Homogeneous Balance Equations for Measures: Errors of the Stochastic Solution

N. Golyandina and V. Nekrutkin

Abstract — Stochastic (Monte Carlo) solutions of the general balance ordinary differential equations for measures are considered. For this purpose a wide class of n-particle processes is studied from the viewpoint of the main terms of asymptotical ($n \to \infty$) bias and variance for the corresponding estimates. Boltzmann-like and linear equations are considered as the particular cases of balance equations, several new and well-known algorithms are theoretically compared from the statistical and complexity points of view.

1991 MS classification: primary 65C05, secondary 82C80, 76P05, 34G20.

Key words: differential equations for measures, balance equations, Boltzmann-like equations, linear equations, stochastic solution, measure-valued processes, n-particle processes, bias and variance main terms, algorithm comparison, operator semigroup theory.

1 Introduction

The aim of this paper is to construct a class of stochastic processes solving general (up to some technical restrictions) homogeneous balance ordinary differential equations (o.d.e.'s) for measures and to study the corresponding approximation errors. By the balance o.d.e. we understand the equation

$$\frac{d\mu_t}{dt} = B(\mu_t) - \mu_t, \qquad \mu_0 = \mu \in H, \tag{1.1}$$

where μ_t is a measure-valued function, H is the set of distributions on a measurable space $(D, \mathcal{B})$ and B is a mapping $H \mapsto H$. Therefore, if there exists a solution $\mu_t(\mu)$ of (1.1), it is a probability measure for any t.

Examples of balance equations for measures are well-known. The first is the linear o.d.e. with $B(\nu) = \int_D T(\,\cdot\,; u)\nu(du)$, where $T(\,\cdot\,; u)$ is a distribution on $\mathcal{B}$ for any u. Then (1.1) is a sort of the Kolmogorov equation for the purely jump Markov process $\zeta(t)$ with the initial distribution μ, jump law

T and exponential (with the unit rate) time intervals between jumps. This process solves (1.1) in the sense that $\mathcal{L}(\zeta(t)) = \mu_t$ (from now on $\mathcal{L}(\xi)$ stands for the distribution of a random variable ξ). Therefore if $\zeta_1(t), \ldots, \zeta_n(t)$ are independent copies of $\zeta(t)$, we use $n^{-1}\sum_{i=1}^{n} g(\zeta_i(t))$ as the standard unbiased estimator of $\psi(\mu_t) = \int_D g\, d\mu_t$.

Boltzmann-like equations are much more interesting example of balance equations for measures. Here (dealing with bounded total cross sections and using the proper time rescaling) $B(\nu) = \int_{D^2} T(\,\cdot\,; u_1, u_2)\nu(du_1)\nu(du_2)$, where $T(\,\cdot\,; u_1, u_2)$ is a distribution on $\mathcal{B}$ for any $u_1, u_2 \in D$. It is well-known that such an equation has the unique solution $\mu_t(\mu) \in H$ for any $\mu \in H$. Moreover, as in the linear case, there exists the process $\zeta(t)$ with $\mathcal{L}(\zeta(t)) = \mu_t$ (see, for example, [5, ch. 4], where $\zeta(t)$ is called a *collisional* process). Note that we study homogeneous Boltzmann-like equations and therefore deal with homogeneous variants of corresponding processes and algorithms. Algorithmically, the collisional process may be easily simulated, but its simulational complexity grows exponentially in t.

The other approach to the stochastic solution of Boltzmann-like equations is the use of n-particle processes $\eta^{(n)}(t) = (\zeta_1^{(n)}(t), \ldots, \zeta_n^{(n)}(t))$. Such processes may be described as the special jump Markov processes with the phase space D^n and the initial distribution $\mu^{\otimes n} = \mu \otimes \ldots \otimes \mu$ (n times). Two main variants of n-particle processes are thoroughly investigated: one is called a *Nanbu-like* or a *simple mean-field* and the other — a *Bird-like* or a *binary mean-field* process. Numerous results, starting with [13], were obtained for these processes (and many of them are valid in the general non-homogeneous case). Restricting ourselves to fixed $t > 0$, the first assertion to be proved here is the Law-of-Large-Numbers result

$$\widehat{\psi}_n(g, t) = \frac{1}{n}\sum_{i=1}^{n} g\left(\zeta_i^{(n)}(t)\right) \xrightarrow[n\to\infty]{P} \psi(\mu_t) = \int_D g\, d\mu_t, \tag{1.2}$$

which must be valid for a wide class of bounded measurable functions g. The second problem is the investigation of the statistical properties for $\widehat{\psi}_n(g, t)$ as the estimator of $\psi(\mu_t)$, that is, its bias $\mathbf{E}\widehat{\psi}_n(g, t) - \psi(\mu_t)$ and mean square error $\mathbf{E}(\widehat{\psi}_n(g, t) - \psi(\mu_t))^2$. Generally speaking, answers for both questions are at hand both for Nanbu and Bird-like processes (see [18] and [12]): for any bounded measurable g the convergence (1.2) holds; the main terms have a form $c(t)/n$, where the coefficients $c(t)$ may be explicitly written down, but are expressed through certain complicated series. Thus it is difficult to analyze these coefficients and, say, to compare different variants of n-particle processes from the viewpoint of their computational complexity.

Note that these results are valid for more general non-homogeneous Boltzmann-like equations, where particles are moving between jumps with trajectories guided by general Markov families. As for functional analogies of (1.2), they may be found in [14], [11] and [15], where the strong convergence in the space of signed measures on the corresponding Skorohod space as well as the rate of such a convergence is investigated. The polynomial structure of B is considerably used by all these approaches.

For linear equations, we may gather up the independent processes $\zeta_i(t)$ in the process $\eta^{(n)}(t) = (\zeta_1(t), \ldots, \zeta_n(t)) \in D^n$, which occurs to be jump and Markov. Therefore the traditional Monte Carlo procedure for linear equations may be expressed in terms of n-particle processes (see Section 8).

Now let us turn to the general balance equations. Can one construct a sequence of n-particle processes $\eta^{(n)}(t) = (\zeta_1^{(n)}(t), \ldots, \zeta_n^{(n)}(t))$, solving (1.1) in the sense of (1.2)? What is the way to describe a class of such processes? What can we say about statistical errors of the corresponding estimates? Can we theoretically compare these errors for different n-particle processes? That is the area of problems we are to solve here.

For this purpose one must choose the proper mathematical technique, but before discussing this choice, it is convenient to display one (the simplest) variant of the n-particle process, solving (1.1). For any fixed distribution ν, $B(\nu) = B(\,\cdot\,;\nu)$ is also the distribution on $\mathcal{B}$. If we denote by δ_u the Dirac measure, concentrated at the point u, and take

$$\nu = \nu(\mathbf{u}) = n^{-1}\sum_{i=1}^{n}\delta_{u_i}, \qquad \mathbf{u} = (u_1, \ldots, u_n) \in D^n,$$

then $\mathcal{P}(\,\cdot\,;\mathbf{u}) \stackrel{def}{=} B(\,\cdot\,;\nu)$ is a distribution for any $\mathbf{u} \in D^n$ and, supposing we can simulate this distribution, the whole process may be described as follows:

First, we take independent $\zeta_i^{(n)}(0)$ with $\mathcal{L}(\zeta_i^{(n)}(0)) = \mu$. Then we choose the exponential rate of $\eta^{(n)}(t)$ (that is, the parameter of exponential distribution of its first jump moment) as $\lambda_n = n$ and its jump law as

$$T(dv_1 \ldots dv_n; u_1, \ldots, u_n) = n^{-1}\sum_{i=1}^{n}\mathcal{P}(dv_i,; u_1, \ldots, u_n) \otimes \bigotimes_{j\neq i}\delta_{u_j}(dv_j).$$

Algorithmically, if the process has coordinates $u_1, \ldots, u_n$ just before a jump, we randomly choose a particle, and it jumps to a new position due to the distribution $\mathcal{P}(\,\cdot\,; u_1, \ldots, u_n)$, while all other particles stay still. It occurs (see Section 6) that under smoothness conditions on B, the convergence (1.2) holds for any continuous bounded g and $\mathbf{E}(\widehat{\psi}_n(g,t) - \psi(\mu_t))^2 = V(t)/n + o(n^{-1})$ with $V(t) = V_0(t) + V_1(t)$,

$$V_0(t) = \frac{1}{2} \int\limits_{D^2} f_1^2(t, \mu; v, u)\mu(dv)\mu(du), \tag{1.3}$$

and

$$V_1(t) = \int\limits_0^t d\tau \int\limits_{D^2} f_1^2(\tau, \mu_{t-\tau}; v, u) B(dv; \mu_{t-\tau})\mu_{t-\tau}(du). \tag{1.4}$$

As for f_1, it may be explained as follows (see Sections 2 and 4 for details). If we take $\mu_t(\nu)$ as the smooth solution of (1.1) depending on the initial value ν, we may consider its derivative mapping

$$\partial^1 \mu_t(\nu; \,\cdot\,) = \frac{d\mu_t(\nu)}{d\nu}\,(\,\cdot\,),$$

which is the linear mapping on the corresponding space of signed measures (i.e., a sort of the differential, which is expressed through the first derivative, but is a linear function by definition). Then

$$f_1(t, \nu; v, u) \stackrel{def}{=} \int\limits_D g\, d\Big(\partial^1 \mu_t(\nu;\, \delta_v - \delta_u\,)\Big) = \frac{d\psi(\mu_t(\nu))}{d\nu}\,(\delta_v - \delta_u). \tag{1.5}$$

The technique applied to achieve such results is placed around the Trotter-Kato theorem and the Duhamel perturbation formula for contraction operator semigroups on Banach spaces (e.g., [6], [4]). Roughly speaking, the Trotter-Kato theorem tells us that a sequence of contraction operator semigroups converges to its limit semigroup if the corresponding sequence of generators converges to the limit generator, while the Duhamel formula expresses the rate of semigroup convergence through the rate of generator convergence.

As generators of jump Markov families have very simple structure (as well as generators of semigroups, associated with o.d.e.'s), it is natural to apply this technique to them. Such an approach is usual if we replace H by a certain subset of the euclidean space (this corresponds to finite D). One may consult [1] for details (see [17], [16] for the simulational viewpoint). Note that dealing with this technique we obtain all results not only for linear functions (i.e., integrals), but for any sufficiently smooth functional ψ on H.

When applying these ideas to the equation (1.1), several difficulties come to light. First, the natural Banach space here is the space E of all finite signed measures on $(D, \mathcal{B})$ with the variation norm, which induces the strong topology on E. But this space has bad topological features, for example, it may be non-separable and the set H of all probability measures may be non-compact with respect to this topology. From the practical viewpoint, one can not approximate a non-discrete measure with the help of discrete ones in variation.

Therefore we use two topologies on E: the first is the strong topology, and the second is the usual weak topology induced by the weak convergence of measures (by the way, this means that $\mathcal{B}$ is the Borel σ-algebra here). Note that if D is finite, both topologies coincide.

Then, the processes under consideration must be measure-valued. But an arbitrary measure (we use this term instead of 'signed measure' for the brevity) is not a constructive object from the Monte Carlo viewpoint. Say, if we deal with a jump Markov process in $\mathbb{R}^d$, there occurs no principal difficulties for the simulation of a jump, as its value is a random d-dimensional vector with a known distribution. But if the process is measure-valued, then its jump value is a random (signed) measure and, generally, one can't explicitly simulate such a random object. Therefore we must restrict ourselves to a class of suitable jump Markov measure-valued processes.

Here we consider the *empirical* n-particle processes with the phase space H_n, consisting of discrete distributions with n atoms of equal weights. Such processes may be obtained from the 'usual' processes $\eta^{(n)}(t) \in D^n$ by

$$\xi_n(t) = n^{-1} \sum_{i=1}^{n} \delta_{\zeta_i(t)} \in H_n, \qquad \eta^{(n)}(t) = (\zeta_1(t), \ldots, \zeta_n(t)) \in D^n,$$

and therefore are in a way constructive.

Thus our aim may be formulated as follows. Dealing with a (more or less) general homogeneous balance o.d.e. for measures, we want to describe a class of empirical jump Markov n-particle processes, solving this equation in the sense of (1.2) for any bounded continuous function g. Moreover we are to obtain explicit expressions for the main terms of both bias and mean square deviation of the estimate $\widehat{\psi}_n(g, t)$.

As Boltzmann-like and linear equations are the important particular cases of balance equations, it is interesting to compare old and new methods of their stochastic solution from the viewpoint of simulational complexity.

Now let us describe the contents of the paper. It is divided into two parts, the first devoted to general balance equations (Sections 2 – 6) and the second (Sections 7 and 8) — to the Boltzmann-like and linear ones.

In Section 2 derivatives with respect to measures as well as two (strong and weak) topologies on measure spaces are studied. General equations for measures are considered, and, under certain smoothness conditions on their solutions, generators of the corresponding semigroups are obtained (Lemma 2.3). The important restraint on the equation is that its solution $\mu_t(\mu)$ does not leave certain weak compact V with non-empty strong interior, if the initial value μ belongs to V.

Section 3 is devoted to the preliminary approximation results, based on a variant of the Trotter-Kato theorem and the Duhamel formula. Theorem 3.2 gives us general sufficient conditions for the convergence

$$\sup_{\nu\in V_n} |\mathbf{E}_\nu \psi(\xi_n(t)) - \psi(\mu_t(\nu))| \underset{n\to\infty}{\longrightarrow} 0,$$

where V_n approximates V, $\psi \in \mathbf{C}(V)$ and $\xi_n(t)$ is a measure-valued Markov process from the Markov family with the phase space V_n and distributions $\mathbf{P}_\nu$. These conditions are specialized for jump Markov families (see Theorem 3.3) in terms of their infinitesimal characteristics, and Theorem 3.4 gives the suitable residual term of such an approximation for $\psi \in \mathbf{C}^2(V)$.

These results are somewhat preliminary. Section 4 starts the investigation of the balance equation (1.1). The existence/uniqueness result under the smoothness conditions on its right-hand side G is proved (Lemmas 4.1 and 4.2). Though the set H of all distributions has the empty strong interior and is not a linear space, it is shown that all main approximation propositions are valid here, if one considers H-derivative mappings instead of the usual derivatives. Theorem 4.1 (which is a reformulation of Theorem 3.4 for balance equations) is the basis of the further residual analysis.

The empirical jump measure-valued Markov families, approximating the balance equations are considered in Section 5. First, such families are put into correspondence with the jump Markov families in D^n (Lemma 5.1). Then, under mild invariance and continuity conditions, the class of (n,k)-particle families is described and the general approximation result (Theorem 5.1) is proved. More precisely, (5.12) gives us the suitable form of the difference $\Delta_n(\psi,\nu) = \mathbf{E}_\nu \psi(\xi_n(t)) - \psi(\mu_t(\nu))$.

This representation is the base of Theorem 5.3 (with Corollary 5.1 and Remarks 5.2, 5.3). Here the main term of $\Delta_n(\psi,\nu)$ is specificated for a wide class of (n,k)-particle processes. This term has the order n^{-1} with the constant, expressed through the second (and, maybe, the first) H-derivative of $\mu_t(\mu)$ with respect to the initial value μ.

The resulting general theoretical propositions of the paper are placed in Section 6. Here we take into account simulation of the initial distribution μ and turn to the statistical properties of estimates for $\psi(\mu_t)$ with smooth ψ.

