

PROPAGATION OF CHAOS IN ENTROPY

ROBERT AEBI

Abstract

A notion of convergence ‘propagation of chaos’, called by McKean [18], is defined in terms of the relative entropy. Our goal is to show ‘the propagation of chaos *in entropy*’ for clouds of interacting particles with prescribed initial and terminal distributions. It is shown that, as the number of particles in the clouds tends to infinity, particles in the clouds become, in the sense of relative entropy, asymptotically independent with an identical distribution Q . The limiting distribution Q is known to be a diffusion process related to the Schrödinger equation.

1. Introduction

The propagation of chaos for systems of interacting diffusion processes is introduced by McKean [15], [16] and investigated in Gutkin and Kac [10], Tanaka [30], Kusuoka and Tamura [15], Shiga and Tanaka [28], Oelschläger [26], Sznitman [29] and Dawson and Gärtner [8]. Our discussion is related to problems considered by Nagasawa and Tanaka [17], [18], [19] and Aebi and Nagasawa [1].

In this paper we formulate ‘propagation of chaos’ in terms of relative entropy which has been established as a natural indicator for the mutual randomness of probability measures. See for example Csiszar [5], [6] in connection with large deviations, Boltzmann [4] and Lanford [16] who deal with statistical mechanics and Khinchin [12] for an approach to information theory.

DEFINITION 1. A sequence of systems $((X_1, \dots, X_n), \mathbf{Q}^{(n,k)})$ of interacting diffusion processes performs *propagation of chaos in entropy* with limiting distribution Q as n and k tend to infinity, if for each $m \in \mathbf{N}$, the marginal distribution $Q_m^{(n,k)}$ of $\mathbf{Q}^{(n,k)}$ on Ω^m and the empirical distribution $L_n = 1/n \sum_{i=1}^n \delta_{\omega_i}$, $\omega_i \in \Omega$, $i \in \mathbf{N}$, satisfy

Key words: propagation of chaos, relative entropy, interacting diffusion processes.
AMS subject classification: primary 82C22; secondary 60J60, 81B05.
Received December 21, 1994; revised April 5, 1996.

$$(1) \quad \lim_{k \nearrow \infty} \lim_{n \nearrow \infty} H(Q^m | Q_m^{(n, k)}) = 0$$

$$(2) \quad \lim_{k \nearrow \infty} \lim_{n \nearrow \infty} H(Q | Q^{(n, k)} [L_n]) = 0$$

where k is called a modeling parameter.

The definition given above is the key point of the present article as will be explained: Nagasawa's definition given in section 8.2 of [24] consists of two steps; namely, $H(Q_m^{(n, k)} | Q_k^m)$ as well as $H(Q^{(n, k)} [L_n] | Q_k)$ are required to vanish as n tends to infinity, where $(Q_k)_{k \in N}$ is a family of distributions approximating Q , indexed by the modeling parameter k , and then $H(Q | Q_k)$ is required to vanish as k tends to infinity. As shown in [1], [24], the definition has technically no problem, since it implies 'the propagation of chaos in variation-metric' (cf. inequality (16) below). It is, nonetheless, mathematically not satisfactory. Our Definition 1 is a modification which is more natural than Nagasawa's definition, since it avoids the two-steps-argument. This causes, however, mathematical problems, because of the non-symmetry of the relative entropy $H(Q | P)$ in $\{Q, P\}$ (cf. definition (5)), which will be resolved in Section 3. Our goal is to prove Theorem 1 given in Section 2, which improves Theorem 8.3 in [24] and Theorem 4.1 in [1].

The meaning of Definition 1 is this: First of all, property (1) implies that for arbitrary but fixed m any subset of m -interacting diffusion processes is asymptotically independent and that Q is the asymptotic distribution of each diffusion process in the subset. Property (2) indicates a mixing of the involved diffusion processes by means of the empirical distribution L_n (cf. definition (6)), namely, the empirical distribution L_n under $Q^{(n, k)}$ converges to Q in entropy as the total number n of participating diffusion processes and the modeling parameter k tend to infinity.

In Section 2 a theorem on 'the propagation of chaos in entropy' in the sense of Definition 1 will be given. This kind of limit theorem was first investigated by Schrödinger in [27], who was motivated by an analogy of wave functions in quantum mechanics and a pair of diffusion equations in duality: He considers a cloud of particles independent and identically distributed according to an initial distribution q_a at an initial time a , and then he requires that the terminal distribution at terminal time b , $a < b$, must be a prescribed distribution q_b . To realize such movements of particles, it is therefore necessary to introduce interactions between participating particles at intermediate time t , $a < t < b$, through conditioning adapted to the prescribed distributions q_a and q_b , which are assumed to be observable. The 'controlled' cloud of particles is obtained as the 'most probable' one in the sense of large deviations (cf. [1]). To discuss large deviations we adopt, as the substantial mathematical tools, the asymptotic quasi-independence as well as the I -('Csiszar') projection given by Csiszar [5], [6].

Section 3 is devoted to a proof of 'the propagation of chaos in entropy' in the setting of Section 2. In Section 4 frequently used known results are briefly quoted for the reader's convenience. It should be noticed that the limit (41)

stated in Proposition 2 holds only in this revised form which is essentially weaker than (2.13) in [1].

Corollary 2 of Theorem 1 is to indicate that the Schrödinger equation can be considered as a kind of 'Boltzmann equation' (cf. [23]), i.e. an equation for systems of interacting particles in the spirit of Boltzmann. To investigate this aspect was the main motivation of Schrödinger in [27].

2. A situation for propagation of chaos

To formulate our main theorem we recall some notations and definitions (cf. [1], [24] for details). Let $\Omega = C([a, b], \mathbf{R}^d)$, $-\infty < a < b < \infty$, be the space of continuous paths taking values in \mathbf{R}^d , d fixed, with the Borel σ -field $\sigma(\Omega)$. $M_1(\Omega)$ denotes the set of probability measures on Ω . The state of a path $\omega \in \Omega$ at time t , $a \leq t \leq b$, is described by the projection $X_t(\omega) = X(t, \omega) = \omega(t)$.

