filters : Algorithms and Theoretical Analysis*, Hans R. Kunsch

3. *Feynman-Kac Formulae : Genealogical and Interacting Particle Systems with Applications*,Pierre Del Moral.

4. *Central Limit Theorem for Sequential Monte Carlo Methods and its Applications to Bayesian Inference,* N. Chopin

- Key point to make : Since the particles are interacting one can not straightforwardly apply standard classical limit theorems which rely on statistically independent samples.

- When considering convergence analysis, as pointed out in 1, we should ask questions such as :

  – Does the particle filter converge asymptotically, in the number of particles N or in time n or in both, to the optimal filter and in what sense does this convergence occur ?

- Do the standard rates established for Monte Carlo techniques apply ?

- Does the error accumulate with time and what effect does this have on convergence ?

- Can asymptotic results such as CLT, Berry Essen, Bias, Propogation of Chaos and Large deviations be established and in what forms of SMC algorithm are they valid ?

- The mathematical justification of SMC algorithms generally focuses on the simple result of the strong law of large numbers, which states for a measurable test function $\varphi(x)$ it can be shown that

$$\int \varphi_n(x_{1:n})\,\widehat{\pi}_n^N(dx_{1:n}) \underset{N\to\infty}{\overset{a.s}{\longrightarrow}} \int \varphi_n(x_{1:n})\,\pi_n(dx_{1:n})$$

where $\widehat{\pi}_n^N (dx_{1:n})$ is the empirical estimate of the target measure $\pi_n (dx_{1:n})$.

- This SLLN result is important to practitioners but not of very much practical use as it does not provide a rate of convergence of the estimate or a variance for the estimate.

- Majority of convergence results in engineering literature focuss on the filtering formulation in which the target distribution $\pi_n (dx_{1:n})$ is interpreted as the conditional distribution $P (dx_{1:n}|y_{1:n})$ of an unobserved state sequence $X_{1:n}$ which evolves according to a nonlinear state space model.

- The conditioning is made on a noisy observation process $Y_{1:n} = y_{1:n}$ which is assumed fixed and is generally a known nonlinear function of the state variable and a noise sequence.

Review of Almost Sure Convergence Results

(Reference 1)

- Preliminary Definitions : Notation consistent with paper usefull for later reading....
  - State space - E endowed with a metric d to make metric space (E, d)
  - $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ are sequences of continous functions $a_n, b_n : E \rightarrow E$
  - k_n and $k_{1:n}$ are functions defined as

$$\begin{aligned} k_n &\triangleq a_n \circ b_n \\ k_{1:n} &\triangleq k_n \circ k_{n-1} \circ ... \circ k_1 \end{aligned}$$

- In a filtering context we now associate these abstract mappings to the SMC framework :
  - E will be defined as $\mathcal{P}\left(\mathbb{R}^{n_x}\right)$ which is the space of all probability measures over n_x-dimensional Euclidean space
  - b_n is considered as the map taking $\pi_{n-1|n-1} \longrightarrow \pi_{n|n-1}$ (ie. prediction / mutation mapping)
  - a_n is the correction map which takes $\pi_{n|n-1} \longrightarrow \pi_{n|n}$ via Bayes rule
  - k_n is now the map from $\pi_{n-1|n-1} \longrightarrow \pi_{n|n}$
  - $k_{1:n}$is the map from $\pi_{0|0} \longrightarrow \pi_{n|n}$

– c_n^N is the (not necessarily continous) discreteization/selection mapping from $E \longrightarrow E$.

- Now the discreteization / selection stage must be considered. This is introduced as a perturbation c_n^N to the above mappings, as follows

$$\begin{aligned} \widehat{k}_n^N &\triangleq c_n^N \circ a_n \circ c_n^N \circ b_n \\ \widehat{k}_{1:n}^N &\triangleq \widehat{k}_n^N \circ \widehat{k}_{n-1}^N \circ \ldots \circ \widehat{k}_1^N \end{aligned}$$

  Effectively c_n^N takes a measure to a random empirical measure, of N samples, which approximates the original measure.

- In paper 1 some constraints are placed on the mapping c_n^N. At a given time n it is required that as $N \longrightarrow \infty$ then c_n^N must converge in a UNIFORM manner to the identity mapping ie.

$$\forall n \text{ and } \forall e_N, e \in E \text{ we have } \lim_{N\to\infty} e_N = e \Rightarrow \lim_{N\to\infty} c_n^N(e_N) = e$$

If these assumptions are satisfied then one may establish the following almost sure convergence result :

$$\lim_{N\to\infty} k_n^N = k_n \quad \text{and} \quad \lim_{N\to\infty} k_{1:n}^N = k_{1:n} \tag{1}$$

and k_n^N and $k_{1:n}^N$ satisfies the following :

$$\begin{aligned} \lim_{N\to\infty} e_N &= e \Rightarrow \lim_{N\to\infty} k_n^N(e_N) = k_n(e) \\ \lim_{N\to\infty} e_N &= e \Rightarrow \lim_{N\to\infty} k_{1:n}^N(e_N) = k_{1:n}(e) \end{aligned} \tag{2}$$

- The proof of these statements follows by proving 2 since this implies 1. An induction argument on n is used and also the fact that a_n, b_n are continous mappings and we have assumed that c_n^N converges in a Uniform manner to the identity, assymptotically in N.