Theorem 6.1 shows that under the conditions of the semigroup convergence and assuming that $\mathcal{L}(\xi_n(0)) \stackrel{w}{\to} \delta_\nu$, the distributions of the measure-valued processes $\xi_n(\,\cdot\,)$ weakly converge in the space of probability measures on the Skorohod space $\mathrm{D}([0,\infty), D)$ to the distribution, concentrated at the measure-valued function $\mu_\cdot(\nu)$ (this result is the corollary of Theorem 2.11 [6, Ch. 4]). Therefore, among other things, the corresponding Monte Carlo procedure gives the consistent estimate of $\psi(\mu_t)$ for any weakly continuous ψ.

Lemma 6.1 tells that the mean square error of the estimate may be decomposed into two summands (such as (1.3) and (1.4)), the first responsible for the simulation of the initial distribution and the second — for the simulation of the process itself. Theorem 6.2 gives the explicit form of the main term for the first summand if the initial value $\xi_n(0)$ of the process $\xi_n(t)$ is the empirical measure produced by i.i.d. random variables, and $\psi \in \mathbf{C}^3(H)$. The order of this term is $O(n^{-1})$. Theorem 6.3 shows that the proper variant of stratification reduces this order to $o(n^{-1})$ (proofs of both theorems may be found in [10]). Finally, Theorem 6.4 with Remark 6.3 gives the entire main terms of bias and mean square error for a wide class of (n,k)-particle processes, general balance o.d.e.'s for measures and smooth nonlinear functionals.

The second part of the paper (Sections 7 and 8) is devoted to examples of general theory implementation: Boltzmann-like and linear equations.

For Boltzmann-like equations, we proof the existence/uniqueness result of their smooth solutions (Lemma 7.1) and describe several algorithms for their stochastic estimation: the Simplest algorithm, the Nanbu-like algorithm (both with $k=1$) and a class of Bird-like algorithms (with $k=2$, that is, with binary collisions). Dealing with mean square errors (moreover, assuming that the proper stratification is used when simulating the initial distribution), we obtain the corresponding *asymptotical variance coefficients*, that is, the values $\lim n\mathbf{E}(\psi(\xi_n(t)) - \psi(\mu_t))^2$, $n \to \infty$.

Lemma 7.2 gives comparison results for the Nanbu-like and several Bird-like variance coefficients. The most interesting inequality tells that any Bird-like algorithm has at most twice greater asymptotical variance coefficient than the Nanbu-like one. This means that if the computational cost of a binary collision is basically the same as that of a single particle jump in the Nanbu algorithm, the complexity of the Bird-like algorithm is smaller then that of the Nanbu one. Such situation occurs in rarefied gas dynamics, where $T(\,\cdot\,; u_1, u_2)$ stands for the uniform distribution on the sphere of radius $R = |u_1 - u_2|/2$ with the center $U = (u_1 + u_2)/2$, and binary collisions have a form $(u_1, u_2) \mapsto (U + \omega R, U - \omega R)$ with the random direction ω.

Other results of Section 7 deal with the random number of colliding particles (Theorem 7.1), locally compact D (Theorem 7.2), and the example of Boltzmann-like equation (see [2, 19]), where the long-term behavior of variance coefficients is investigated (Theorem 7.3). This example shows that the variance coefficient for a Bird-like process may be greater than that for the Nanbu one. Yet the first method is better from the complexity viewpoint.

Finally, linear equations are the theme of Section 8. Here, as we consider the linear functional $\psi(\nu) = \int g d\nu$ with the continuous g, the whole problem becomes linear and all terms of interest are expressed through the fundamental solution $\mu_t(\delta_v)$ of the equation. For example, $f_1(t, \nu; v, u) = \psi(\mu_t(\delta_v)) - \psi(\mu_t(\delta_u))$ and does not dependent on ν here (compare this with the general expression (1.5)). Therefore asymptotical variance coefficients have a simple structure and may be investigated rather thoroughly. It is interesting that artificial collisions may diminish simulational complexity of the algorithm in the same way as the Bird algorithm with binary collisions may have an advantage over the Nanbu one for Boltzmann-like equations.

Technical restrictions of the paper generally result from the peculiarities of the operator semigroup theory, used as the base of all considerations. It seems that the main is the compactness of D, others are certain continuity demands on functions, derivatives etc. Though simulation examples (see Sections 7 and 8) show that these requirements are not necessary, strong mathematical proofs of such statements are beyond the scope of this work.

Several results of the paper were published without proofs in [7] – [9].

2 Differential equations on measure spaces

2.1 Derivatives with respect to measures

Let us start with some general terms and notation. For Banach spaces E_1, E_2 with norms $\|\cdot\|_1$ and $\|\cdot\|_2$ we denote by $\mathbf{L}(E_1, E_2)$ the Banach space of linear operators acting from E_1 to E_2 and by $|\!|\!|\cdot|\!|\!|$ the corresponding norm in $\mathbf{L}(E_1, E_2)$. Let U be an open subset of E_1. The mapping $g' : U \mapsto \mathbf{L}(E_1, E_2)$ is called the *derivative mapping* (or, briefly, the *derivative*) of the mapping

$g : U \mapsto E_2$, if for any $\mu, \widetilde{\mu} \in U$ with $\|\widetilde{\mu} - \mu\|_1 \to 0$

$$\|g(\widetilde{\mu}) - g(\mu) - g'(\mu)(\widetilde{\mu} - \mu)\|_2 = o(\|\widetilde{\mu} - \mu\|_1). \tag{2.1}$$

Remark 2.1 1. *The derivative g' may be considered as a mapping $U \times E_1 \mapsto E_2$, which is linear in the second argument.*
2. *As usual, if the closure of the interior of $V \subset E_1$ includes V, the equality* (2.1) *defines the derivative $g' : V \mapsto \mathbf{L}(E_1, E_2)$.*
3. *The space $\mathbf{L}(\mathbf{R}, E_2)$ may be identified with E_2 by the natural isometry* [3].

Let us consider derivatives of the higher order, using the second derivative as an example. By definition, the second derivative mapping is a mapping $g'' : U \mapsto \mathbf{L}(E_1 \times E_1, E_2)$, which satisfies the relation analogous to (2.1). Therefore we use the notation $g''(\mu)$ or $d^2 g(\mu)/d\mu^2$ for the value of this mapping in the point $\mu \in U$ and may consider the norm of this value in $\mathbf{L}(E_1 \times E_1, E_2)$.

On the other hand one may deal with the second derivative as a mapping $U \times E_1 \times E_1 \mapsto E_2$, which is linear in its last two arguments and is invariant under their permutation. Then we denote the second derivative by $\partial^2 g$. Thus, if $\mu \in U$ and $\nu_1, \nu_2 \in E_1$, then $g''(\mu)(\nu_1, \nu_2)$, $(d^2 g(\mu)/d\mu^2)(\nu_1, \nu_2)$ and $\partial^2 g(\mu; \nu_1, \nu_2)$ signify the same point of E_2.

Now we suppose that (D, ρ) is a complete metric separable compact space, $\mathcal{F}^\rho$ is the topology induced by ρ, $\mathcal{B}$ is the corresponding Borel σ-algebra and E is the linear space of signed finite measures (or, more briefly, measures) on $(D, \mathcal{B})$. It is well known that the space E with the norm $\|\cdot\| = \mathrm{Var}(\cdot)$ is a Banach space. We denote by $\mathcal{F}_s$ the strong topology induced by this norm and use the sign $\stackrel{s}{\to}$ for the corresponding convergence.

Let $\mathcal{F}_w$ be a topology on E induced by the weak convergence of measures, (namely, μ_n converges weakly to μ, $\mu_n \stackrel{w}{\to} \mu$, if $\int f d\mu_n \to \int f d\mu$ for any bounded continuous function $f : D \mapsto \mathbf{R}$). Generally speaking, $(E, \mathcal{F}_s)$ is not separable, but $(E, \mathcal{F}_w)$ is a complete metrizable separable space.

As two topologies on the same space E may produce an uncertainty, we use the following terminology. We consider mappings whose domains and ranges are allowed to be E, its subsets and/or their products. If we equip such domains and ranges with the strong topology $\mathcal{F}_s$, the continuous mapping is called *s-continuous* or *strongly continuous*; in the case we consider the topology $\mathcal{F}_w$ instead of $\mathcal{F}_s$, the mapping is *w-continuous* (*weakly continuous*).

For example, a mapping g is *twice w-continuously differentiable*, if there exists g'' and it is continuous as a mapping $(U, \mathcal{F}_w) \times (E, \mathcal{F}_w) \times (E, \mathcal{F}_w) \mapsto (E, \mathcal{F}_w)$. Note that s-continuous mappings may be not w-continuous and vice versa. As a rule we deal with the weak topology $\mathcal{F}_w$ and mean this topology unless another is mentioned.

Let $V \subset E$ be a weak compact. Suppose that its s-interior is not empty and the s-closure of this interior includes V. Then (see Remark 2.1) one can consider differentiable mappings on V.

Definition 2.1 The function $\psi : V \mapsto \mathbf{R}$ is called *w-continuously differentiable*, if there exists the derivative ψ', which is continuous as a mapping $(V, \mathcal{F}_w) \times (E, \mathcal{F}_w) \mapsto \mathbf{R}$.

The following functional spaces are used throughout the paper:

1. $\mathbf{B}(V)$ is the space of bounded $\sigma(\mathcal{F}_w)$-measurable functions;
2. $\mathbf{C}(V)$ is the Banach space of functions, continuous on $(V, \mathcal{F}_w)$, with the norm $\|\psi\| = \sup\limits_{\mu \in V} |\psi(\mu)|$ (see [3], p. 10 – 11);
3. $\widetilde{\mathbf{C}}^1(V)$ is the set of all w-continuously differentiable functions;
4. $\mathbf{C}^1(V) = \mathbf{C}(V) \cap \widetilde{\mathbf{C}}^1(V)$;
5. $\widetilde{\mathbf{C}}^2(V)$ is the set of all twice w-continuously differentiable functions;
6. $\mathbf{C}^2(V) = \mathbf{C}^1(V) \cap \widetilde{\mathbf{C}}^2(V)$.

The analogous definitions may be done for $\mathbf{C}^k(V)$, $k \geq 3$.

Lemma 2.1 *The set $\mathbf{C}^1(V)$ is dense in $\mathbf{C}(V)$.*

Proof. We need to check the Stone-Weierstrass conditions, that is, to prove that a) $\mathbf{C}^1(V)$ is the algebra, b) for any $\mu_1, \mu_2 \in V$, $\mu_1 \neq \mu_2$, there exists a function $\psi \in \mathbf{C}^1(V)$ such that $\psi(\mu_1) \neq \psi(\mu_2)$.

The condition a) is fulfilled as $(\psi_1\psi_2)' = \psi_1'\psi_2 + \psi_1\psi_2'$. Let us check the second condition. If $\mu_1 \neq \mu_2$, then there exists a bounded continuous function g such that $\int g d\mu_1 \neq \int g d\mu_2$. Denote $\psi(\lambda) = \int g d\lambda$, $\lambda \in V$. Evidently, $\psi \in \mathbf{C}(V)$ and $\psi(\mu_1) \neq \psi(\mu_2)$. Moreover ψ has a derivative $\partial^1\psi(\lambda; \nu) = \int g d\nu$, $\lambda \in V$, $\nu \in E$. Thus the derivative mapping $\partial^1\psi$ is w-continuous. $\square$

The following proposition gives the useful representation for derivative mappings of smooth functions.

Lemma 2.2 *Let $\psi \in \mathbf{C}^1(V)$. Then there exists an $\mathcal{F}_w \times \mathcal{F}^\rho$-continuous function $\psi' : V \times D \mapsto \mathbf{R}$ such that $\partial^1\psi(\mu; \nu) = \int \psi'(\mu, v)\nu(dv)$ for any $\nu \in E$.*

Proof. Let $u \in D$ and put $\psi'(\mu, u) \stackrel{def}{=} \partial^1\psi(\mu; \delta_u) = \int_D \psi'(\mu, v)\delta_u(dv)$. As $\psi \in \mathbf{C}^1(V)$, the function ψ' is continuous in the collection of its arguments. Besides, as D is a compact, the function $\psi'(\mu, \cdot)$ is bounded for any fixed μ. Evidently, $\partial^1\psi(\mu; \nu_n) = \int_D \psi'(\mu, v)\nu_n(dv)$ for any discrete measure $\nu_n = \sum_{i=1}^{n} \alpha_i^{(n)} \delta_{u_i^{(n)}}$. Let $\nu \in E$. As the linear span of the set of Dirac measures is dense in $(E, \mathcal{F}_w)$, there exists a sequence ν_n such that $\nu_n \stackrel{w}{\to} \nu$ and the result is evident. □

Similar definitions and properties are valid for higher derivatives. For example, the analogue of Lemma 2.2 for $\mathbf{C}^2(V)$ may be stated as follows:

Lemma 2.3 *Let $\psi \in \mathbf{C}(V)$. Then there exists an $\mathcal{F}_w \times \mathcal{F}^\rho \times \mathcal{F}^\rho$-continuous function $\psi'' : V \times D \times D \mapsto \mathbf{R}$ such that for any $\nu_1, \nu_2 \in E$*

$$\partial^2\psi(\mu; \nu_1, \nu_2) = \int\limits_{D^2} \psi''(\mu, v_1, v_2)\nu_1(dv_1)\nu_2(dv_2).$$

2.2 Differential equations and associated semigroups

Let $V \subset E$, $G : V \mapsto E$, $s > 0$, $I = [0, s]$, and consider an ordinary differential equation for measures,

$$\frac{d\mu_t}{dt} = G(\mu_t). \tag{2.2}$$

The differential equation (2.2) is said to be *defined on* V. The mapping $\mu_\cdot : I \mapsto V$ is a solution of the equation (2.2) on I, if μ_t is differentiable in $t \in I$ and (2.2) holds. Note that the derivative mapping $d\mu_t/dt$ is considered as an element of E here (see Remark 2.1).

Suppose that for any $\mu \in V$ and $t \geq 0$ there exists the unique solution $\mu_t(\mu)$ of (2.2) with the initial value μ. Let the closure of the interior of V (with respect to the strong topology) include V, then one can deal with the strong and weak properties (such as the continuity and the differentiability) of $G(\mu)$ and $\mu_t(\mu)$ as functions of μ.

From here on we suppose that the solution μ_t of the equation (2.2) satisfies the following conditions (may be, not all of them simultaneously).

Smoothness conditions:

$B0$. for any $\mu \in V$ there exists the unique solution $\mu_t(\mu) \in V$ of the equation (2.2) and $\sup\limits_{\mu \in V, t \geq 0} \|d\mu_t(\mu)/dt\| \leq M < \infty$;

$B1$. the mapping $\mu_t : V \mapsto E$ is s-continuous and differentiable, the derivative mapping $\mu'_t : V \times E \mapsto E$ is also s-continuous;

$B2$. there exist $A > 0$ and B such that $\sup\limits_{\mu \in V} \|d\mu_t(\mu)/d\mu\| \leq A\exp(Bt)$;

$B3$. μ_t and μ'_t are w-continuous;

$B4$. μ_t is twice differentiable and $\mu''_t : V \times E \times E \mapsto E$ is s-continuous;

$B5$. μ''_t is w-continuous.

It follows from $B0 - B3$, that G is w-continuous and $\sup\limits_{\mu \in V} \|G(\mu)\| \leq M$.