For a given pair of probability measures (q_a, q_b) on \mathbf{R}^d , we consider

$$(3) \quad A_{a,b} = \{P \in M_1(\Omega) : P \circ X_r^{-1} = q_r \text{ for } r = a, b\} \subset M_1(\Omega)$$

which is a class of continuous stochastic processes on \mathbf{R}^d with prescribed marginal distributions q_a and q_b at finite initial and final times a and b , respectively.

Let $\bar{P} \in M_1(\Omega)$ be a Markov process which is going to serve as a reference measure on the path space Ω . We assume that

$$(4) \quad \exists P \in A_{a,b} : H(P | \bar{P}) < \infty$$

where the relative entropy $H(P | \bar{P})$ of P with respect to \bar{P} is defined as

$$(5) \quad H(P | \bar{P}) = \int \log\left(\frac{dP}{d\bar{P}}\right) dP, \text{ if } P \ll \bar{P}, \\ = \infty, \text{ otherwise.}$$

In most of the interesting known situations and hence also in our investigations, the measure \bar{P} itself is not an element of the set $A_{a,b}$. In applications of Theorem 1 which will be given below, a typical reference measure is the so-called 'renormalization' of a measure with creation and killing (cf. [1], [24]).

Let (Ω^n, \bar{P}) be n independent copies of (Ω, \bar{P}) , i.e. \bar{P} is the n -product of the probability measure \bar{P} . Denoting the empirical distribution of $\omega = (\omega_1, \dots, \omega_n) \in \Omega^n$ by

$$(6) \quad L_n(\omega) = \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i}$$

we have $L_n(\omega) \in M_1(\Omega)$ for each such ω . To define a reasonable conditioning, we must enlarge the subset $A_{a,b}$, since L_n takes only discrete values and hence $\{L_n \in A_{a,b}\}$ might be empty in general. Let us take any sequence of finite measurable partitions $\mathcal{P}_k(\mathbf{R}^d) = \{B_1, \dots, B_k\}$ of \mathbf{R}^d , $k \in \mathbf{N}$, such that

$$(7) \quad \sigma(\mathcal{P}_k(\mathbf{R}^d)) \subset \sigma(\mathcal{P}_{k+1}(\mathbf{R}^d)) \quad \text{and} \quad \sigma(\mathcal{P}_k(\mathbf{R}^d)) \nearrow \sigma(\mathbf{R}^d) \quad \text{as } k \nearrow \infty.$$

We define a family of subsets $A(\varepsilon, k)$ of $M_1(\Omega)$ for $\varepsilon > 0$ and $k \in \mathbf{N}$ in terms of the partitions $\mathcal{P}_k(\mathbf{R}^d)$ as

$$(8) \quad A(\varepsilon, k) = \{P \in M_1(\Omega) : |P(X_r \in B_i) - q_r(B_i)| \leq \frac{\varepsilon}{2^k}, \quad \forall B_i \in \mathcal{P}_k(\mathbf{R}^d),$$

$$\text{and } P \circ X_r^{-1} \ll \bar{P} \circ X_r^{-1} \text{ on } \sigma(\mathcal{P}_k(\mathbf{R}^d)) \text{ for } r = a, b\}.$$

Thus

$$(9) \quad \bar{P}(L_n \in A(\varepsilon, k)) > 0$$

for $\varepsilon > 0$, $k \in \mathbf{N}$ and $n \in \mathbf{N}$ large enough. In fact, since $L_n(\omega) \circ X_r^{-1}$ for $\omega \in \Omega^n$ is concentrated on $\omega_j(r) \in \mathbf{R}^d$, $j = 1, \dots, n$, $\{L_n \in A(\varepsilon, k)\}$ is the set of those $\omega \in \Omega^n$ for which

$$\left| \frac{\#_r(B_i)}{n} - q_r(B_i) \right| \leq \frac{\varepsilon}{2^k} \quad \text{for } i = 1, \dots, k; \quad r = a, b$$

where $\#_r(B_i)$ is the number of $\omega_j(r)$, $j = 1, \dots, n$, contained in the set $B_i \in \mathcal{P}_k(\mathbf{R}^d)$. Because of (4), there exists $P \in A_{a,b}$ in (3) such that $P \ll \bar{P}$. Hence (9) is a consequence of the law of large numbers which claims that $P^n(L_n \in A(\varepsilon, k))$ tends to 1 as n increases to infinity.

The conditioning of \bar{P} on the set $A(\varepsilon, k)$ in (8) by means of the empirical distribution L_n in (6) can now be defined as

$$(10) \quad \bar{P}^{(n,k)}(\cdot) = \bar{P}(\cdot | L_n \in A(\varepsilon, k)).$$

We do not indicate ε on the LHS of (10), since it is a purely technical parameter (cf. Lemma 4 in Section 4). It might be illustrative to notice that the marginals of $\bar{P}^{(n,k)}$ on Ω belong to $A(\varepsilon, k)$ in (8). In fact, the convex combination of elements in $A(\varepsilon, k)$

$$(11) \quad \int_{\Omega^n} L_n(\omega)(\cdot) d\bar{P}(\omega | L_n(\omega) \in A(\varepsilon, k)) = \frac{1}{n} \sum_{i=1}^n \bar{P}(\omega_i \in \cdot | L_n \in A(\varepsilon, k))$$

$$= \bar{P}(\omega_j \in \cdot | L_n \in A(\varepsilon, k))$$

is just the j -th marginal distribution of $\bar{P}^{(n,k)}$ on Ω , $j = 1, \dots, n$. Following (i) of Lemma 4 given in Section 4, $A(\varepsilon, k)$ is a convex set and hence contains the expression (11). Further discussions will be provided by Lemma 1 and Remark 2.

Let us now formulate our main result.