- Before studying the Central Limit Theorem it is useful to present how most of the analysis is developed :

- This will start with more overview of the models and then present an important mathematical break down of the models (briefly mentioned above) which allows the CLT to proceed.

Description of Models

- Have a sequence of target distributions $(\pi_n(dx_{0:n}))_{n\in\mathcal{N}}$ defined on $E_n = E^n$, each admiting a density $\pi_n(x_{0:n})$ with respect to dominating measure $dx_{0:n}$.

- Assumed that these densities may be evaluated pointwise upto a normalising constant.

- At time $n-1$ we have the weighted particles $\left\{W_{n-1}^{(i)}, X_{1:n-1}^{(i)}\right\}$ which approximate $\pi_{n-1}(dx_{0:n-1})$ using the empirical measure

$$\widehat{\pi}_{n-1}^{N}(dx_{0:n-1}) = \sum_{i=1}^{N} W_{n-1}^{(i)} \delta_{X_{0:n-1}^{(i)}}(dx_{1:n-1})$$

- Interacting particle method, used to approximate these target distributions, is presented in three stages.
  - **Mutation** stage involves mutation of the particles $X_{0:n-1}^{(i)}$ by a Markov kernel $M_n\left(X_{0:n-1}^{(i)}, dx_{0:n}\right)$ to obtain new paths $X_{0:n}^{(i)}$.
  - In the Bayesian filtering formulation this kernel may depend on $y_{1:n}$.
  - The new particles are distributed according to

$$\widetilde{\pi}_n(dx_{0:n}) = \int \pi_{n-1}(dx_{0:n-1}) M_n(x_{0:n-1}, dx_{0:n})$$

  - **Correction** stage involves importance sampling to produce a consistent empirical estimate which corrects for the difference between the sampling distribution $\widetilde{\pi}_n(dx_{0:n})$ and target distribution $\pi_n(dx_{0:n})$.

– The unnormalised incremental importance weight is given by

$$w_n^{(i)} = w\left(X_{0:n}^{(i)}\right) \propto \frac{\pi_n\left(X_{0:n}^{(i)}\right)}{\widetilde{\pi}_n\left(X_{0:n}^{(i)}\right)} = \frac{\pi_n\left(X_{0:n}^{(i)}\right)}{\pi_{n-1}\left(X_{0:n-1}^{(i)}\right) M_n\left(X_{n-1}^{(i)}, X_n^{(i)}\right)} \tag{3}$$

which produces the weighted empirical measure.

$$\widehat{\pi}_n^N\left(dx_{0:n}\right) = \sum_{i=1}^{N} W_n^{(i)} \delta_{X_{0:n}^{(i)}}\left(dx_{0:n}\right) \tag{4}$$

Where the normalised importance weights are given by

$$W_n^{(i)} \propto W_{n-1}^{(i)} w_n\left(X_{0:n}^{(i)}\right), \sum_{i=1}^{N} W_n^{(i)} = 1$$

– **Resampling** stage is where the weighted particles are resampled to obtain $\left\{\frac{1}{N}, X_{0:n}^{(i)}\right\}$.

- If you read Del Moral's work he considers SMC algorithms in a more general framework which involves two stages of mutation and selection.

- Selection stage is represented by the updating Markov transition on E_n which maps the space of probability distributions $\mathcal{P}(E_n)$ into itself :

$$S_{n,\widetilde{\pi}_n^N}\left(X_n^{(i)}, dz_n\right) = \epsilon_n w_n^{(i)} \delta_{X_n^{(i)}}(dz_n) + \left(1 - \epsilon_n w_n^{(i)}\right) \sum_{i=1}^{N} \frac{w_n^{(i)}}{\sum_{j=1}^{N} w_n^{(j)}} \delta_{X_n^{(i)}}(dz_n)$$

where ϵ_n represents any possibly null constant such that $\|\epsilon_n w_n^{(i)}\| \leq 1$.

- This version of update and resampling stages is interpreted as having $X_n^{(i)} = X_{n-1}^{(i)}$ with weight $\epsilon_n w_n^{(i)}$, which means particle i remains in its current position after mutation at time $n-1$ or particle i may be resampled multinomially with weight $\left(1 - \epsilon_n w_n^{(i)}\right)$.

- Advantage of this is explained by fact that if a particle is in a position in the state space which lends considerable support to the posterior of interest $\pi_n\left(dx_n\right)$, as shown by the incremental weight $w_n^{(i)}$ then it should not have to take part in the resampling stage.

- Hence one obtains a reduction in the variance as demonstrated in the CLT proofs established by DelMoral, Doucet and Peters.

- Another point to mention is that the resampling stage, although required to provide the interactions within the particle system, should be carried out as infrequently as possible in order to help reduce the variance of the particle weights.

- There are also many resampling schemes which have been developed, as mentioned earlier, some have the advantage of being minimum variance whilst others are easy to implement.

- The minimum variance resampling is stratefied or systematic resampling which ensures the variance of any empirical estimates obtained after resampling is minimised.

- Proofs of CLT variance have only been carried out successfully to this date for multinomial resampling although Chopin in 4 has made attempts at such proofs for residual resampling.

End of Part 1