Relations between differential equations and the corresponding contraction operator semigroups are well known. Nevertheless we prove here all necessary statements owing to certain peculiarities of the spaces under consideration.

Suppose that conditions $B0 - B3$ are fulfilled for the equation (2.2). For any $t \in [0, +\infty)$ let us define the operator $T^t : \mathbf{B}(V) \mapsto \mathbf{B}(V)$ by $T^t\psi(\mu) = \psi(\mu_t(\mu))$, where $\psi \in \mathbf{B}(V)$, $\mu_t(\mu)$ is the solution of (2.2) and $\mu_0(\mu) = \mu \in V$.

Lemma 2.4 *The operator family T^t, $t \geq 0$, has the following properties:*
1. $T^t(\mathbf{C}(V)) \subset \mathbf{C}(V)$;
2. *T^t, $t \geq 0$, is a strongly continuous semigroup of linear contraction operators on $\mathbf{C}(V)$;*
3. $T^t(\mathbf{C}^1(V)) \subset \mathbf{C}^1(V)$;
4. *if $\psi \in \mathbf{C}^1(V)$, then $\sup\limits_{\mu \in V} \|dT^t\psi(\mu)/d\mu\| \leq A_\psi e^{ct}$, where A_ψ does not depend on t and c does not depend on t and ψ.*

Proof. Statements 1 and 3 follow from $B3$. For statement 2, the strong continuity of T^t is a consequence of the uniform continuity of the function $\phi(t, \mu) = T^t\psi(\mu) - \psi(\mu)$ on the compact set $[0, s] \times V$, $s > 0$.

As both $\{\nu : \|\nu\| \leq 1\}$ and V are weak compacts (and therefore any w-continuous function is bounded on them), we obtain from $B2$ that

$$\|dT^t\psi(\mu)/d\mu\| \leq \|d\psi(\mu_t(\mu))/d\mu_t(\mu)\| \cdot \|d\mu_t(\mu)/d\mu\| \leq A_\psi e^{ct}.$$

Thus the fourth statement is proved. □

Denote by K the strong generating operator (generator) of the semigroup T^t and by D_K the domain of K.

Lemma 2.5 *The following propositions hold:*
1. $\mathbf{C}^1(V) \subset D_K$ *and* $K(\mathbf{C}^1(V)) \subset \mathbf{C}(V)$.
2. *If* $\psi \in \mathbf{C}^1(V)$ *then* $K\psi(\mu) = \partial^1\psi(\mu; G(\mu))$.

Proof. Let $\psi \in \mathbf{C}^1(V)$. If $t \downarrow 0$ then, due to $B1$, we obtain the pointwise convergence

$$\frac{T^t\psi(\mu) - \psi(\mu)}{t} \to \partial^1\psi\Big(\mu_t(\mu); d\mu_t(\mu)/dt\Big)\Big|_{t=0} = \partial^1\psi(\mu; G(\mu)). \tag{2.3}$$

Let us prove that $\partial^1\psi(\mu; G(\mu)) \in \mathbf{C}(V)$. Surely, if $\mu^{(n)} \xrightarrow{w} \mu$ as $n \to \infty$ then

$$\begin{aligned}
&\left|\partial^1\psi\left(\mu^{(n)}; G(\mu^{(n)})\right) - \partial^1\psi\Big(\mu; G(\mu)\Big)\right| \\
&\quad \leq \|\psi'(\mu^{(n)}) - \psi'(\mu)\|\, \|G(\mu^{(n)})\| + \left|\partial^1\psi\left(\mu^{(n)}; G(\mu^{(n)}) - G(\mu)\right)\right| \\
&\quad = I_1 + I_2.
\end{aligned}$$

As $\sup_{\lambda\in V}\|G(\lambda)\| \leq M$ and both V and $\{\nu : \|\nu\| \leq 1\}$ are compact sets in $(E, \mathcal{F}_w)$, then $I_1 \to 0$. The second term tends to zero as G is w-continuous.

Finally, all mappings in (2.3) are w-continuous in μ, V is a compact, and the convergence in (2.3) holds uniformly in μ. □

3 Approximation results

3.1 General conditions

We start with a variant of the Trotter–Kato theorem ([6, Th. 6.1, Ch. 1], [1]). Consider a topological space $\mathbf{X}$ and denote by $\mathbf{C} = \mathbf{C}(\mathbf{X})$ the space of bounded continuous functions on $\mathbf{X}$ with the norm $\|\psi\| = \sup_{x\in\mathbf{X}} |\psi(x)|$. For any $n \geq 1$ let $\mathbf{X}_n$ be a closed subset of $\mathbf{X}$, $\mathbf{C}_n = \mathbf{C}|_{\mathbf{X}_n}$ and $\|\psi\|_n = \sup_{x\in\mathbf{X}_n} |\psi(x)|$. Denote by $\mathfrak{p}_n : \mathbf{C} \mapsto \mathbf{C}_n$ the restriction operator $\mathfrak{p}_n\psi = \psi|_{\mathbf{X}_n}$. For the brevity, we write $\psi_n \underset{n\to\infty}{\longrightarrow} \psi$ instead of $\|\psi_n - \mathfrak{p}_n\psi\|_n \underset{n\to\infty}{\longrightarrow} 0$, if $\psi_n \in \mathbf{C}_n$ and $\psi \in \mathbf{C}$.

Let T^t be a contraction operator semigroup on $\mathbf{C}$ with generator K. For any n consider a contraction operator semigroup T_n^t on $\mathbf{C}_n$ with generator K_n. Suppose that

1. the sequence $\mathbf{X}_n$ *approximates* $\mathbf{X}$ in the sense that $\|\psi\|_n \to \|\psi\|$ for any $\psi \in \mathbf{C}$ and $n \to \infty$;

2. the semigroups T_n^t (and T^t) are strongly continuous on $\mathbf{C}_n$ (and on $\mathbf{C}$).

Trotter–Kato Theorem. *Let there exists $\mathbf{C}' \subset \mathbf{C}$ such that $\mathbf{C}' \subset D_K$ and $\mathbf{C}'_n \stackrel{def}{=} \mathbf{C}'|_{\mathbf{X_n}} \subset D_{K_n}$. If (a) the closure of the restriction K to $\mathbf{C}'$ is identical to K and (b) $K_n\mathfrak{p}_n\psi \to K\psi$ for any $\psi \in \mathbf{C}'$, then $T_n^t\mathfrak{p}_n\psi \to T^t\psi$ for any $\psi \in \mathbf{C}$.*

The following statement (see [1] for the analogous proposition related to the Euclidean space) gives the sufficient conditions for the Trotter–Kato theorem result. Consider

1. $V_n \subset V$ — the sequence of weakly closed sets, approximating V;

2. $\mathbf{X} = (V, \mathcal{F}_w)$, $\mathbf{X}_n = (V_n, \mathcal{F}_w)$;

3. $\mathbf{C}_n = \mathbf{C}(V)|_{V_n}$, $\mathbf{C}_n^1 = \mathbf{C}^1(V)|_{V_n}$.

Theorem 3.1 *Let $T^t(\mathbf{C}^1(V)) \subset \mathbf{C}^1(V)$, $\mathbf{C}^1(V) \subset D_K$, $\mathbf{C}_n^1 \subset D_{K_n}$ and suppose that for any $\psi \in \mathbf{C}^1(V)$ and $t \geq 0$*

$$\sup_{\mu \in V} \| dT^t\psi(\mu)/d\mu \| \leq A_\psi e^{bt}, \tag{3.1}$$

where A and b do not depend on t and b does not depend on ψ. If $K_n\mathfrak{p}_n\psi \to K\psi$ for $\psi \in \mathbf{C}^1(V)$, then the hypotheses of the Trotter-Kato theorem hold.

Proof. As T^t is strongly continuous on $\mathbf{C}(V)$, then $K(D_K) = \mathbf{C}(V)$. Moreover, the resolvent operator $R_b\phi = \int\limits_0^\infty e^{-bt}T^t\phi\, dt$ transfers $\mathbf{C}(V)$ to D_K, is bounded and invertible on these spaces and has the form $R_b = (b\mathbf{I} - K)^{-1}$. Denote $\mathbf{C}' = R_b(\mathbf{C}^1(V))$. As $\mathbf{C}^1(V)$ is dense in $\mathbf{C}(V)$ (see Lemma 2.1), then $\mathbf{C}' \subset D_K$ and $\mathbf{C}'$ is dense in D_K.

Let us check the condition (a): the closure of $\{(\psi, K\psi), \psi \in \mathbf{C}'\}$ coincides with $\{(\psi, K\psi), \psi \in D_K\}$. Suppose $\psi \in D_K$. As $\phi = R_b^{-1}\psi \in \mathbf{C}(V)$, then by Lemma 2.1 there exist $\phi_m \in \mathbf{C}^1(V)$ such that $\|\phi_m - (\mathbf{I} - K)\psi\| \to 0$ as $m \to \infty$. Moreover $\psi_m = R_b\phi_m \in \mathbf{C}'$ and $\psi_m \to R_b\phi = \psi$ by the boundness of the resolvent operator. Finally, as $K\psi_m = b\psi_m - \phi_m$ and $K\psi = b\psi - \phi$, then $K\psi_m \to K\psi$. Thus the assertion (a) is proved.

Let us turn to (b). If $\psi \in \mathbf{C}'$, then $\psi = R_b\phi$ for some $\varphi \in \mathbf{C}^1(V)$. We must prove that $\psi \in \mathbf{C}^1(V)$. Evidently, $\psi \in \mathbf{C}(V)$ and the integral

$$\int\limits_0^\infty e^{-bt} \| dT^t\psi(\mu)/d\mu \| dt$$

converges uniformly in μ by (3.1). Thus the derivative ψ' exists and

$$\partial^1\psi(\mu;\nu) = \int_0^\infty e^{-bt}\partial^1 T^t\psi(\mu;\nu)dt.$$

As $T^t(\mathbf{C}^1(V)) \subset \mathbf{C}^1(V)$, ψ' is w-continuous. □

Now let us turn to the approximation with the help of measure-valued Markov families. For each $n \geq 1$ consider the following objects:

1. $\Xi_n = \{\Omega, \xi_n(t), \mathbf{P}_\mu^{(n)}, \mathrm{F}_t^{(n)}\}$ is a homogeneous Markov family with the phase space V_n. In what follows we omit the upper index n in the notation of distributions $\mathbf{P}_\mu^{(n)}$ and the corresponding mathematical expectations $\mathbf{E}_\mu^{(n)}$;

2. T_n^t, $t \geq 0$, is the operator semigroup defined by $T_n^t\psi(\mu) = \mathbf{E}_\mu\psi(\xi_n(t))$ on the space $\mathbf{B}_n = \mathbf{B}(V_n)$ of bounded and $\sigma(\mathcal{F}_w)$-measurable functions;

3. K_n is the strong generator of the semigroup T_n^t with domain $D_{K_n} \subset \mathbf{B}_n$.

The following approximation conditions are supposed to be fulfilled.

Approximation conditions:

$C1$. V_n approximates V;

$C2$. $T_n^t(\mathbf{C}_n) \subset \mathbf{C}_n$ and T_n^t is strongly continuous on $\mathbf{C}_n$;

$C3$. $\mathbf{C}_n^1 \subset D_{K_n}$.

The resulting statement is an immediate consequence of Theorem 3.1 due to proposition 4 of Lemma 2.4. As earlier, T^t is the semigroup generated by the differential equation (2.2).

Theorem 3.2 *Let the smoothness conditions $B0-B3$ and the approximation conditions $C1-C3$ be fulfilled. If $K_n\mathfrak{p}_n\psi \to \partial^1\psi(\,\cdot\,;G(\cdot))$ for any $\psi \in \mathbf{C}^1(V)$, then $T_n^t\mathfrak{p}_n\psi \to T^t\psi$ for any $\psi \in \mathbf{C}(V)$.*

3.2 Jump Markov family approximation

3.2.1 The convergence conditions

Consider homogeneous Markov families $\Xi_n = \{\Omega, \xi_n(t), \mathbf{P}_\mu, \mathrm{F}_t^{(n)}\}$ with the phase spaces V_n and denote by $T_n^t\psi(\mu) = \mathbf{E}_\mu\psi(\xi_n(t))$, $n \geq 1$, the corresponding sequence of operator semigroups defined on $\mathbf{B}(V_n)$. Suppose that the domains D_{K_n} of the generators K_n coincide with $\mathbf{B}(V_n)$, and

$$K_n\psi(\mu) = \lambda_n(\mu)\left(\int\limits_{V_n} \pi_n(\mu, d\nu)\psi(\nu) - \psi(\mu)\right), \tag{3.2}$$

where $\psi \in \mathbf{B}(V_n)$, $\mu \in V_n$, λ_n is a certain bounded nonnegative measurable function, and π_n is a measurable function in the first argument and a probability measure in the second one. We assume that $q_n(\mu) \stackrel{def}{=} \pi_n(\mu, \{\mu\}) < 1$.

As it is well-known, the Markov family Ξ_n may be considered as a purely jump one. That is, the trajectories of $\xi_n(t)$ are *cadlag* and piecewise constant. Moreover, if we denote by $\widetilde{\tau}_n$ the first jump moment of $\xi_n(\cdot)$, then the $\mathbf{P}_\mu$-distribution of $\widetilde{\tau}_n$ is exponential with the rate $\widetilde{\lambda}_n(\mu) = \lambda_n(\mu)(1 - q_n(\mu))$ and

$$\mathbf{P}_\mu(\xi_n(\widetilde{\tau}_n + 0) \in \,\cdot\,) = \widetilde{\pi}_n(\mu, \,\cdot\,) = \frac{\pi_n(\mu, \cdot) - \delta_\mu(\cdot)q_n(\mu)}{1 - q_n(\mu)}.$$

Denote by $\pi_n : \mathbf{B}(V_n) \mapsto \mathbf{B}(V_n)$ the integral operator with the kernel $\pi_n(\mu, \,\cdot\,)$.

Lemma 3.1 *If $\lambda_n \in \mathbf{C}_n$ and $\pi_n(\mathbf{C}_n) \subset \mathbf{C}_n$, then $T_n^t(\mathbf{C}_n) \subset \mathbf{C}_n$ and the semigroup T_n^t is strongly continuous on $\mathbf{C}_n$.*

Proof. For $\phi \in \mathbf{C}_n$ and $t \in [0, s]$ consider the integral equation

$$u_t(\mu) = e^{-\lambda_n(\mu)t}\phi(\mu) + \int\limits_0^t \lambda_n(\mu)e^{-\lambda_n(\mu)(t-s)} \int\limits_{V_n} \pi_n(\mu, d\nu)u_s(\nu)ds.$$

Evidently, $u_t = T^t\phi$ is the unique solution of this equation in the class of all strongly continuous in t bounded functions. On the other hand, the function $w_0(t, \mu) = e^{-\lambda_n(\mu)t}\phi(\mu)$ is continuous on $W_s = [0, s] \times V_n$ with respect to the product topology, and the operator $\mathbf{L} : \mathbf{B}(W_s) \mapsto \mathbf{B}(W_s)$, defined by

$$\mathbf{L}\psi(t, \mu) = \int\limits_0^t \lambda_n(\mu)e^{-\lambda_n(\mu)(t-s)} \int\limits_{V_n} \pi_n(\mu, d\nu)\psi(s, \nu)ds,$$

acts from $\mathbf{C}(W_s)$ to itself. As the norm of this operator is less than 1, $u_t = T_n^t \phi \in \mathbf{C}_n$ for $\phi \in \mathbf{C}_n$. The strong continuity of the semigroup T_n^t is an immediate consequence of the definition. □

The property $\pi_n(\mathbf{C}_n) \subset \mathbf{C}_n$ is called the *weak continuity* of π_n. This term is applied to other integral operators in the analogous sense.