THEOREM 1. *Let the set $A_{a,b}$ in (3) satisfy condition (4). Thus there exists Csiszar's projection $Q \in A_{a,b}$ defined by*

$$(12) \quad H(Q | \bar{P}) = \min_{P \in A_{a,b}} H(P | \bar{P}).$$

Moreover, there exists a Markovian modification $Q^{(n,k)}$ of the conditional process $\bar{P}^{(n,k)}$ defined in (10) such that

$$(13) \quad H(\mathbf{Q}^{(n,k)} | Q_{\varepsilon,k}^n) \leq H(\bar{\mathbf{P}}^{(n,k)} | Q_{\varepsilon,k}^n)$$

where $Q_{\varepsilon,k}$ is Csiszar's projection of $\bar{\mathbf{P}}$ on $A(\varepsilon, k)$ in (8) defined analogously to (12).

Then the sequence of systems $((X_1, \dots, X_n), \mathbf{Q}^{(n,k)})$ performs 'propagation of chaos in entropy' of Definition 1 with Csiszar's projection Q in (12) as limiting distribution when n and k tend to infinity, provided $\log(\mathbf{Q}_m^{(n,k)}/Q_{\varepsilon,k}^m)$, $n, k \in \mathbf{N}$, is uniformly integrable w.r.t. Q^m for $m \in \mathbf{N}$.

COROLLARY 1. *The results of Theorem 1 for $m \in \mathbf{N}$ are also true, if for all sequences $(n(\nu), k(\nu))$, $\nu \in \mathbf{N}$, and $\forall \delta > 0$, $\exists M < \infty$, $\exists \delta_1 > 0$, $\exists \nu(m, \delta, M, \delta_1) \in \mathbf{N}$ such that*

$$(14) \quad \int_{B_{\nu, M}^{(m)}} \left(\frac{d\mathbf{Q}_m^{(n(\nu), k(\nu))}}{dQ_{\varepsilon, k(\nu)}^m} \right)^t dQ_{\varepsilon, k(\nu)}^m < \delta$$

for all $t \in (-\delta_1, \delta_1)$ and for all $\nu > \nu(m, \delta, M, \delta_1)$ where

$$(15) \quad B_{\nu, M}^{(m)} = \left\{ \left(\frac{d\mathbf{Q}_m^{(n(\nu), k(\nu))}}{dQ_{\varepsilon, k(\nu)}^m} \right)^\theta > M \text{ for } \theta = 1 \text{ or } -1 \right\}.$$

The proof is an immediate consequence of Lemma 3 providing $Q^m \ll \mathbf{Q}_m^{(n(\nu), k(\nu))}$, $\nu \in \mathbf{N}$, as well as a way to deal with (35) in the proof of Theorem 1. The integrability condition (14) shows the required accuracy of the approximation $A(\varepsilon, k)$ in (8) of $A_{a,b}$ in (3) with respect to the reference measure $\bar{\mathbf{P}}$.

Theorem 1 has some significance in quantum physics. In fact, we learn from [2], [19], [20], [21] that Schrödinger equations are related to so-called Schrödinger processes and vice versa, as far as solutions exist and the corresponding Schrödinger processes can be constructed. Moreover, [1] claims that Schrödinger processes are certain Csiszar projections. As a consequence, Theorem 1 yields

COROLLARY 2. *In the situation of Theorem 1, the distribution of a 'typical particle' under the law $\mathbf{Q}^{(n,k)}$ as n and k tend to infinity is determined by the Schrödinger process Q and hence by the related Schrödinger equation. In other words, a Schrödinger equation is a kind of 'Boltzmann equation' for a system of interacting particles represented by a system of interacting diffusion processes $((X_1, \dots, X_n), \mathbf{Q}^{(n,k)})$ as n and k tend to infinity.*

We can consider Schrödinger processes Q as Gibbs states of microscopic systems represented by $\mathbf{Q}^{(n,k)}$ for n and k tending to infinity. In fact, Q determines the rate function of the large deviation principle (40) called approximate Sanov property. Hence, Q is the 'most probable' diffusion process under the given circumstances and consequently, it is the limiting distribution postulated by the fundamental hypothesis of statistical mechanics. This illustrates the analogy to classical statistical mechanics, cf. e.g. [16].

Remark 1. Csiszar [5], Kemperman [11] and Kullback [14] show that the variation distance is dominated by the relative entropy according to

$$(16) \quad |P_1 - P_2|_{\text{var. } R} \leq \sqrt{2H(P_1|P_2)}$$

where $P_1, P_2, R \in M_1(\Omega)$ with $P_1 \ll P_2 \ll R$. Hence, Theorem 1 provides a comparably strong result. In fact, propagation of chaos in entropy yields propagation of chaos in variation because of (16) and propagation of chaos in variation obviously yields propagation of chaos in weak convergence.

3. Proof of Theorem 1

We denote a partial empirical distribution by

$$(17) \quad L_{m,n}(\omega) = \frac{1}{n-m} \sum_{j=m+1}^n \delta_{\omega_j}, \quad m < n, \quad \omega = (\omega_1, \dots, \omega_n) \in \Omega^n.$$

LEMMA 1. *Let us assume condition (4). (i) If $n > m2^k/\varepsilon$ is chosen large enough such that*

$$\frac{n-m}{n} L_{m,n}(\omega) \in A\left(\varepsilon - \tilde{\varepsilon}\left(\frac{m}{n}\right), k\right)$$

where

$$\tilde{\varepsilon}\left(\frac{m}{n}\right) = 2^k \frac{m}{n}$$

then

$$L_n(\omega) \in A(\varepsilon, k)$$

for \bar{P}^n -a.a. $\omega \in \Omega^n$.

(ii) *Moreover, there exists $n(m, \varepsilon, k) \in \mathbb{N}$, $n(m, \varepsilon, k) > m2^k/\varepsilon$, such that*

$$(18) \quad Q^m \ll \bar{P}_m^{(n,k)}$$

for all $n \geq n(m, \varepsilon, k)$, where $\bar{P}_m^{(n,k)}$ is the m -dimensional marginal distribution of $\bar{P}^{(n,k)}$ in (10).