Corollary 3.1 *If $\lambda_n \in \mathbf{C}_n$ and $\pi_n(\mathbf{C}_n) \subset \mathbf{C}_n$, then the Markov family with the generator* (3.2) *satisfies the approximation conditions $C2$, $C3$.*

Let us introduce some new notation and definitions. First we put

$$\widetilde{p}_n(\mu, \cdot) \stackrel{def}{=} \mathbf{P}_\mu(\xi_n(\widetilde{\tau}_n + 0) - \xi_n(\widetilde{\tau}_n) \in \cdot) = \widetilde{\pi}_n(\mu, \cdot + \mu).$$

Thus $\widetilde{p}_n(\mu, \cdot)$ is the $\mathbf{P}_\mu$-distribution of the first jump value of the process $\xi_n(t)$. Surely, $\widetilde{p}_n(\mu, \mathbf{0}) = 0$. If we denote $p_n(\mu, \cdot) = \pi_n(\mu, \cdot + \mu)$ then

$$p_n(\mu, \cdot) = (1 - q_n(\mu))\widetilde{p}_n(\mu, \cdot) + q_n(\mu)\delta_{\mathbf{0}}(\cdot)$$

and $p_n(\mu, \mathbf{0}) = q_n(\mu)$. This corresponds to 'false jumps' which may occur with the probability $q_n(\mu)$. Therefore we call λ_n and π_n the *infinitesimal characteristics* of the Markov family Ξ_n, and $\widetilde{\lambda}_n$ and $\widetilde{\pi}_n$ its *canonical infinitesimal characteristics*.

The following remark concerns the usage of the mathematical expectation sign $\mathbf{E}$. As usual, $\mathbf{E}_\mu$ denotes the mathematical expectation, corresponding to the Markov family. At the same time the sign $\mathbf{E}$ (without any index) means the mathematical expectation applied to a random variable, which is defined on a certain abstract probability space (different in different situations). The probability signs $\mathbf{P}_\mu$ and $\mathbf{P}$ are used in the analogous sense.

Let $\theta_\mu^{(n)}$ be a random measure with $\mathcal{L}(\theta_\mu^{(n)}) = p_n(\mu, \cdot)$. Then, subject to the last remark, the generator (3.2) may be represented as

$$K_n\psi(\mu) = \lambda_n(\mu)\mathbf{E}\Big(\psi(\mu + \theta_\mu^{(n)}) - \psi(\mu)\Big). \tag{3.3}$$

Now let us turn to the convergence $T_n^t \to T^t$ with T_n^t guided by generators (3.3), and T^t — by the differential equation (2.2). We suppose that the solution of (2.2) satisfies conditions $B0 - B3$ of Section 2.2. Additionally we suppose that the approximation condition $C1$ is fulfilled, and therefore V_n approximates V. For a random measure $\xi \in V_n$ denote by $\mathcal{E}(\xi)$ the measure ν such that $\int f\, d\nu = \mathbf{E} \int f\, d\xi$ for any $f \in \mathbf{C}(D)$.

Theorem 3.3 *Let λ_n and π_n be w-continuous. Suppose that for any $n \geq 1$ there exists a nonnegative function $\rho_n : V_n \mapsto \mathbb{R}$ such that*

$$\sup_{\mu\in V_n} \rho_n(\mu) \underset{n\to\infty}{\longrightarrow} 0, \quad \sup_{\mu\in V_n} \lambda_n(\mu)\rho_n(\mu) \leq C < +\infty \tag{3.4}$$

and

$$\sup_{\mu\in V_n} \Big((\lambda_n(\mu)+1)\mathbf{P}(\|\theta_\mu^{(n)}\| \geq \rho_n) \Big) \underset{n\to\infty}{\longrightarrow} 0. \tag{3.5}$$

If additionally

$$\sup_{\mu\in V_n} \int\limits_D f\, d\left(\lambda_n(\mu)\mathcal{E}(\theta_\mu^{(n)}) - G(\mu) \right) \underset{n\to\infty}{\longrightarrow} 0 \tag{3.6}$$

for any $f \in \mathbf{C}(D)$, then

$$\sup_{\mu\in V_n} |T_n^t \mathfrak{p}_n \psi(\mu) - T^t\psi(\mu)| \underset{n\to\infty}{\longrightarrow} 0, \quad \psi \in \mathbf{C}(V). \tag{3.7}$$

Proof. Due to Corollary 3.1 and Theorem 3.2 it is sufficient to prove that

$$L(\psi, n) \overset{def}{=} \sup_{\mu\in V_n} |K_n\mathfrak{p}_n\psi(\mu) - K\psi(\mu)| \underset{n\to\infty}{\longrightarrow} 0, \qquad \psi \in \mathbf{C}^1(V).$$

Denote $A_n(\mu) = \lambda_n(\mu)\Big(\psi(\mu+\theta_\mu^{(n)}) - \psi(\mu)\Big) - K\psi(\mu)$. Then

$$\begin{aligned} L(\psi, n) = \sup_{\mu\in V_n} &\left| \mathbf{E}\left(A_n(\mu),\ \|\theta_\mu^{(n)}\| \geq \rho_n \right) \right| \\ &+ \sup_{\mu\in V_n} \left| \mathbf{E}\left(A_n(\mu),\ \|\theta_\mu^{(n)}\| < \rho_n \right) \right| = I_1 + I_2. \end{aligned}$$

As ψ and $K\psi$ are bounded, $I_1 \to 0$ by (3.5). On the other hand as $K\psi(\mu) = \partial^1\psi(\mu; G(\mu))$ and

$$\psi(\mu+\nu) - \psi(\mu) = \partial^1\psi(\mu;\nu) + \alpha(\mu,\nu)\|\nu\|$$

with $\sup\limits_{\mu\in V_n,\ \|\nu\|<\varepsilon} |\alpha(\mu,\nu)| \underset{\varepsilon\to 0}{\longrightarrow} 0$, then up to order $o(1)$

$$\begin{aligned} I_2 \leq &\sup_{\mu\in V_n} \left| \mathbf{E}\Big(\lambda_n(\mu)\partial^1\psi(\mu;\theta_\mu^{(n)}) \Big) - \partial^1\psi(\mu; G(\mu)) \right| \\ &+ \sup_{\mu\in V_n} \mathbf{E}\Big(\lambda_n(\mu) \left|\alpha(\mu,\theta_\mu^{(n)})\right| \|\theta_\mu^{(n)}\|,\ \|\theta_\mu^{(n)}\| < \rho_n \Big) = I_3 + I_4. \end{aligned}$$

Finally, I_3 tends to zero due to Lemma 2.2 and (3.6), while I_4 — by (3.4).□

Remark 3.1 *The convergence $\sup\limits_{\mu\in V_n} \|\lambda_n(\mu)\mathcal{E}(\theta_\mu^{(n)}) - G(\mu)\| \underset{n\to\infty}{\longrightarrow} 0$ produces the sufficient condition for (3.6).*

3.2.2 First results on the approximation errors

Suppose that conditions $B0 - B5$ (Section 2.2) are fulfilled for the equation (2.2). Note that $T^t(\mathbf{C}^2(V)) \subset \mathbf{C}^2(V)$ due to $B4$ and $B5$.

Consider the sequence of jump Markov families with the phase spaces V_n, processes $\xi_n(t)$, generators (3.2) and semigroups $T_n^t \mathfrak{p}_n \psi(\mu) = \mathbf{E}_\mu \psi(\xi_n(t))$, and suppose that the approximation condition $C1$ holds for V_n, V. The following proposition makes Theorem 3.3 more precise if $\psi \in \mathbf{C}^2(V)$.

Theorem 3.4 *Let $\psi \in \mathbf{C}^2(V)$. If the condition* (3.6) *is valid and*

$$\beta_n = \sup_{\mu \in V_n} \lambda_n(\mu) \mathbf{E} \|\theta_\mu^{(n)}\|^2 \underset{n\to\infty}{\longrightarrow} 0, \tag{3.8}$$

then uniformly in $\mu \in V_n$

$$\begin{aligned} &T_n^t \mathfrak{p}_n \psi(\mu) - T^t \psi(\mu) \\ &\quad = \int_0^t T_n^{t-\tau} \mathfrak{p}_n \Big(\partial^1 T^\tau \psi \left(\mu; \lambda_n(\mu) \mathcal{E}(\theta_\mu^{(n)}) - G(\mu) \right) \\ &\quad + \frac{1}{2} \lambda_n(\mu) \mathbf{E} \partial^2 T^\tau \psi \left(\mu; \theta_\mu^{(n)}, \theta_\mu^{(n)} \right) \Big) d\tau + o(\beta_n). \end{aligned} \tag{3.9}$$

Proof. Let us apply the Duhamel perturbation formula ([4, p. 82], [6, Lemma 6.2, Ch. 1]): if $\psi \in \mathbf{C}^1(V)$, then

$$T_n^t \mathfrak{p}_n \psi - \mathfrak{p}_n T^t \psi = \int_0^t T_n^{t-\tau} (K_n \mathfrak{p}_n - \mathfrak{p}_n K) T^\tau \psi \, d\tau. \tag{3.10}$$

Now, as $\varphi = T^\tau \psi \in \mathbf{C}^2(V)$, then

$$\begin{aligned} (K_n \mathfrak{p}_n - \mathfrak{p}_n K)\varphi(\mu) &= \lambda_n(\mu) \mathbf{E} \Big(\varphi(\mu + \theta_\mu^{(n)}) - \varphi(\mu) \Big) - \partial^1 \varphi(\mu; G(\mu)) \\ &= \mathbf{E} \partial^1 \varphi \Big(\mu; \lambda_n(\mu) \theta_\mu^{(n)} - G(\mu) \Big) + \frac{\lambda_n(\mu)}{2} \mathbf{E} \partial^2 \varphi(\mu; \theta_\mu^{(n)}, \theta_\mu^{(n)}) + \alpha_n, \end{aligned} \tag{3.11}$$

where the first term on the right-hand side of (3.11) tends to zero uniformly in $\mu \in V_n$ due to (3.6), the second has the form $O(\beta_n)$ and the third is of the order $o(\beta_n)$. Applying (3.10), we obtain (3.9). □

Remark 3.2 *As $\mathbf{C}^2(V)$ is dense in $\mathbf{C}(V)$, Theorems* 3.3 *and* 3.4 *give the sufficient conditions for the semigroup convergence on $\mathbf{C}(V)$. It may be seen that condition $B2$ has not been applied when demonstrating Theorem* 3.4, *while conditions $B4$ and $B5$ are not used in Theorem* 3.3.

The following assertion is the useful refinement of Theorem 3.4. Suppose that $\psi \in \mathbf{C}^2(V)$ and conditions $B0 - B5$ are fulfilled.

Theorem 3.5 *Let $V_n = V$, $\lambda_n(\mu) = n\lambda(\mu) > 0$, $n\mathcal{E}(\theta_\mu^{(n)}) = G(\mu)/\lambda(\mu)$ for sufficiently large n, and $n \sup\limits_{\mu \in V} \|\theta_\mu^{(n)}\| < \infty$ $\mathbf{P}$-a.s. Then*

$$\limsup_n n\mathbf{E}_\mu \Big(\psi(\xi_n(t)) - T^t\psi(\mu)\Big)^2 \geq \int\limits_0^t \frac{1}{\lambda(\mu_\tau)} \Big(\partial^1 T^{t-\tau}\psi(\mu_\tau; G(\mu_\tau))\Big)^2 d\tau \quad (3.12)$$

with $\mu_\tau = \mu_\tau(\mu)$. Moreover, if $n\theta_\mu^{(n)} = G(\mu)/\lambda(\mu)$ $\mathbf{P}$-a.s. for sufficiently large n, then $n\mathbf{E}_\mu\Big(\psi(\xi_n(t)) - T^t\psi(\mu)\Big)^2$ tends to the right-hand side of (3.12) *as $n \to \infty$, and this convergence is uniform in $\mu \in V$.*

Proof. First, (3.6) and (3.8) are valid here and $\beta_n = O(1/n)$. Therefore (see Remark 3.2) we get from (3.9) that up to order $o(1/n)$

$$T_n^t\psi(\mu) - T^t\psi(\mu) = \frac{1}{2n} \int\limits_0^t T^{t-\tau}\Big(\lambda(\mu)\mathbf{E}\partial^2 T^\tau\psi\left(\mu; n\theta_\mu^{(n)}, n\theta_\mu^{(n)}\right)\Big) d\tau. \quad (3.13)$$

It follows from (3.13) that $m_0(\psi) \stackrel{def}{=} \mathbf{E}_\mu\psi(\xi_n(t)) - \psi(\mu_t(\mu)) = O(1/n)$ uniformly in μ. Surely, $\mathbf{E}_\mu\Big(\psi(\xi_n(t)) - \psi(\mu_t(\mu))\Big)^2 = m_0(\psi^2) - 2\psi(\mu_t(\mu))\, m_0(\psi)$.

It may be easily seen that for any $F \in \mathbf{C}^2(V)$, $\mu \in V$ and $\nu_1, \nu_2 \in E$

$$\partial^2 F^2(\mu; \nu_1, \nu_2) = 2\partial^1 F(\mu; \nu_1)\, \partial^1 F(\mu; \nu_2) + 2F(\mu)\, \partial^2 F(\mu; \nu_1, \nu_2).$$

Thus as $T^\tau\psi \in \mathbf{C}^2(V)$ and $T^{t-\tau} \circ T^\tau = T^t$, we obtain that

$$\begin{aligned} &\mathbf{E}_\mu\Big(\psi(\xi_n(t)) - \psi(\mu_t(\mu))\Big)^2 \\ &\qquad = \frac{1}{n} \int\limits_0^t T^{t-\tau}\Big(\lambda(\mu)\mathbf{E}\Big(\partial^1 T^\tau\psi\left(\mu; n\theta_\mu^{(n)}\right)\Big)^2\Big)\, d\tau + o(1/n). \end{aligned}$$

Due to the integral form of $\partial^1 T^\tau\psi$ (see Lemma 2.2),

$$\mathbf{E}\Big(\partial^1 T^\tau\psi(\mu; n\theta_\mu^{(n)})\Big)^2 \geq \Big(\partial^1 T^\tau\psi(\mu; n\mathcal{E}(\theta_\mu^{(n)}))\Big)^2 = \frac{\Big(\partial^1 T^\tau\psi(\mu; G(\mu))\Big)^2}{\lambda^2(\mu)}$$

and the proof is complete. $\square$

Theorem 3.5 tells us that for a positive continuous λ optimal jumps are not random: if μ is the position of the optimal jump process just *before* a jump, its position *after* the jump is equal to $\mu + n^{-1}G(\mu)/\lambda(\mu)$.

Now let us turn to the example, elucidating the conditions and results of Theorems 3.4 and 3.5. Let $D = \{1, \ldots, d\}$ and consider the discrete topology on D. Surely, now E consists of all discrete (signed) measures with d atoms and each $\mu \in E$ can be identified with a vector $\mathbf{z} = (z_1, \ldots, z_d) \in \mathbb{R}^d$ by $\mu(\{i\}) = z_i$. Moreover both strong and weak topologies coincide here and are equivalent to the usual topology on $\mathbb{R}^d$, derivatives with respect to measures correspond to the usual derivatives on $\mathbb{R}^d$ etc.