Proof. Let $\mathcal{P}_k = \{B_1, \dots, B_k\}$ be a partition of \mathbf{R}^d as introduced in (7). In view of the definition of $A(\varepsilon, k)$ in (8), we first notice for $r = a, b$ that $L_n(\omega) \circ X_r^{-1}$ is not absolutely continuous with respect to $\bar{P} \circ X_r^{-1}$ on the set $\{\omega \in \Omega^n : \exists j = 1, \dots, n, \exists i = 1, \dots, k \text{ such that } \omega_j \in X_r^{-1}(B_i) \text{ where } \bar{P}(X_r^{-1}(B_i)) = 0\}$ which is however a \bar{P}^n -zero set. In terms of (17) we receive

$$\begin{aligned} & |L_n(\omega)(X_r^{-1}(B_i)) - q_r(B_i)| \\ & \leq \left| \frac{1}{n} \sum_{j=1}^m \delta_{\omega_j}(X_r^{-1}(B_i)) \right| + \left| \frac{1}{n} \sum_{j=m+1}^n \delta_{\omega_j}(X_r^{-1}(B_i)) - q_r(B_i) \right| \\ & = \left| \frac{m}{n} L_m(\omega)(X_r^{-1}(B_i)) \right| + \left| \frac{n-m}{n} L_{m,n}(\omega)(X_r^{-1}(B_i)) - q_r(B_i) \right| \end{aligned}$$

for $i=1, \dots, k$ and $r=a, b$. Hence (i) follows for \bar{P}^n -almost every $\omega \in \Omega^n$, if n is so large that

$$\frac{\varepsilon - \bar{\varepsilon}\left(\frac{m}{n}\right)}{2^k} = \frac{\varepsilon}{2^k} - \frac{m}{n}$$

is positive.

Let $n > m2^k/\varepsilon$ and let $B^{(m)} \in \sigma(\Omega^m)$. Because of $Q \ll \bar{P}$ in Lemma 4 (ii), part (i) implies

$$\begin{aligned} (19) \quad & Q^n(B^{(m)} \cap \{L_n \in A(\varepsilon, k)\}) \\ & \geq Q^n\left(B^{(m)} \cap \left\{\frac{n-m}{n}L_{m,n} \in A\left(\varepsilon - \bar{\varepsilon}\left(\frac{m}{n}\right), k\right)\right\}\right) \\ & = Q^m(B^{(m)}) Q^n\left(\frac{n-m}{n}L_{m,n} \in A\left(\varepsilon - \bar{\varepsilon}\left(\frac{m}{n}\right), k\right)\right). \end{aligned}$$

Since the law of large numbers claims that

$$Q^n\left(\frac{n-m}{n}L_{m,n} \in A\left(\varepsilon - \bar{\varepsilon}\left(\frac{m}{n}\right), k\right)\right) \rightarrow 1 \quad \text{as } n \nearrow \infty$$

there exists $n(m, \varepsilon, k) > m2^k/\varepsilon$ such that for all $n \geq n(m, \varepsilon, k)$

$$Q^n\left(\frac{n-m}{n}L_{m,n} \in A\left(\varepsilon - \bar{\varepsilon}\left(\frac{m}{n}\right), k\right)\right) > 0.$$

If $Q^m(B^{(m)}) > 0$, then inequality (19) yields

$$Q^n(B^{(m)} \cap \{L_n \in A(\varepsilon, k)\}) > 0$$

for all $n \geq n(m, \varepsilon, k)$ where $n(m, \varepsilon, k)$ is independent of the particular set $B^{(m)}$. As a consequence of $Q \ll \bar{P}$,

$$\bar{P}^n(B^{(m)} \cap \{L_n \in A(\varepsilon, k)\}) > 0$$

which implies

$$\bar{P}_m^{(n, k)}(B^{(m)}) = \bar{P}^{(n, k)}(B^{(m)}) > 0$$

concluding the proof.

LEMMA 2. *Let us assume (4). Then the limit*

$$\lim_{k \nearrow \infty} H(Q | Q_{\varepsilon, k}) = 0$$

in (38) of Csiszar projections Q and $Q_{\varepsilon, k}$ in Lemma 4 (ii) implies convergence in variation, i.e.

$$(20) \quad \int \left| \frac{dQ_{\varepsilon, k}}{d\bar{P}} - \frac{dQ}{d\bar{P}} \right| d\bar{P} \rightarrow 0 \quad \text{as } k \nearrow \infty.$$

Moreover, if $B_k \in \sigma(\Omega)$ for $k \in \mathbf{N}$, then

$$(21) \quad \lim_{k \nearrow \infty} \int_{B_k} \log\left(\frac{dQ}{dQ_{\varepsilon, k}}\right) dQ = 0.$$

Proof. The limit (20) follows from (38) because of (16).

We notice that the function $h(x) = x \log x - (x-1)$, $x \geq 0$, satisfies

$$(22) \quad \int h\left(\frac{dQ}{dQ_{\varepsilon, k}}\right) dQ_{\varepsilon, k} = \int \log\left(\frac{dQ}{dQ_{\varepsilon, k}}\right) \frac{dQ}{dQ_{\varepsilon, k}} dQ_{\varepsilon, k} = H(Q | Q_{\varepsilon, k})$$

which vanishes in (38) as k tends to infinity. Thus

$$\int_{B_k} \log\left(\frac{dQ}{dQ_{\varepsilon, k}}\right) dQ = \int_{B_k} h\left(\frac{dQ}{dQ_{\varepsilon, k}}\right) dQ_{\varepsilon, k} + \int_{B_k} \left(\frac{dQ}{dQ_{\varepsilon, k}} - 1\right) dQ_{\varepsilon, k}$$

also vanishes as k tends to infinity. In fact, the first term on the RHS can simply be estimated by (22) since h is non-negative, and the second term on the RHS tends to zero as a consequence of (20).

PROPOSITION 1. *Let (4) be provided and let Q and $Q_{\varepsilon, k}$ be the processes in (12) and in Lemma 4 (ii), respectively. Then there exists a Markovian modification $\mathbf{Q}^{(n, k)}$ of $\bar{\mathbf{P}}^{(k, k)}$ in (10) with property (13). Its m -dimensional marginal distributions $\mathbf{Q}_m^{(n, k)}$, $m \in \mathbf{N}$, satisfy*

$$(23) \quad Q^m \ll \mathbf{Q}_m^{(n, k)}$$

for all $n > n(m, \varepsilon, k)$ in Lemma 1 (ii), and

$$(24) \quad \lim_{k \nearrow \infty} \lim_{n \nearrow \infty} H(\mathbf{Q}_m^{(n, k)} | Q_{\varepsilon, k}^m) = 0$$

which yields

$$(25) \quad \lim_{k \nearrow \infty} \lim_{n \nearrow \infty} \int \left| \frac{dQ_{\varepsilon, k}^m}{d\bar{P}^m} - \frac{d\mathbf{Q}_m^{(n, k)}}{d\bar{P}^m} \right| d\bar{P}^m = 0$$

Moreover, if $B_{n, k}^{(m)} \in \sigma(\Omega^m)$ for $n, k \in \mathbf{N}$, then

$$(26) \quad \lim_{k \nearrow \infty} \lim_{n \nearrow \infty} \int_{B_{n, k}^{(m)}} \log(g_m^{(n, k)}) d\mathbf{Q}_m^{(n, k)} = 0$$

where

$$(27) \quad g_m^{(n, k)} = \frac{d\mathbf{Q}_m^{(n, k)}}{dQ_{\varepsilon, k}^m} \quad \text{for } n, k \in \mathbf{N}.$$

Proof. The existence of $\mathbf{Q}^{(n, k)}$ with (13) is provided by Lemma 4.2 in [1] because of (37) and (41). We notice that the marginal distributions of $\bar{\mathbf{P}}^{(n, k)}$ and $\mathbf{Q}^{(n, k)}$ on $\Omega_{m\nu+1} \times \dots \times \Omega_{m(\nu+1)}$ for $\nu = 0, 1, 2, \dots$ coincide because of the symmetry of $\bar{\mathbf{P}}^{(n, k)}$ in ω_i , $i = 1, \dots, n$. Thus Lemma 1 (ii) yields (23). Moreover,

$$(28) \quad H(\mathbf{Q}_m^{(n, k)} | Q_{\varepsilon, k}^m) \leq m \frac{n}{n-r} \frac{1}{n} H(\mathbf{Q}^{(n, k)} | Q_{\varepsilon, k}^n)$$

by Lemma 11.3 in [24] where $0 \leq r \leq m-1$. Hence (24) is a consequence of (13) and (41) which cause the RHS of (28) to vanish as n and k tend to infinity. Convergence in variation (25) follows from (24) by means of (16).

The limit (26) is received analogously to the proof of Lemma 2. In fact,

$$(29) \quad \int \log\left(\frac{dQ_m^{(n,k)}}{dQ_{\varepsilon,k}^m}\right) \frac{dQ_m^{(n,k)}}{dQ_{\varepsilon,k}^m} dQ_{\varepsilon,k}^m = \int h\left(\frac{dQ_m^{(n,k)}}{dQ_{\varepsilon,k}^m}\right) dQ_{\varepsilon,k}^m$$

where $h(x) = x \log x - (x-1)$ for $x \geq 0$. The LHS of (29) equals $H(Q_m^{(n,k)} | Q_{\varepsilon,k}^m)$ which vanishes in (24) as n and k tend to infinity. As a consequence,

$$\int_{B_{n,k}^{(m)}} \log\left(\frac{dQ_m^{(n,k)}}{dQ_{\varepsilon,k}^m}\right) dQ_m^{(n,k)} = \int_{B_{n,k}^{(m)}} h\left(\frac{dQ_m^{(n,k)}}{dQ_{\varepsilon,k}^m}\right) dQ_{\varepsilon,k}^m + \int_{B_{n,k}^{(m)}} \left(\frac{dQ_m^{(n,k)}}{dQ_{\varepsilon,k}^m} - 1\right) dQ_{\varepsilon,k}^m$$

vanishes as n and k tend to infinity. In fact, the first term on the RHS is estimated by (29) since h is non-negative, and the second term on the RHS tends to zero because of (25).

LEMMA 3. *Let us assume (4) and (14) in Corollary 1. If for $m \in \mathbb{N}$, $g_m^{(n(\nu), k(\nu))}$, $\nu \in \mathbb{N}$, is given in (27), then $\forall \delta > 0$, $\exists M < \infty$, $\exists \nu(\delta, M) \in \mathbb{N}$ such that*

$$(30) \quad \int_{\tilde{B}_{\nu, M}^{(m)}} |\log(g_m^{(n(\nu), k(\nu))})| dQ^m < \delta, \quad \forall \nu > \nu(\delta, M)$$

with $\tilde{B}_{\nu, M}^{(m)} = \{|\log g_m^{(n(\nu), k(\nu))}| > M\}$.