As for the o.d.e. (2.2), it is now the system of d ordinary differential equations

$$\frac{d\mathbf{z}}{dt} = G(\mathbf{z}), \quad \mathbf{z} = \mathbf{z}(t) \in V \subset \mathbb{R}^d, \quad G : V \mapsto \mathbb{R}^d,$$

conditions $B0 - B5$ may be rewritten in the terms of G, and one can produce numerous examples when these conditions hold.

Let us take $\lambda(\mu) = \lambda(\mathbf{z}) \equiv 1$ for the simplicity. Then the *optimal* jump process of Theorem 3.5 corresponds to the deterministic jumps $\mathbf{z} \mapsto \mathbf{z} + G(\mathbf{z})/n$ and there is no difficulties in the simulation of such processes [16].

Quite different situation may occur in the general case. For the optimal simulation (assuming $\lambda(\mu) \equiv 1$) we must produce the deterministic jump $\mu \mapsto \mu + G(\mu)/n$, where $\mu \in V$ and $G(\mu) \in E$. But the arbitrary measures are not constructive objects from the simulation viewpoint, and even if μ is a discrete measure, $G(\mu)$ may be non-discrete. Therefore the procedure described in Theorem 3.5 is optimal but (generally speaking) is not constructive for the equations for measures. The assumptions of Theorem 3.4 are not so restrictive and therefore this theorem is the base of further considerations.

4 Balance equations

In the previous sections general approximation problems of the stochastic solution for o.d.e. defined on $V \subset E$ were considered, and jump Markov families with the phase spaces $V_n \subset V$ were used as approximating families. Now let us suppose that the equation (2.2) is defined on H, where $H \subset E$ stands for the set of all probability measures on $(D, \mathcal{B})$. Thus we assume that G has the domain H and for each $\mu \in H$ the solution $\mu_t(\mu)$ also belongs to H. Then it is natural to approximate such a kind of differential equation by Markov families with the phase spaces $H_n \subset H$.

Surely, we can't apply immediately all previous results here as H has the empty s-interior. On the other hand H is a weak compact and a convex subset of E. Therefore, slightly modifying the derivative definition, we automatically get assertions, analogous to ones demonstrated in Theorem 3.4.

Let $Z = \{\nu \in E : \nu(D) = 0\}$. The mapping $g'_H : H \mapsto \mathbf{L}(Z, E)$ is called an *H-derivative* of $g : H \mapsto E$, if

$$\|g(\widetilde{\mu}) - g(\mu) - g'_H(\mu)(\widetilde{\mu} - \mu)\| = o(\|\widetilde{\mu} - \mu\|), \quad \mu, \widetilde{\mu} \in H, \ \|\widetilde{\mu} - \mu\| \to 0. \tag{4.1}$$

The higher H-derivatives may be defined in the same way as well as functional spaces $\mathbf{C}^k(H)$ (see Section 2.1).

It may be easily seen that the definition (4.1) is correct and H-derivatives have the same properties as the usual ones defined by (2.1). To demonstrate such assertions one must extend g to a mapping $\widetilde{g}$ defined on H^C, where H^C consists of positive measures with $c_1 \le \|\mu\| \le c_2$ and $0 < c_1 < 1 < c_2 < \infty$. This extension may be defined by $\widetilde{g}(\nu) = \|\nu\| g(\nu/\|\nu\|)$. It is easy to prove that $\widetilde{g} \in \mathbf{C}^k(H^C)$ if $g \in \mathbf{C}^k(H)$ and the equalities of the kind $\partial^k \widetilde{g}\,|_H = \partial^k_H g$ hold. Thus all necessary properties of g may be reformulated in terms of $\widetilde{g}$ and vice versa.

Now let us turn to o.d.e.'s for measures. Let $G : H \mapsto Z$, i.e. $G(\mu)(D) = 0$ for any $\mu \in H$. Consider the *balance differential equation*

$$\frac{d\mu_t}{dt} = G(\mu_t) \tag{4.2}$$

and discuss the conditions we put on its right-hand side.

Conditions on the balance equation:

*A*0. there exists $\varepsilon_0 > 0$ such that for any $\mu \in H$ the measure $\mu + \varepsilon_0 G(\mu)$ is positive;

*A*1. the mapping $G : H \mapsto Z$ is s-continuous, H-differentiable and its H-derivative G'_H is s-continuous;

*A*2. there exists $M_1 > 0$ such that $\sup \|G'_H(\mu)(\nu)\| \le M_1$, where the supremum sign acts over all pairs (ν, μ) with $\nu \in Z$, $\|\nu\| \le 1$ and $\mu \in H$;

*A*3. the mappings G and G'_H are w-continuous;

*A*4. the mapping G is twice H-differentiable, the derivative mapping G''_H is s-continuous and there exists $M_2 > 0$ such that $\sup \|G''_H(\mu)(\nu_1, \nu_2)\| \le M_2$, where the supremum sign acts over all collections (ν_1, ν_2, μ) with $\nu_1, \nu_2 \in Z$, $\max(\|\nu_1\|, \|\nu_2\|) \le 1$ and $\mu \in H$;

$A5$. the mapping G''_H is w-continuous.

Lemma 4.1 *Under conditions $A0-A3$, there exists the unique solution $\mu_t(\mu)$ of the o.d.e.* (4.2) *with the initial data $\mu \in H$. This solution satisfies conditions $B0 - B3$, replacing V by H and derivatives by H-derivatives.*

Proof. The existence result follows from the standard polygon line Euler method. Note that condition $A0$ is used here to demonstrate that all polygon lines (and their limit) belong to H for any $t \in [0, +\infty)$. By conditions $A0$ and $A1$ we see that $\|G(\mu)\| \leq M < \infty$ for any $\mu \in H$. Therefore $\sup\limits_{\mu \in H, t \geq 0} \|d\mu_t(\mu)/dt\| \leq M < \infty$ and $B0$ is valid.

Let $B(\mu) = \varepsilon_0 G(\mu) + \mu$, then $B(\mu)$ is positive due to $A0$ and therefore $B(\mu) \in H$. Thus the equation (4.2) gets the form

$$\frac{d\mu_t}{dt} = cB(\mu_t) - c\mu_t, \quad c = 1/\varepsilon_0. \tag{4.3}$$

From this point on we put $c = 1$ as it has a sense of the time unit: if $\mu_t^{(c)}$ denotes a solution of (4.3) with a certain $c > 0$, then $\mu_t^{(c)} = \mu_{ct}^{(1)}$. Note that the mapping B satisfies conditions $A1 - A3$.

Now let us turn to $B1 - B3$. Note that if a sequence of functions $d_t^{(n)}$ satisfies inequalities

$$d_t^{(n)} \leq \gamma_1 e^{-t} + \gamma_2 \int\limits_0^t e^{-(t-s)} d_s^{(n-1)} ds, \qquad d_t^{(0)} = \gamma_1 e^{-t}$$

with $\gamma_1, \gamma_2 > 0$, then

$$d_t^{(n)} \leq \gamma_1 e^{-t} \sum_{k=0}^{n} \frac{(\gamma_2 t)^k}{k!} \leq \gamma_1 e^{(\gamma_2 - 1)t}.$$

Consider the following equation

$$\mu_t = e^{-t}\mu + \int\limits_0^t e^{-(t-s)} B(\mu_s)\, ds, \tag{4.4}$$

where the integral sign stands for the Riemann integral. By $A1$ and $A2$ one may check in a standard way that μ_t is the strong limit of $\mu_t^{(n)}$, where

$$\mu_t^{(n)} = e^{-t}\mu + \int\limits_0^t e^{-(t-s)} B\left(\mu_s^{(n-1)}\right) ds, \quad \mu_t^{(0)} = e^{-t}\mu, \tag{4.5}$$

and B is extended to a mapping defined on the set E^+ of positive measures by

$$B(\nu) = \|\nu\| B(\nu/\|\nu\|). \tag{4.6}$$

Let us fix $t > 0$ and put $V_t = \{\nu \in E^+, e^{-t} \le \|\nu\| \le 1\}$. By (4.6),

$$\sup_{\nu \in V_t} \| dB(\nu)/d\nu \| = \gamma < +\infty,$$

where $\| \cdot \|$ stands for the norm in the Banach space of linear mappings.

It may be easily checked that $B1$ and $B3$ hold for $\mu_t^{(n)}$. Thus to prove these properties for μ_t we must demonstrate the boundness of $\|\mu_t^{(n)}(\mu)\|$ and $\|d\mu_t^{(n)}(\mu)/d\mu\|$ in n and μ for fixed t.

Let $a_t^{(n)} = \|\mu_t^{(n)}\|$. Then $a_t^{(0)} = e^{-t}$ and, as $\|B(\nu)\| = \|\nu\|$ for $\nu \in E^+$,

$$a_t^{(n)} \le e^{-t} + \int_0^t e^{-(t-s)} a_s^{(n-1)} ds.$$

Thus $a_t^{(n)} \le 1$, and μ_t is weakly continuous in the initial data.

Let us turn to the derivative $\partial_H^1 \mu_t^{(n)}(\mu;\cdot) = d_H \mu_t^{(n)}(\mu)/d\mu$. Using $A2$ we obtain that both sides of the identity (4.5) may be differentiated with respect to the initial data μ, and

$$\partial_H^1 \mu_t^{(n)}(\mu;\nu) = e^{-t}\nu + \int_0^t e^{-(t-s)}\, \partial^1 B\left(\mu_s^{(n-1)}(\mu); \partial_H^1 \mu_s^{(n-1)}(\mu;\nu)\right) ds$$

for $\nu \in Z$. Denoting $b_t^{(n)} = \|d_H \mu_t^{(n)}(\mu)/d\mu\|$, we get that

$$b_t^{(n)} \le e^{-t} + \gamma \int_0^t e^{-(t-s)} b_s^{(n-1)} ds, \qquad b_t^{(0)} = e^{-t}.$$

Thus $b_t^{(n)} \le e^{(\gamma-1)t}$, $B1 - B3$ hold and the proof is complete. □

Now we turn to the properties of the second derivative $d_H^2 \mu_t(\mu)/d\mu^2$.

Lemma 4.2 *Let us suppose that G satisfies conditions $A0 - A5$. Then for any $t \ge 0$ the solution $\mu_t(\mu)$ of (4.2) with the initial data $\mu \in H$ is twice H-differentiable in μ and satisfies conditions $B4$, $B5$.*

Proof. Generally the proof is the same as that of Lemma 4.1. It is sufficient to demonstrate that $c_t^{(n)} \stackrel{def}{=} \sup_{\mu\in H} \|d_H^2\mu_t^{(n)}(\mu)/d\mu^2\|$ is bounded in n for fixed t. As before, we have from (4.6) that $\sup_{\nu\in V_t}\|d^2B(\nu)/d\nu^2\| = \gamma_2 < +\infty$, and (4.5) with $A4$ give us the identity

$$\begin{aligned}
&\partial_H^2\mu_t^{(n)}(\mu;\nu_1,\nu_2)\\
&= \int_0^t e^{-(t-s)}\Big(\partial^2 B\left(\mu_s^{(n-1)}(\mu);\partial_H^1\mu_s^{(n-1)}(\mu;\nu_1),\partial_H^1\mu_s^{(n-1)}(\mu;\nu_2)\right)\\
&+\partial^1 B\left(\mu_s^{(n-1)}(\mu);\partial_H^2\mu_s^{(n-1)}(\mu;\nu_1,\nu_2)\right)\Big)ds, \qquad \nu_1,\nu_2\in Z.
\end{aligned}$$

Let $\gamma_1 > \max(1,\gamma)$ and denote $b_t^{(n)} = \|d_H\mu_t^{(n)}(\mu)/d\mu\|$. As $b_t^{(n)} \le e^{(\gamma_1-1)t}$, then

$$\begin{aligned}
c_t^{(n)} &\le \int_0^t e^{-(t-s)}\left(\gamma_2(b_s^{(n-1)})^2 + \gamma_1 c_s^{(n-1)}\right)ds\\
&\le \gamma_2\int_0^t e^{-(t-s)}e^{2(\gamma_1-1)s}ds + \gamma_1\int_0^t e^{-(t-s)}c_s^{(n-1)}ds\\
&\le \gamma_2 e^{-t}(e^{(2\gamma_1-1)t}-1)/(2\gamma_1-1) + \gamma_1\int_0^t e^{-(t-s)}c_s^{(n-1)}ds\\
&\le Ae^{Ct} + \gamma_1\int_0^t e^{-(t-s)}c_s^{(n-1)}ds,
\end{aligned}$$

where $C = 2\gamma_1 - 2 > 0$ and $A = \gamma_2/C$. As $c_t^{(0)} = 0$ and $c_t^{(1)} \le Ae^{Ct}$, then

$$c_t^{(n)} \le Ae^{Ct}\sum_{k=1}^n \frac{(\gamma_1 t)^k}{C^k k!} \le A\exp(Ct + \gamma_1 t/C).$$

The proof is complete. □

Let now $H_n \subset H$ be the closed sets, approximating H, and $\mathfrak{p}_n$ stand for the corresponding restriction operators. As usual, we denote by T^t the semigroup of the balance equation (4.2), and consider the jump Markov families Ξ_n with w-continuous infinitesimal characteristics λ_n and π_n, phase spaces H_n and semigroups T_n^t. As above denote by $\theta_\mu^{(n)}$ the jump random measures (for the exact definition see (3.3)).

All previous approximation results may be automatically reformulated for the case of balance equations defined on H. Here we give the assertion, analogous to Theorem 3.4, as it is the base of all the further considerations. For the brevity, we omit H in the name and notation of H-derivatives.

Theorem 4.1 *Let conditions A0 – A5 be valid for the balance equation* (4.2)*. Consider the jump Markov families with w-continuous infinitesimal characteristics λ_n, π_n and phase spaces H_n. If*

$$\sup_{\mu\in H_n} \int\limits_D f\, d\Big(\lambda_n(\mu)\mathcal{E}(\theta_\mu^{(n)}) - G(\mu)\Big) \underset{n\to\infty}{\longrightarrow} 0 \tag{4.7}$$

for any $f \in \mathbf{C}(D)$ and

$$\beta_n = \sup_{\mu\in H_n} \lambda_n(\mu)\mathbf{E}\|\theta_\mu^{(n)}\|^2 \underset{n\to\infty}{\longrightarrow} 0, \tag{4.8}$$

then

$$\sup_{\mu\in H_n} |T_n^t \mathfrak{p}_n \psi(\mu) - T^t\psi(\mu)| \underset{n\to\infty}{\longrightarrow} 0 \tag{4.9}$$

for $\psi \in \mathbf{C}(H)$; if $\psi \in \mathbf{C}^2(H)$ then (3.9) *holds uniformly in $\mu \in H_n$.*

5 Empirical jump Markov families

Let $H_n = \{n^{-1}\sum_{i=1}^n \delta_{v_i},\ v_i \in D\}$. Measure-valued processes with the phase space H_n are said to be *empirical*. Note that H_n is w-closed and the sequence H_n approximates H.