Proof. Referring to (27) we notice that for $\tilde{M} \in (0, \infty)$

$$\tilde{B}_{\nu, \tilde{M}}^{(m)} = \{|\log g_m^{(n(\nu), k(\nu))}| > \tilde{M}\} = \{(g_m^{(n(\nu), k(\nu))})^\theta > e^{\tilde{M}} \text{ for } \theta = 1 \text{ or } -1\}$$

which corresponds to $B_{\nu, M}^{(m)}$ in (15) with $M = e^{\tilde{M}}$. If, as provided by (14),

$$\begin{aligned} & \int_{B_{\nu, M}^{(m)}} \exp\{t \log(g_m^{(n(\nu), k(\nu))})\} dQ_{\varepsilon, k(\nu)}^m \\ &= \int_{B_{\nu, M}^{(m)} \cap \{(\log(g_m^{(n(\nu), k(\nu))}))^+ > 0\}} \exp\{t(\log(g_m^{(n(\nu), k(\nu))}))^+\} dQ_{\varepsilon, k(\nu)}^m \\ & \quad + \int_{B_{\nu, M}^{(m)} \cap \{(\log(g_m^{(n(\nu), k(\nu))}))^- > 0\}} \exp\{-t(\log(g_m^{(n(\nu), k(\nu))}))^-\} dQ_{\varepsilon, k(\nu)}^m < \frac{\tilde{\delta}}{2} \end{aligned}$$

for all $t \in (-\delta_1, \delta_1)$ where $1 < M < \infty$, then

$$(31) \quad \int_{B_{\nu, M}^{(m)}} \exp\{t |\log(g_m^{(n(\nu), k(\nu))})|\} dQ_{\varepsilon, k(\nu)}^m < \tilde{\delta}$$

for all $t \in (-\delta_1, \delta_1)$. In order to arrive at (30), the inequality

$$ab < a \log a + e^b \quad \text{for } a, b \geq 0$$

in section 15 of [3] is applied to $a = dQ^m / dQ_{\varepsilon, k(\nu)}^m$ and $b = t |\log(g_m^{(n(\nu), k(\nu))})|$ for $t \in (0, \delta_1)$. Thus

$$\begin{aligned} & \int_{B_{\nu, M}^{(m)}} |\log(g_m^{(n, \nu, k(\nu))})| dQ^m \\ &= \frac{1}{t} \int_{B_{\nu, M}^{(m)}} t |\log(g_m^{(n, \nu, k(\nu))})| \frac{dQ^m}{dQ_{\varepsilon, k(\nu)}^{(m)}} dQ_{\varepsilon, k(\nu)}^m \\ &< \frac{1}{t} \int_{B_{\nu, M}^{(m)}} \frac{dQ^m}{dQ_{\varepsilon, k(\nu)}^{(m)}} \log\left(\frac{dQ^m}{dQ_{\varepsilon, k(\nu)}^{(m)}}\right) dQ_{\varepsilon, k(\nu)}^m \\ & \quad + \frac{1}{t} \int_{B_{\nu, M}^{(m)}} \exp\{t |\log(g_m^{(n, \nu, k(\nu))})|\} dQ_{\varepsilon, k(\nu)}^m < \delta \end{aligned}$$

since the first term on the RHS vanishes in (21) as $k(\nu)$ tends to infinity while the second term on the RHS can be made arbitrarily small by a large M on the basis of (31).

Remark 2. In order to verify propagation of chaos, i.e. (1) and (2), we essentially have for use the limit (38) which is a consequence of (37) and (39) and the limit (24) which is a consequence of (13), (28) and (41). While (24) provides some integrability properties in Proposition 1, Lemma 3 shows some consequences of (38) under assumption (14). Let us briefly resume our approach. Because of (37) and (38), $Q \ll Q_{\varepsilon, k} \ll \bar{P}$, where $Q_{\varepsilon, k}$ is Csiszar's projection of \bar{P} on $A(\varepsilon, k)$ in Lemma 4 (ii). Following Proposition 1, $Q^m \ll Q_m^{(n, k)}$ for $n > n(m, \varepsilon, k)$, where $Q^{(n, k)}$ is the Markovian modification of $\bar{P}^{(n, k)}$ in (10). We notice that (20) and (25) yield

$$\lim_{k \nearrow \infty} \lim_{n \nearrow \infty} \int \left| \frac{dQ_m^{(n, k)}}{d\bar{P}^m} - \frac{dQ^m}{d\bar{P}^m} \right| d\bar{P}^m = 0$$

i.e. convergence in variation. In order to get $\lim_{k \nearrow \infty} \lim_{n \nearrow \infty} H(Q^m | Q_m^{(n, k)}) = 0$, i.e. (1), $Q_m^{(n, k)}$ has to satisfy additionally a uniform integrability condition with respect to the limiting distribution Q^m . Lemma 2 and Proposition 1 refine (20) and (25) on the basis of (38) and (24), respectively. In Corollary 1, the assumption (14) is rather a condition on the approximating sequence $Q_{\varepsilon, k}$ than on Q itself.

Proof of Theorem 1. Because of (4), Lemma 4 provides the unique Csiszar projections Q and $Q_{\varepsilon, k}$ of \bar{P} on $A_{a, b}$ and on $A(\varepsilon, k)$, $\varepsilon > 0$, $k \in \mathbf{N}$, respectively. Moreover, Proposition 1 yields a Markovian modification $Q^{(n, k)}$ of $\bar{P}^{(n, k)}$ which satisfies (13). Following Proposition 2, $Q^{(n, k)}$ and $Q_{\varepsilon, k}^n$ satisfy (41) because of (13). In our situation based on (10), which is particularly symmetrical in ω_i , $i=1, \dots, n$, property (1) for $m=1$ implies property (2) in Definition 1.

In order to show (1) for any $m \in \mathbf{N}$, we notice first that $H(Q^m | Q_m^{(n, k)})$ is well-defined. In fact, $Q^m \ll Q_m^{(n, k)}$, i.e. $n \geq n(m, \varepsilon, k)$ in Lemma 1 (ii), holds as a consequence of (38) and the assumed uniform integrability of $\log(dQ_m^{(n, k)}/dQ_{\varepsilon, k}^m)$ w.r.t. Q^m . Since $Q_m^{(n, k)}$ takes different positions in the relative entropy expressions in (1) and (24), respectively, we consider

$$(32) \quad \frac{dQ^m}{d\mathbf{Q}_m^{(n,k)}} = \frac{dQ^m}{dQ_{\varepsilon,k}^m} \frac{dQ_{\varepsilon,k}^m}{d\mathbf{Q}_m^{(n,k)}}, \quad Q^m\text{-a.s.}$$

for $n \geq n(m, \varepsilon, k)$ where the first factor on the RHS of (32) exists because of (38) and the second factor on the RHS of (32) exists on $\text{supp } \mathbf{Q}_m^{(n,k)}$ because of (24). As a consequence of (32),

$$(33) \quad H(Q^m | \mathbf{Q}_m^{(n,k)}) = mH(Q | Q_{\varepsilon,k}) - \int \log(g_m^{(n,k)}) dQ^m$$

where $g_m^{(n,k)}$, $n, k \in \mathbf{N}$, is defined in (27).

Let us investigate the limit behaviour of the second term of the RHS of (33). Proposition 1 claims that

$$(34) \quad \lim_{k \nearrow \infty} \lim_{n \nearrow \infty} \int_{\{|\log(g_m^{(n,k)})| > M\}} \log(g_m^{(n,k)}) d\mathbf{Q}_m^{(n,k)} = 0$$

for any $M > 0$. Thus we are ready to claim that

$$(35) \quad \left| \int \log(g_m^{(n,k)}) dQ^m - \int \log(g_m^{(n,k)}) d\mathbf{Q}_m^{(n,k)} \right| \\ \leq \left| \int \log(g_m^{(n,k)}) dQ^m - \int_{\{|\log(g_m^{(n,k)})| \leq M\}} \log(g_m^{(n,k)}) dQ^m \right| \\ + \left| \int_{\{|\log(g_m^{(n,k)})| \leq M\}} \log(g_m^{(n,k)}) dQ^m - \int_{\{|\log(g_m^{(n,k)})| \leq M\}} \log(g_m^{(n,k)}) dQ_{\varepsilon,k}^m \right| \\ + \left| \int_{\{|\log(g_m^{(n,k)})| \leq M\}} \log(g_m^{(n,k)}) dQ_{\varepsilon,k}^m - \int_{\{|\log(g_m^{(n,k)})| \leq M\}} \log(g_m^{(n,k)}) d\mathbf{Q}_m^{(n,k)} \right| \\ + \left| \int_{\{|\log(g_m^{(n,k)})| \leq M\}} \log(g_m^{(n,k)}) d\mathbf{Q}_m^{(n,k)} - \int \log(g_m^{(n,k)}) d\mathbf{Q}_m^{(n,k)} \right| \rightarrow 0$$

as $n, k \nearrow \infty$. In fact, the first term on the RHS of (35) becomes arbitrarily small for M large because of the supposed uniform integrability of $\log(g_m^{(n,k)})$, $n, k \in \mathbf{N}$, w.r.t. Q^m . The fourth term on the RHS of (35) can be made small for large M because of (34) following from Proposition 1. In case of the second and the third term on the RHS of (35) we refer to (20) and (25), respectively.

Finally we obtain (1) through (33) by means of

$$\begin{aligned} & |H(Q^m | \mathbf{Q}_m^{(n,k)})| \\ & \leq m |H(Q | Q_{\varepsilon,k})| + \left| \int \log(g_m^{(n,k)}) dQ^m - \int \log(g_m^{(n,k)}) d\mathbf{Q}_m^{(n,k)} \right| \\ & \quad + |H(\mathbf{Q}_m^{(n,k)} | Q_{\varepsilon,k}^m)| \rightarrow 0 \end{aligned}$$

as $n, k \nearrow \infty$. In fact, the first term on the RHS vanishes in (38) as k tends to infinity, the second term becomes small as n and k tend to infinity because of (35) and the third term vanishes as shown in (24).

4. Csiszar’s projection and the approximate Sanov property

Here we briefly arrange some of the quoted results for the reader’s convenience.

LEMMA 4 (Csiszar [6], Aebi and Nagasawa [1]). *Suppose that any reference measure $\bar{P} \in M_1(\Omega)$ and $A_{a,b}$ in (3) satisfy (4). Then:*

(i) $A_{a,b}$ and $A(\varepsilon, k)$ in (8) are convex, variation closed and satisfy

$$(36) \quad A_{a,b} = \bigcap_{k \in \mathbf{N}} A(\varepsilon, k), \quad \forall \varepsilon > 0.$$

(ii) *There exist uniquely the Csiszar projections Q and $Q_{\varepsilon,k}$ of \bar{P} on $A_{a,b}$ and on $A(\varepsilon, k)$, $k \in \mathbf{N}$, $\varepsilon > 0$, respectively, defined according to (12). They satisfy*

$$(37) \quad \lim_{k \nearrow \infty} H(Q_{\varepsilon,k} | \bar{P}) = H(Q | \bar{P}), \quad \forall \varepsilon > 0$$

and

$$(38) \quad \lim_{k \nearrow \infty} H(Q | Q_{\varepsilon,k}) = 0, \quad \forall \varepsilon > 0.$$

Proof. Details are found in Lemma 3.5 of [1]. In case of the crucial convergence in (38), we refer to inequality (2.14) in [6]. It provides

$$(39) \quad H(Q | \bar{P}) - H(Q_{\varepsilon,k} | \bar{P}) \geq H(Q | Q_{\varepsilon,k})$$

because of $Q \in A_{a,b} \subset A(\varepsilon, k)$. As a consequence of (37), the LHS of (39) vanishes as k tends to infinity.

PROPOSITION 2 (Aebi and Nagasawa [1], revised). *Let (4) be provided. Then:*

(i) *The set $A_{a,b}$ possesses the approximate Sanov property*

$$(40) \quad \lim_{k \nearrow \infty} \lim_{n \nearrow \infty} \frac{1}{n} \log \bar{P}(L_n \in A(\varepsilon, k)) = -H(Q | \bar{P}), \quad \forall \varepsilon > 0$$

where Q is Csiszar’s projection of \bar{P} on the set $A_{a,b}$ defined in (12).