Let the mapping $Q : D^n \mapsto H_n$ be defined by

$$Q\mathbf{v} = n^{-1}\sum_{i=1}^n \delta_{v_i}, \qquad \mathbf{v} = (v_1, \ldots, v_n) \in D^n.$$

This mapping is a continuous one if we consider the usual product topology on D^n and the weak topology $\mathcal{F}_w$ on H_n. If $\mathfrak{M}$ is a certain measure on D^n, then it goes under Q to the measure $\mathfrak{m} = \mathfrak{M}Q^{-1}$, defined on $(H_n, \sigma(\mathcal{F}_w))$ by $\mathfrak{m}(A) = \mathfrak{M}(Q^{-1}(A))$, $A \in \sigma(\mathcal{F}_w)$.

The mapping Q produces a bijection Q between the set of functions $f : D^n \mapsto \mathbb{R}$, invariant under permutations of their arguments, and the set of functions $F : H_n \mapsto \mathbb{R}$. This bijection may be defined by the equality

$f(v_1, \dots, v_n) = F(n^{-1} \sum_{i=1}^{n} \delta_{v_i})$, or $F = \mathrm{Q}f$. Evidently, Q is a gomeomorphism between the set of all invariant continuous functions $D^n \mapsto \mathbb{R}$ and the set of all w-continuous functions $H_n \mapsto \mathbb{R}$.

For $f : D^n \mapsto \mathbb{R}$ denote by f_s its symmetrization by the averaging upon all permutations of its arguments. Thus we obtain the mapping $f \mapsto \mathrm{Q}_s f$,

$$\mathrm{Q}_s f(\nu) \stackrel{def}{=} \mathrm{Q} f_s(\nu) = \frac{1}{\mathrm{card}(Q^{-1}\nu)} \sum_{\mathbf{y} \in Q^{-1}\nu} f(\mathbf{y}), \quad \nu \in H_n.$$

Surely, if $f \in \mathbf{C}(D^n)$ then $\mathrm{Q}_s f$ is w-continuous.

Now let us turn to the empirical Markov jump families on H_n. Note that if $\mathbf{v}_n(t) = (v_1^{(n)}(t), \dots, v_n^{(n)}(t)) \in D^n$ is a random process, then the measure-valued random process $\xi_n(t) = Q\mathbf{v}_n(t) = n^{-1} \sum_{i=1}^{n} \delta_{v_i^{(n)}(t)}$ is the empirical one.

Consider a jump Markov family $\{\Omega, \mathbf{v}_n(t), \mathbf{P}_{\mathbf{y}}, F_t = \sigma(\mathbf{v}_n(s), s \le t)\}$ with the phase space $(D^n, \mathcal{B}^n)$, *cadlag* piecewise constant trajectories $\mathbf{v}_n(t)$, and infinitesimal characteristics $\Lambda_n(\mathbf{y})$ and $\Pi_n(\mathbf{y}, d\mathbf{x})$. Let $\xi_n(t) \stackrel{def}{=} Q\mathbf{v}_n(t)$ and

$$\mathbf{P}_\nu = \frac{1}{\mathrm{card}(Q^{-1}\nu)} \sum_{\mathbf{y} \in Q^{-1}\nu} \mathbf{P}_{\mathbf{y}}, \qquad \nu \in H_n. \tag{5.1}$$

Lemma 5.1 *The collection $\{\Omega, \xi_n(t), \mathbf{P}_\nu, \mathfrak{F}_t = \sigma(\xi_n(s), s \le t)\}$ determines the jump Markov family with the phase space $(H_n, \sigma(\mathcal{F}_w))$, cadlag piecewise constant trajectories $\xi_n(t)$, and infinitesimal characteristics $\lambda_n = \mathrm{Q}_s\Lambda_n$ and*

$$\pi_n(\nu, \cdot\,) = \frac{1}{\mathrm{card}(Q^{-1}\nu)} \sum_{\mathbf{y} \in Q^{-1}\nu} \Pi_n Q^{-1}(\mathbf{y}, \cdot\,). \tag{5.2}$$

Proof. The trajectory properties are the direct consequence of the definition. Let us prove the Markov property. First of all we have that $\mathbf{P}_{\mathbf{y}}$–a. s. $\mathbf{E}_{\mathbf{y}}(\phi(\mathbf{v}_n(t)) \,|\, \mathfrak{F}_t) = \phi_s(\mathbf{v}_n(t))$ for any bounded measurable ϕ (this identity follows from the fact that $\mathbf{E}_{\mathbf{y}}(\phi(\widetilde{\mathbf{v}}_n(t)) \,|\, \mathfrak{F}_t) = \mathbf{E}_{\mathbf{y}}(\phi(\mathbf{v}_n(t)) \,|\, \mathfrak{F}_t)$, where $\widetilde{\mathbf{v}}_n(t)$ stands for an arbitrary permutation of the coordinates of $\mathbf{v}_n(t)$).

Then, for any bounded measurable function $F : H_n \mapsto \mathbb{R}$,

$$\mathbf{E}_\nu(F(\xi_n(t+s)) \,|\, \mathfrak{F}_t) = \frac{1}{\mathrm{card}(Q^{-1}\nu)} \sum_{\mathbf{y} \in Q^{-1}\nu} \mathbf{E}_{\mathbf{y}}(\Phi(\mathbf{v}_n(t+s)) \,|\, \mathfrak{F}_t)$$

with $\Phi = F \circ Q$. As $F_t \supset \mathfrak{F}_t$, then $\mathbf{P}_{\mathbf{y}}$-a. s.

$$\begin{aligned}
\mathbf{E}_{\mathbf{y}}(\Phi(\mathbf{v}_n(t+s)) \,|\, \mathfrak{F}_t) &= \mathbf{E}_{\mathbf{y}}(\mathbf{E}_{\mathbf{y}}(\Phi(\mathbf{v}_n(t+s)) \,|\, F_t) \,|\, \mathfrak{F}_t) \\
&= \mathbf{E}_{\mathbf{y}}(\mathbf{E}_{\mathbf{v}_n(t)}\Phi(\mathbf{v}_n(s)) \,|\, \mathfrak{F}_t) = \mathbf{E}_{\xi_n(t)}\Phi(\mathbf{v}_n(s)) = \mathbf{E}_{\xi_n(t)}F(\xi_n(s)).
\end{aligned}$$

Therefore $\mathbf{E}_\nu(F(\xi_n(t+s))\,|\,\mathfrak{F}_t) = \mathbf{E}_{\xi_n(t)}F(\xi_n(s))$ $\mathbf{P}_\nu$-a. s.

Now let us consider the infinitesimal operator of this family. For an arbitrary bounded measurable function $F: H_n \mapsto \mathbb{R}$ and $\Phi = F \circ Q$

$$\frac{\mathbf{E}_\nu F(\xi_n(t)) - F(\nu)}{t} = \frac{1}{\mathrm{card}(Q^{-1}\nu)} \sum_{\mathbf{y}\in Q^{-1}\nu} \frac{\mathbf{E}_\mathbf{y}\Phi(v_n(t)) - \Phi(\mathbf{y})}{t}$$
$$\underset{t\downarrow 0}{\longrightarrow} \frac{1}{\mathrm{card}(Q^{-1}\nu)} \sum_{\mathbf{y}\in Q^{-1}\nu} \Lambda_n(\mathbf{y}) \int\limits_{D^n} \Pi_n(\mathbf{y}, d\mathbf{z})(\Phi(\mathbf{z}) - \Phi(\mathbf{y}))$$
$$= \frac{1}{\mathrm{card}(Q^{-1}\nu)} \sum_{\mathbf{y}\in Q^{-1}\nu} \Lambda_n(\mathbf{y}) \int\limits_{D^n} \Pi_n(\mathbf{y}, d\mathbf{z})(F(Q\mathbf{z}) - F(\nu))$$
$$= \lambda_n(\nu) \int\limits_{H_n} \pi_n(\nu, d\widetilde{\nu})(F(\widetilde{\nu}) - F(\nu)),$$

and the convergence is uniform in $\mathbf{y}$ and therefore in ν. $\square$

Remark 5.1 *The transformation $\mathbf{v}_n(t) \mapsto \xi_n(t) = Q\mathbf{v}_n(t)$ may modify the jump frequency of the process $\mathbf{v}_n(t)$ due to the symmetrization of its coordinates. Thus canonical infinitesimal characteristics of both processes may be not compatible; in particular,* (5.2) *may be wrong for them.*

Surely, if the Markov family of the process $\mathbf{v}_n(t)$ is amplified by the initial distribution $\mathfrak{M}$, then $\mathfrak{M}$ corresponds to the initial distribution $\mathfrak{m} = \mathfrak{M}Q^{-1}$ for the Markov family of the process $\xi_n(t)$. Therefore to deal with the process $\xi_n(t)$, it is sufficient to consider the process $\mathbf{v}_n(t)$.

Now our aim is to produce a class of jump Markov families in D^n such that the conditions (4.7), (4.8) and the approximation result (3.9) are valid for them. Let $k \geq 1$. For any collection $J_k = (i_1, \ldots, i_k)$, $i_j = 1, \ldots, n$, such that $i_j \neq i_l$ for $j \neq l$, and for any $\mathbf{v} \in D^n$ consider the probabilities $p_{i_1\ldots i_k}(\mathbf{v})$ with $\sum_{J_k} p_{i_1\ldots i_k}(\mathbf{v}) = 1$, where $\sum_{J_k}$ stands for the summation over all multiindices J_k. In the same way consider the probability measures $T_{i_1\ldots i_k}(\,\cdot\,;\mathbf{v})$ defined on $\mathcal{B}^k$, which may be called *transformation functions.*

For fixed k and n the random process $\mathbf{v}_n(t) \in D^n$ (as well as the corresponding process $\xi_n(t) = Q\mathbf{v}_n(t) \in H_n$) is said to be (n,k)*-particle process,* if its infinitesimal characteristic $\Pi_n = \Pi_n^{(k)}$ has a form

$$\Pi_n^{(k)}(\mathbf{v}, A) = \int\limits_A \sum_{J_k} p_{i_1\ldots i_k}(\mathbf{v}) T_{i_1\ldots i_k}(dv'_{i_1}\ldots dv'_{i_k};\mathbf{v}) \bigotimes_{j\notin\{i_1,\ldots,i_k\}} \delta_{v_j}(dv'_j), \qquad (5.3)$$

$A \in \mathcal{B}^n$. The algorithmic description of such a process may be given in the following manner. If at the time moment preceding to the jump the phase coordinate of the process is $\mathbf{v} \in D^n$, then we choose (with the probability $p_{i_1\ldots i_k}(\mathbf{v})$) a collection of k "particles" from $\mathbf{v}$ with the numbers $i_1, \ldots, i_k$. Then these particles get their new positions according to the transformation function $T_{i_1\ldots i_k}(\,\cdot\,;\mathbf{v})$. All other particles stay still.

Impose the following restrictions on $p_{i_1\ldots i_k}$ and $T_{i_1\ldots i_k}$.

Invariance conditions:

$U1$. there exist functions $p^{(k)} : D^k \times D^n \mapsto \mathbf{R}$ and $T^{(k)} : \mathcal{B}^k \times D^k \times D^n \mapsto \mathbf{R}$, invariant under arbitrary permutation of the last n arguments, such that for any $J_k = (i_1, \ldots, i_k)$ and $\mathbf{v} = (v_1, \ldots, v_n)$

$$p_{i_1\ldots i_k}(\mathbf{v}) = p^{(k)}(v_{i_1}, \ldots, v_{i_k}; \mathbf{v})$$

and for any $A^{(k)} \in \mathcal{B}^k$

$$T_{i_1\ldots i_k}(A^{(k)}; \mathbf{v}) = T^{(k)}(A^{(k)}; v_{i_1}, \ldots, v_{i_k}; \mathbf{v}); \tag{5.4}$$

$U2$. $\Lambda_n^{(k)}$ is invariant under an arbitrary permutation of its arguments.

Lemma 5.2 *If $A \in \mathcal{B}^n$ is symmetrical and (5.4) holds, then $\Pi_n^{(k)}(\mathbf{v}, A)$ is invariant under any permutation of coordinates of its first argument.*

Proof. Let $\mathbf{v} = (v_1, \ldots, v_n)$. Denote by π a permutation of the coordinates of $\mathbf{v}$ and put $\widetilde{\mathbf{v}} = \pi\mathbf{v} = (\widetilde{v}_1, \ldots, \widetilde{v}_n)$. For fixed $i_1, \ldots, i_k$ denote by $j_1, \ldots, j_k$ such numbers that $\pi(v_{i_1}, \ldots, v_{i_k}) = (\widetilde{v}_{j_1}, \ldots, \widetilde{v}_{j_k})$. Then

$$p_{i_1\ldots i_k}(\mathbf{v}) = p^{(k)}(v_{i_1}, \ldots, v_{i_k}; \mathbf{v}) = p^{(k)}(\widetilde{v}_{j_1}, \ldots, \widetilde{v}_{j_k}; \widetilde{\mathbf{v}}) = p_{j_1\ldots j_k}(\widetilde{\mathbf{v}}), \tag{5.5}$$

and the analogous identity holds for the transformation function $T_{i_1\ldots i_k}$. Therefore

$$\begin{aligned} T_{i_1\ldots i_k}(dv'_{i_1}, \ldots, dv'_{i_k}; \mathbf{v}) \bigotimes_{m \notin \{i_1,\ldots,i_k\}} \delta_{v_m}(dv'_m) = \\ T_{j_1\ldots j_k}(dv'_{i_1}, \ldots, dv'_{i_k}; \widetilde{\mathbf{v}}) \bigotimes_{m \notin \{j_1,\ldots,j_k\}} \delta_{\widetilde{v}_m}(dv'_m). \end{aligned} \tag{5.6}$$

By (5.5), (5.6) and symmetry of A we obtain that for each summand in the sum (5.3), corresponding to $\Pi_n^{(k)}(\widetilde{\mathbf{v}}, A)$, there exists the unique summand in the sum, corresponding to $\Pi_n^{(k)}(\mathbf{v}, A)$, such that these terms are equal. $\square$

If conditions $U1$ and $U2$ are valid for the (n,k)-particle family, it is called an *invariant* family. The infinitesimal characteristic (5.2) of the invariant family has a form

$$\begin{aligned}\pi_n(\nu,B) &= \pi_n^{(k)}(\nu,B)\\ &= \int\limits_{Q^{-1}B}\sum_{J_k} p_{i_1\ldots i_k}(\mathbf{v})T_{i_1\ldots i_k}(dv'_{i_1}\ldots dv'_{i_k};\mathbf{v})\bigotimes_{j\notin\{i_1,\ldots,i_k\}}\delta_{v_j}(dv'_j),\end{aligned} \tag{5.7}$$

where $\nu\in H_n$ and $\mathbf{v}=(v_1,\ldots,v_n)\in Q^{-1}\nu$. Similarly, $\lambda_n(\nu)=\Lambda_n^{(k)}(\mathbf{v})$.

Let us turn to the continuity properties of the transition functions.

Weak continuity conditions:

$P1$. $\Lambda_n^{(k)},\ p_{i_1\ldots i_k}\in\mathbf{C}(D^n)$;

$P2$. the transformation functions $T_{i_1\ldots i_k}(\,\cdot\,;\mathbf{v})$ are w-continuous in $\mathbf{v}$.

Due to the gomeomorphism Q between the set of w-continuous functions $H_n\mapsto\mathbf{R}$ and the set of all invariant continuous functions $D^n\mapsto\mathbf{R}$, the weak continuity of $\pi_n^{(k)}(\nu,\,\cdot\,)$ follows from $P1$ and $P2$, and therefore the infinitesimal characteristics of the family are also w-continuous.

The following proposition is a variant of Theorem 4.1 for the processes under consideration. Suppose that the o.d.e. (4.2) satisfies conditions $A0-A5$, and conditions $P1,P2$ are valid for the invariant (n,k)-particle families.

Let $\psi\in\mathbf{C}^2(H)$. Denote $\Phi_1(\tau,\mu;\nu)=\partial^1T^\tau\psi(\mu;\nu)$ and $f_1(\tau,\mu;v,u)=\Phi_1(\tau,\mu;\delta_v-\delta_u)$ for $\mu\in H$, $\nu\in Z$, $u,v\in D$. Similarly we put $\Phi_2(\tau,\mu;\nu_1,\nu_2)=\partial^2T^\tau\psi(\mu;\nu_1,\nu_2)$ and

$$f_2(v_1,u_1;v_2,u_2)=f_2(\tau,\mu;v_1,u_1;v_2,u_2)=\partial^2T^\tau\psi(\mu;\delta_{v_1}-\delta_{u_1};\delta_{v_2}-\delta_{u_2}) \tag{5.8}$$

with $u_1,u_2,v_1,v_2\in D$ and $\nu_1,\nu_2\in Z$. Sometimes we include ψ in the denotation of f_1 and f_2. For $\mathbf{u}=(u_1,\ldots,u_n)\in D^n$, $J_k=(i_1,\ldots,i_k)$, $\mathbf{v}^{J_k}=(v_{i_1},\ldots,v_{i_k})$ and $\nu=n^{-1}\sum\limits_{i=1}^n\delta_{u_i}$, let $G_n^{(k)}(\nu)=G_n^{(k)}(\,\cdot\,;\nu)$,

$$\begin{aligned}G_n^{(k)}(\,A\,;\nu)\\ =\frac{\lambda_n^{(k)}(\nu)}{n}\sum_{J_k}p_{J_k}(\mathbf{u})\sum_{m=1}^k\Big(T_{J_k}(D^{m-1}\times A\times D^{k-m};\mathbf{u})-\delta_{u_{i_m}}(A)\Big),\end{aligned} \tag{5.9}$$

$$a_k^{(n)}(\tau,\nu)=\Phi_1\left(\tau,\nu;G_n^{(k)}(\nu)-G(\nu)\right), \tag{5.10}$$

$$m_k(\tau,\nu) = \frac{\lambda_n^{(k)}(\nu)}{2n} \sum_{J_k} p_{J_k}(\mathbf{u}) \int_{D^k} T_{J_k}(d\mathbf{v}^{J_k};\mathbf{u}) \sum_{l,m=1}^{k} f_2(v_{i_l},u_{i_l};v_{i_m},u_{im}) \quad (5.11)$$

and $\Delta_n(\psi,\nu) = T_n^t \rho_n \psi(\nu) - T^t\psi(\nu)$.

Theorem 5.1 *Let $n\to\infty$, $k=k(n)$.*
1. *If $\alpha_n = \sup_{\nu\in H_n} \lambda_n^{(k)}(\nu)k^2/n^2 \to 0$ and $\gamma_n = \sup_{\nu\in H_n} \|G_n^{(k)}(\nu) - G(\nu)\| \to 0$, then the convergence* (4.9) *holds on $\mathbf{C}(H)$.*
2. *If $\alpha_n = O(1/n)$ and $\psi\in\mathbf{C}^2(H)$, then uniformly in $\nu\in H_n$*

$$\Delta_n(\psi,\nu) = \int_0^t \mathbf{E}_\nu\Big(a_k^{(n)}(\tau,\xi_n(t-\tau)) + \frac{1}{n} m_k(\tau,\xi_n(t-\tau))\Big)d\tau + o(n^{-1}). \quad (5.12)$$

3. *If $k(n)=O(1)$, $\sup_{\nu\in H_n} \lambda_n^{(k)}(\nu) \asymp n$ and $\gamma_n = O(1/n)$, then $\Delta_n(\psi,\nu) = O(1/n)$.*

Proof. First of all let us demonstrate that

$$n^{-1}\, m_k(\tau,\nu) = \frac{\lambda_n^{(k)}(\nu)}{2} \mathbf{E}\partial^2 T^\tau \psi(\nu; n\theta_\nu^{(n)}, n\theta_\nu^{(n)}) \quad (5.13)$$

for $\nu\in H_n$. We see by (5.7) that

$$\mathbf{E}\Phi_2(\tau,\nu;n\theta_\nu^{(n)}, n\theta_\nu^{(n)}) = \sum_{J_k} p_{J_k}(\mathbf{u}) \int_{D^k} T_{J_k}(d\mathbf{v}^{J_k};\mathbf{u})\, \Phi_2\left(\tau,\nu; \sum_{l=1}^{k}(\delta_{v_{i_l}} - \delta_{u_{i_l}}), \sum_{m=1}^{k}(\delta_{v_{i_m}} - \delta_{u_{i_m}})\right).$$

As Φ_2 is linear in its last two arguments, (5.13) may be easily checked.

Now let us demonstrate that $\xi_n(t) = Q\mathbf{v}_n(t)$ satisfies the hypotheses of Theorem 4.1. Note that $\|\theta_\nu^{(n)}\| \le 2k/n$ for $\nu\in H_n$ and $\beta_n \le \alpha_n \to 0$.

As $G_n^{(k)}(\nu) = \lambda_n^{(k)}(\nu)\mathcal{E}(\theta_\nu^{(n)})$ by (5.7), the condition (4.7) is valid and the first proposition is proved (as well as (3.9)). The second proposition is a consequence of (5.13) and the identity $T_n^{t-\tau}\rho_n\psi(\nu) = \mathbf{E}_\nu\psi(\xi_n(t-\tau))$. The third follows from (5.11) as $\lambda_n^{(k)}/n$, f_1 and f_2 are uniformly continuous. □

Let us turn to the example of (n,k)-particle families that satisfy the hypotheses of Theorem 5.1 and provide simplicity of the expression on the right-hand side of (5.12).

Let $k \geq 1$, and consider distributions $B_k(\,\cdot\,;\mu)$ on $\mathcal{B}^k$, which are weakly continuous in $\mu \in H$. Suppose that for any measurable C_i, $1 \leq i \leq k$, $B_k(C_1 \times \ldots \times C_k;\mu)$ are invariant under permutations of these sets for any $\mu \in H$. Denote $B_{k,1}(dv;\mu) = B_k(dv \times D^{k-1};\mu)$ and $B_{k,2}(dv_1 dv_2;\mu) = B_k(dv_1 dv_2 \times D^{k-2};\mu)$ for $k \geq 2$. Surely, $B_{1,1} = B_1$ and $B_{2,2} = B_2$.

Note that $G(\mu) = B(\mu) - \mu$, where $B(\mu) = B(\,\cdot\,;\mu) \in H$ for $\mu \in H$.

Theorem 5.2 *Let us take k not depending on n, and put $\lambda_n^{(k)} = n/k$, $p_{i_1\ldots i_k}(\mathbf{u}) = 1/A_n^k$ with $A_n^k = n!/(n-k)!$ and $T_{i_1\ldots i_k}(\,\cdot\,;\mathbf{u}) = B_k(\,\cdot\,;Q\mathbf{u})$. Suppose that $B_{k,1}(\,\cdot\,;\mu) = B(\,\cdot\,;\mu)$ for any $\mu \in H$. Then*
1. The hypotheses of Theorem 5.1 hold with $\gamma_n \equiv 0$.
2. If $\nu \in H_n$ and $\psi \in \mathbf{C}^2(H)$, then $\Delta_n(\psi,\nu)$ satisfy (5.12) with $a_k^{(n)} \equiv 0$ and

$$m_1(\tau,\nu) = \frac{1}{2}\int\limits_{D^2} f_2(\tau,\nu;v,u;v,u)B(dv;\nu)\nu(du). \tag{5.14}$$

If $k>1$, then up to order $O(1/n)$

$$\begin{aligned} m_k(\tau,\nu) &= m_1(\tau,\nu) \\ &+\frac{k-1}{2}\int\limits_{D^4} f_2(\tau,\nu;v_1,u_1;v_2,u_2)B_{k,2}(dv_1dv_2;\nu)\nu(du_1)\nu(du_2). \end{aligned} \tag{5.15}$$

We omit the proof of this theorem as it is the particular case of Theorem 5.3. One may take $B_k(\,\cdot\,;\mu) = B^{\otimes k}(\mu)$ or $B_k(\,\cdot\,;\mu) = \mathcal{L}(\eta,\ldots,\eta)$ with $\mathcal{L}(\eta) = B(\mu)$ as the examples of distributions B_k.

If $k=1$ then the $(n,1)$-particle process of Theorem 5.2 allows the evident algorithmical description. Here $\lambda_n = n$, and if $\mathbf{u} = (u_1,\ldots,u_n)$ is the position of the process before a jump, then one randomly chooses a particle with a certain coordinate u_l, which jumps to a new position, distributed according to $B(\nu)$ with $\nu = n^{-1}\sum\limits_{i=1}^{n}\delta_{u_i}$. All other particles stay still.

Now let us turn to a more general situation when transformation functions $T_{i_1\ldots i_k}$ depend on $i_1,\ldots,i_k$. Denote $\mathbf{u}^k = (u_1,\ldots,u_k) \in D^k$. Suppose that $B_k(\,\cdot\,;\mathbf{u}^k;\mu)$ are distributions on $\mathcal{B}^k$ and are weakly continuous in $\mu \in H$ and in $\mathbf{u}^k$. Moreover let $B_k(C_1 \times \cdots \times C_k;\mathbf{u}^k;\mu)$ be invariant under all simultaneous permutations of $C_1,\ldots,C_k$ and $u_1,\ldots,u_k$ for any collection (C_i,u_i) and any $\mu \in H$.

Denote $B_{k,1}(dv;\mathbf{u}^k;\mu) = B_k(dv \times D^{k-1};\mathbf{u}^k;\mu)$,

$$\widehat{B}_k(\mu) = \widehat{B}_k(\,\cdot\,;\mu) = \int\limits_{D^k} B_{k,1}(\,\cdot\,;w_1,\ldots,w_k;\mu)\bigotimes_{i=1}^{k}\mu(dw_i),$$

$$B_{k,2}(dv_1dv_2;\mathbf{u}^k;\mu)=B_k(dv_1dv_2\times D^{k-2};\mathbf{u}^k;\mu),$$

$$\widetilde{B}_k(\mu)=\widetilde{B}_k(\,\cdot\,;\mu)=\int\limits_{D^{k-1}}B_{k,1}(\,\cdot\,;w_1,w_1,w_2,\ldots,w_{k-1};\mu)\bigotimes_{i=1}^{k-1}\mu(dw_i),\quad(5.16)$$

$\widehat{G}_n^{(k)}(\mu)=\widehat{B}_k(\mu)-\mu$ and

$$\widetilde{G}_n^{(k)}(\mu)=\frac{k\lambda_n^{(k)}}{n}\frac{n^k}{A_n^k}\left(\widehat{B}_k(\mu)-\frac{1}{n}\frac{k(k-1)}{2}\widetilde{B}_k(\mu)\right)-\frac{k\lambda_n^{(k)}}{n}\mu\,.\qquad(5.17)$$

Note that $\widetilde{B}_1=\widehat{B}_1$ by definition. We say that the $(n,1)$-particle family *corresponds* to the (n,k)-particle one, if

$$B_1(dv;u;\mu)=\int\limits_{D^{k-1}}B_k(dv\times D^{k-1};u,w_1,\ldots,w_{k-1};\mu)\bigotimes_{i=1}^{k-1}\mu(dw_i).\quad(5.18)$$

Finally, denote

$$T_{i_1\ldots i_k}(\,\cdot\,;\mathbf{u})=B_k(\,\cdot\,;u_{i_1},\ldots,u_{i_k};Q\mathbf{u}).\qquad(5.19)$$

Lemma 5.3 *Consider the (n,k)-particle families with fixed k, $p_{i_1\ldots i_k}(\mathbf{u})=1/A_n^k$, $\lambda_n^{(k)}=n/k+O(1)$ and $T_{i_1\ldots i_k}(\,\cdot\,;\mathbf{u})$ defined by (5.19). Then*

$$\sup_{\nu\in H_n}\|\widetilde{G}_n^{(k)}(\nu)-G_n^{(k)}(\nu)\|=o(1/n),\qquad(5.20)$$

where $G_n^{(k)}$ is defined by (5.9). Moreover, uniformly in $\mu\in H$

$$\widetilde{G}_n^{(k)}(\mu)=\widehat{B}_k(\mu)-\mu+O(1/n).\qquad(5.21)$$

Proof. For the (n,k)-particle processes under consideration,

$$G_n^{(k)}(\nu)=\frac{k\lambda_n^{(k)}}{nA_n^k}\sum_{J_k}B_{k,1}(\,\cdot\,;u_{i_1},\ldots,u_{i_k};\nu)-\frac{k\lambda_n^{(k)}}{n}\nu$$

with $\nu=n^{-1}\sum\limits_{i=1}^n\delta_{u_i}$. As $\widehat{B}_k(\nu)=n^{-k}\sum\limits_{i_1,\ldots,i_k=1}^nB_{k,1}(\,\cdot\,;u_{i_1},\ldots,u_{i_k};\nu)$, then

$$n^k\widehat{B}_k(\nu)-\sum_{J_k}B_{k,1}(\,\cdot\,;u_{i_1},\ldots,u_{i_k};\nu)=\sum_{I_k}B_{k,1}(\,\cdot\,;u_{i_1},\ldots,u_{i_k};\nu),$$

where Σ_{I_k} stands for the sum over all collections $(i_1,\ldots,i_k)$ which have at least two coinciding i_j.