(ii) *The system (X_1, \dots, X_n) is asymptotically quasi-independent under $\bar{P}^{(n,k)}$ in (10) with limiting distribution $Q_{\varepsilon,k}$ for each $k \in \mathbf{N}$ and $\varepsilon > 0$, i.e.*

$$(41) \quad \lim_{n \nearrow \infty} \frac{1}{n} H(\bar{P}^{(n,k)} | Q_{\varepsilon,k}^n) = 0, \quad \forall k \in \mathbf{N}, \forall \varepsilon > 0.$$

Acknowledgement. The author would like to thank Professor Tokuzo Shiga for the stimulating discussions and the generous hospitality at the Tokyo Institute of Technology.

REFERENCES

- [1] R. AEBI AND M. NAGASAWA, Large deviations and the propagation of chaos for Schrödinger processes, *Probab. Theory Related Fields*, **94** (1992), 53-68.
- [2] R. AEBI, Diffusions with singular drift related to wave functions, *Probab. Theory Related Fields*, **96** (1993), 107-121.
- [3] E. F. BECKENBACH AND R. BELLMANN, *Inequalities*, Springer-Verlag, Berlin, 1961.
- [4] L. BOLTZMANN, *Vorlesungen über Gastheorie*, J. A. Barth Verlag, Leipzig, 1896.
- [5] I. CSISZAR, Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.*, **2** (1967), 299-318.
- [6] I. CSISZAR, I -divergence geometry of probability distributions and minimization problems, *Ann. Probab.*, **3** (1975), 146-158.
- [7] I. CSISZAR, Sanov property, generalized I -projection and a conditional limit theorem, *Ann. Probab.*, **12** (1984), 768-793.
- [8] D. A. DAWSON AND J. GARTNER, Large deviations, free energy functional and quasi-potential for a mean field model of interacting diffusions, *Mem. Amer. Math. Soc.*, vol. **78** (1989), no. 398.
- [9] H. FÖLLMER, Random fields and diffusion processes, *École d'Été de Probabilités de Saint Flour XV-XVII, 1985-87*, *Lecture Notes in Math.*, **1362**, Springer-Verlag, Berlin, 1988, 101-203.
- [10] E. GUTKIN AND M. KAC, Propagation of chaos and Burger's equation, *SIAM J. Appl. Math.*, **43** (1983), 971-980.
- [11] J. H. B. KEMPERMAN, On the optimum rate of transmitting information, *Probability and Information Theory*, Springer-Verlag, Berlin, 1969, 126-169.
- [12] A. I. KHINCHIN, *Mathematical Foundations of Information Theory*, Dover Publications, New York, 1957.
- [13] J. F. C. KINGMAN AND S. J. TAYLOR, *Introduction to Measure and Probability*, Cambridge University Press, Cambridge, 1977.
- [14] S. KULLBACK, A lower bound for discrimination information in terms of variation, *IEEE Trans. Information Theory*, **IT-13** (1967), 126-127.
- [15] S. KUSUOKA AND Y. TAMURA, Gibbs measures for mean field potentials, *J. Fac. Sci. Univ. Tokyo Sect. IA Math.*, **31** (1984), 223-245.
- [16] O. E. LANFORD, Entropy and equilibrium states in classical statistical mechanics, *Statistical Mechanics and Mathematical Problems* (ed. Lenard, A.), *Lecture Notes in Phys.*, **20**, Springer-Verlag, Berlin, 1973, 1-113.
- [17] H. P. MCKEAN, A class of Markov processes associated with non-linear parabolic equations, *Proc. Nat. Acad. Sci. U. S. A.*, **56** (1966), 1907-1911.
- [18] H. P. MCKEAN, Propagation of chaos for a class of nonlinear parabolic equations, *Lecture Series in Differential Equations*, Catholic Univ. (1967), 41-57.
- [19] M. NAGASAWA AND H. TANAKA, Propagation of chaos for diffusing particles of two types with singular mean field interaction, *Probab. Theory Related Fields*, **71** (1986), 69-83.
- [20] M. NAGASAWA AND H. TANAKA, Diffusion with interactions and collisions between coloured particles and the propagation of chaos, *Probab. Theory Related Fields*, **74** (1987), 161-198.
- [21] M. NAGASAWA AND H. TANAKA, A proof of the propagation of chaos for diffusion processes with drift coefficients not of average form, *Tokyo J. Math.*, **10** (1987), 403-418.

- [22] M. NAGASAWA, Transformations of diffusion and Schrödinger processes, *Probab. Theory Related Fields*, **82** (1989), 109-136.
- [23] M. NAGASAWA, Can the Schrödinger equation be a Boltzmann equation? *Diffusion Processes and Related Problems in Analysis*, M. Pinsky (ed.), *Progr. Probab.*, **22**, Birkhäuser Boston Inc., Boston, 1990, 155-200.
- [24] M. NAGASAWA, *Schrödinger Equations and Diffusion Theory*, *Monogr. Math.*, **86**, Birkhäuser Verlag, Basel, 1993.
- [25] E. NELSON, *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton, 1967.
- [26] K. OELSCHLÄGER, *Many-particle Systems and the Continuum Description of Their Dynamics Habilitation*, Univ. Heidelberg, 1989.
- [27] E. SCHRÖDINGER, Über die Umkehrung der Naturgesetze *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse* (1931), 144-153.
- [28] T. SHIGA AND H. TANAKA, Central limit theorem for a system of Markovian particles with mean field interaction, *Z. Wahrsch. Verw. Gebiete*, **69** (1985), 439-459.
- [29] A.S. SZNITMAN, Topics in propagation of chaos, *École d'Été de Probabilités de Saint Flour XIX-1989, Lecture Notes in Math.*, **1464**, Springer-Verlag, Berlin, 1989, 165-251.
- [30] H. TANAKA, Limit theorems for certain diffusion processes with interaction, *Stochastic Analysis* (ed. Itô, K.), Kinokuniya, Tokyo, 1984, 469-488.

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF BERNE
CH-3012 BERNE
SWITZERLAND
e-mail: aebi@math-stat.unibe.ch