As $\lambda_n^{(k)}k/n = 1+O(1/n)$, $n^k/A_n^k = 1+O(1/n)$ and the number of $(i_1,\ldots,i_k)$ with at least m coinciding i_j has the order $O(n^{k-m+1})$, then

$$\begin{aligned}
&\sum_{I_k} B_{k,1}(\,\cdot\,;u_{i_1},\ldots,u_{i_k};\nu)\\
&\quad=\frac{k(k-1)}{2}\sum_{i=1}^{n}\sum_{J_{k-2}} B_{k,1}(\,\cdot\,;u_i,u_i,u_{i_1},\ldots,u_{i_{k-2}};\nu)+O(n^{k-2})\\
&\quad=\frac{k(k-1)}{2}\sum_{i=1}^{n}\sum_{i_1,\ldots,i_{k-2}=1}^{n} B_{k,1}(\,\cdot\,;u_i,u_i,u_{i_1},\ldots,u_{i_{k-2}};\nu)+O(n^{k-2})\\
&\quad=\frac{k(k-1)n^{k-1}}{2}\widetilde{B}_k(\nu)+O(n^{k-2}),
\end{aligned}$$

and the first proposition is proved. The second one is an immediate consequence of (5.17). The proof is complete. □

Theorem 5.3 *Suppose k does not depend on n, $p_{i_1\ldots i_k}(\mathbf{u}) = 1/A_n^k$, $\lambda_n^{(k)} = n/k+O(1)$, and $T_{i_1\ldots i_k}$ are defined by* (5.19). *Let B_1 and B_k satisfy* (5.18).
1. *If $\sup\limits_{\mu\in H}\|\widetilde{G}_n^{(k)}(\mu)-G(\mu)\| = o(1)$, then propositions* 1 *and* 2 *of Theorem* 5.1 *are valid, and* (5.12) *holds for $\psi\in\mathbf{C}^2(H)$. Moreover, uniformly in $\nu\in H_n$*

$$a_k^{(n)}(\tau,\nu)=\Phi_1\Big(\tau,\nu;\widetilde{G}_n^{(k)}(\nu)-G(\nu)\Big)+o(1/n), \tag{5.22}$$

$$m_1(\tau,\nu)=\frac{1}{2}\int\limits_{D^2} f_2(\tau,\nu;v,u;v,u)B_1(dv;u;\nu)\nu(du) \tag{5.23}$$

and up to order $O(1/n)$

$$\begin{aligned}
&m_k(\tau,\nu)=m_1(\tau,\nu)\\
&\quad+\frac{k-1}{2}\int\limits_{D^4} f_2(\tau,\nu;v_1;u_1,v_2,u_2)\mathfrak{B}_{k,2}(dv_1dv_2;u_1,u_2;\nu)\nu^{\otimes 2}(du_1du_2)
\end{aligned}\tag{5.24}$$

for $k\geq 2$ with

$$\mathfrak{B}_{k,2}(dv_1dv_2;u_1,u_2;\nu)=\int\limits_{D^{k-2}} B_{k,2}(dv_1dv_2;u_1,\ldots,u_k;\nu)\bigotimes_{l=3}^{k}\nu(du_l). \tag{5.25}$$

2. *If additionally $\widehat{B}_k = B$, then $a_k^{(n)}(\tau,\nu)=O(1/n)$ and $\Delta_n(\psi,\nu)=O(1/n)$.*

Proof. It is easy to see that the invariance and w-continuity conditions $U1$, $U2$, $P1$ and $P2$ are valid for the (n,k)-particle processes under consideration. By (5.20), the hypotheses of propositions 1 and 2 of Theorem 5.1 are valid too. Therefore (5.12) holds.

Consider $k=1$ and $\nu=Q\mathbf{u}$. By (5.11) we obtain that

$$m_1(\tau,\nu)=\frac{1}{2n}\sum_{i=1}^{n}\int\limits_{D}B_1(dv;u_i,\nu)f_2(v,u_i;v,,u_i),$$

(see (5.8)) and therefore $m_1(\tau,\nu)$ satisfies (5.23). For $k>1$ we get by (5.11) the identity

$$m_k(\tau,\nu)=\frac{1}{2kA_n^k}\int\limits_{D^k}\sum_{J_k}B_k(dv_{i_1}\dots dv_{i_k};\nu)\sum_{l,m=1}^{k}f_2(v_{i_l},u_{i_l};v_{i_m},u_{i_m}). \quad (5.26)$$

Note that the expression $\int\limits_{D^k}\sum\limits_{J_k}B_k(dv_{i_1}\dots dv_{i_k};\nu)f_2(v_{i_l},u_{i_l};v_{i_m},u_{i_m})$ does not depend on l and m but depends on the fact whether $l=m$ or not. Let us divide the sum on the right-hand side of (5.26) into two parts, the first consisting of summands with $l=m$, and the second of summands with $l\neq m$. As $A_n^k=n^k+O(n^{k-1})$ and the number of $(i_1,\dots,i_k)$ with at least 2 coinciding i_j has the order $O(n^{k-1})$,

$$\begin{aligned}
m_k(\tau,\nu)=&\frac{1}{2A_n^k}\sum_{J_k}\int\limits_{D^k}B_k(dv_{i_1}\dots dv_{i_k};u_{i_1},\dots,u_{i_k};\nu)f_2(v_{i_1},u_{i_1};v_{i_1},u_{i_1})\\
&+\frac{k-1}{2A_n^k}\sum_{J_k}\int\limits_{D^k}B_k(dv_{i_1}\dots dv_{i_k};u_{i_1},\dots,u_{i_k};\nu)f_2(v_{i_1},u_{i_1};v_{i_2},u_{i_2})\\
=&\frac{1}{2n^k}\sum_{i_1,\dots,i_k=1}^{n}\int\limits_{D}B_{k,1}(dv;u_{i_1},\dots,u_{i_k};\nu)f_2(v,u_{i_1};v,u_{i_1})+\\
&+\frac{k-1}{2n^k}\sum_{i_1,\dots,i_k=1}^{n}\int\limits_{D^2}B_{k,2}(dv_1dv_2;u_{i_1},\dots,u_{i_k};\nu)f_2(v_1,u_{i_1};v_2,u_{i_2})+O(n^{-1}).
\end{aligned}$$

Simple manipulations prove the first proposition. As for the second, it follows from the third proposition of Theorem 5.1, as $\gamma_n=O(1/n)$ by (5.21). □

Corollary 5.1 *Let the hypotheses of Theorem* 5.3 *be fulfilled.*
1. *If* $\lambda_n^{(k)}=n/k$, *then uniformly in* $\nu\in H_n$

$$a_k^{(n)}(\tau,\nu)=\frac{k(k-1)}{2n}\,\Phi_1\Big(\tau,\nu;\widehat{B}_k(\nu)-\widetilde{B}_k(\nu)\Big)+\Phi_1\Big(\tau,\nu;\widehat{B}_k(\nu)-B(\nu)\Big)$$

up to order $o(1/n)$.
2. *If $\lambda_n^{(k)} = n^{-(k-1)} A_n^k/k$, then uniformly in $\nu \in H_n$*

$$a_k^{(n)}(\tau,\nu) = \frac{k(k-1)}{2n}\,\Phi_1\Big(\tau,\nu;\nu - \widetilde{B}_k(\nu)\Big) + \Phi_1\Big(\tau,\nu;\widehat{B}_k(\nu) - B(\nu)\Big)$$

up to order $o(1/n)$.

Remark 5.2 1. *Corollary 5.1 shows that under the hypotheses of Theorem 5.3 the main term of $a_k^{(n)}(\tau,\nu)$ depends on the exact form of $\lambda_n^{(k)}$. If we are interested in the order $a_k^{(n)}(\tau,\nu) = O(1/n)$, we need to have the identity $\widehat{B}_k = B$ in both cases of Corollary 5.1. To get the order $o(1/n)$, we must additionally require $\widetilde{B}_k = \widehat{B}_k$ in the first case and $\widetilde{B}_k(\mu) = \mu$ in the second.*
2. *If $\lambda_n^{(k)} = n/k$ and distributions $B_k(\,\cdot\,;u_1,\ldots,u_k;\mu)$ do not depend upon $u_1,\ldots,u_k$, then $\widetilde{B}_k = \widehat{B}_k = B$ and, due to Corollary 5.1, $a_k^{(n)}(\tau,\nu) = o(1/n)$. Moreover, exact calculation shows that $a_k^{(n)} \equiv 0$.*

Now let us turn to the examples, assuming that $\lambda_n^{(k)} = n/k$. If $B(\mu) = \int_D T'(\,\cdot\,;w;\mu)\mu(dw)$, then one may consider the (n,k)-particle process with

$$B_k(\,\cdot\,;u_1,\ldots,u_k;\mu) = \bigotimes_{i=1}^{k} T'(\,\cdot\,;u_i;\mu). \tag{5.27}$$

If we take $k=2$ and suppose that $T'(\,\cdot\,;u;\mu) = \int_D T''(\,\cdot\,;u,w;\mu)\mu(dw)$, then some other $(n,2)$-processes are available. For example, one can take

$$B_2(dv_1 dv_2;u_1,u_2;\mu) = T''(dv_1;u_1,u_2;\mu) \otimes T''(dv_2;u_2,u_1;\mu). \tag{5.28}$$

If $\varphi(u) = \varphi(u;u_1,u_2;\mu)$ is a measurable function satisfying the condition

$$\int_D T''(dw;u_1,u_2;\mu)\delta_{\varphi(w)}(dv) = T''(dv;u_2,u_1;\mu)$$

for any $u_1,u_2 \in D$ and $\mu \in H$, then

$$B_2(dv_1 dv_2;u_1,u_2;\mu) = T''(dv_1;u_1,u_2;\mu) \otimes \delta_{\varphi(v_1)}(dv_2) \tag{5.29}$$

will do. Note that both statements of Theorem 5.3 are valid for these examples. Moreover, dealing with (5.27) we see that $\widetilde{B}_k = B$ and $a_k^{(n)}(\tau,\nu) = o(1/n)$. As for (5.28) and (5.29), we have $\widetilde{B}_2(\mu) = \int_D T''(\,\cdot\,;u,u;\mu)\mu(du)$ here. Therefore $a_2^{(n)}(\tau,\nu) = n^{-1}\,\Phi_1\Big(\tau,\nu;B(\nu) - \widetilde{B}_2(\nu)\Big) + o(1/n)$.

Remark 5.3 *The second proposition of Corollary 5.1 shows that if $k=2$, B_2 is defined by (5.28) or (5.29), $T''(\,\cdot\,;u,u;\mu) = \delta_u$ and $\lambda_n = (n-1)/2$, then $a_2^{(n)}(\tau,\nu) = o(1/n)$. The calculation gives $a_2^{(n)} \equiv 0$.*

6 Estimate properties

Suppose that the w-continuity conditions $P1, P2$ are fulfilled for the invariant (n,k)-particle empirical process $\xi_n(t)$. Now we turn to the statistical properties of random variables $\psi(\xi_n(t))$ as the estimators of $\psi(\mu_t)$.

6.1 The initial distribution

Let $\Xi = \{\Omega, \xi(t), \mathbf{P}_\mu, \mathrm{F}_t\}$ be a Markov family with a phase space $Y \subset H$ and a transition function $P(\mu, t, d\nu)$. Consider a distribution $\mathcal{L}$ on the Borel σ-algebra $\mathcal{B}_Y$. Then one may deal with the distribution

$$\mathbf{P}_{\mathcal{L}}(A) = \int\limits_Y \mathbf{P}_\mu(A)\mathcal{L}(d\mu)$$

and the corresponding mathematical expectation $\mathbf{E}_{\mathcal{L}}$. Surely $\mathcal{L}$ may be interpreted as the initial distribution of the process $\xi(t)$. Thus $(\Omega, \xi(t), \mathbf{P}_{\mathcal{L}}, \mathrm{F}_t)$ is the Markov process with the same transition function $P(\mu, t, d\nu)$ and the initial distribution $\mathcal{L}$. For the brevity, we use the mathematical expectation sign $\mathbf{E}$ instead of $\mathbf{E}_{\mathcal{L}}$.

The following proposition is a consequence of the well-known result [6].

Theorem 6.1 *Let $A0 - A3$ hold for the equation* (4.2). *Under the hypotheses of the first statement of Theorem* 5.1, *if $\mu \in H$ and $\mathcal{L}(\xi_n(0)) \overset{w}{\to} \delta_\mu$, then the distributions of the measure-valued processes $\xi_n(\,\cdot\,)$ weakly converge in the space of probability measures on the Skorohod space* $\mathrm{D}([0,\infty), D)$ *to the Dirac distribution, concentrated at the measure-valued function $\mu.(\mu)$.*

Proof. All we need is to remark that due to Theorem 5.1,

$$\sup_{\nu \in H_n} |T_n^t \mathfrak{p}_n \psi(\nu) - T^t \psi(\nu)| \underset{n\to\infty}{\longrightarrow} 0. \tag{6.1}$$

Now the result follows (see [6, Th. 2.11, Ch. 4]) from (6.1) and the condition $\mathcal{L}(\xi_n(0)) \overset{w}{\to} \delta_\mu$. □

Remark 6.1 *Under the hypotheses of Theorem* 6.1,

$$\mathbf{E}\psi(\xi_n(t)) \underset{n\to\infty}{\longrightarrow} \psi(\mu_t(\mu)), \qquad \psi \in \mathbf{C}(H), \quad t \ge 0. \tag{6.2}$$

Remark 6.1 shows that the random variable $\psi(\xi_n(t))$ is the consistent estimate of $\psi(\mu_t(\mu))$. Let us consider errors of such an estimation: the bias

$\mathbf{E}\psi(\xi_n(t)) - \psi(\mu_t(\mu))$ and the mean square error $\mathbf{E}(\psi(\xi_n(t)) - \psi(\mu_t(\mu)))^2$, and discuss the contribution of the initial distribution $\mathcal{L}(\xi_n(0))$ to these errors.

Suppose that $\nu \in H_n$, $\mu \in H$ and denote $M(\nu) = \mathbf{E}_\nu \psi(\xi_n(t)) - \psi(\mu_t(\nu))$, $S^2(\nu) = \mathbf{E}_\nu \Big(\psi(\xi_n(t)) - \psi(\mu_t(\nu))\Big)^2$,

$$M_0 = M_0(\psi, t, \mu) = \mathbf{E}\psi(\mu_t(\xi_n(0))) - \psi(\mu_t(\mu)) \tag{6.3}$$

and

$$S_0^2 = S_0^2(\psi, t, \nu) = \mathbf{E}\Big(\psi(\mu_t(\xi_n(0))) - \psi(\mu_t(\mu))\Big)^2. \tag{6.4}$$

Standard considerations prove the following assertion.

Lemma 6.1 *Let $\psi \in \mathbf{C}(H)$, then $\mathbf{E}\psi(\xi_n(t)) - \psi(\mu_t(\mu)) = M_0 + \mathbf{E}M(\xi_n(0))$. If $\mathcal{L}(\xi_n(0)) \xrightarrow{w} \delta_\mu$ and $\sup\limits_{\nu \in H_n} |M(\nu)| = O(n^{-1})$, then*

$$\mathbf{E}\Big(\psi(\xi_n(t)) - \psi(\mu_t(\mu))\Big)^2 = S_0^2 + \mathbf{E}S^2(\xi_n(0)) + o(1/n).$$

Let us study the asymptotical ($n \to \infty$) behaviour of M_0 and S_0^2. Let $\mu \in H$, $\xi_n = \xi_n(0) = n^{-1} \sum\limits_{i=1}^{n} \delta_{\zeta_i^{(n)}}$ and suppose that $\mathcal{L}(\zeta_i^{(n)}) = \mu$ and $\mathcal{L}(\zeta_1^{(n)}, \ldots, \zeta_n^{(n)})$ is invariant under any permutation of $\zeta_i^{(n)}$. Denote $\mathcal{L}(\zeta_i^{(n)}, \zeta_j^{(n)}) = \mathcal{P}_2^{(n)}$ for $i \neq j$, and $V_n = \Big\|\mathcal{P}_2^{(n)} - \mu \otimes \mu\Big\|$.

Theorem 6.2 *Let $T^t\psi \in \mathbf{C}^3(H)$. If $V_n = o(1/n)$, then up to order $o(1/n)$*

$$M_0(\psi, t, \mu) = \frac{1}{2n} \int\limits_{D^3} f_2(t, \mu; v_1, u; v_2, u)\mu(du)\mu(dv_1)\mu(dv_2) \tag{6.5}$$

and

$$S_0^2(\psi, t, \mu) = \frac{1}{n} \int\limits_{D^2} f_1^2(t, \mu; v, u)\mu(dv)\mu(du). \tag{6.6}$$

Theorem 6.3 *Let $T^t\psi \in \mathbf{C}^5(H)$. 1. Suppose that*

$$\mathcal{P}_2^{(n)} = \frac{n+1}{n}\, \mu \otimes \mu - \frac{1}{n}\, \nu_n + \gamma_n, \tag{6.7}$$

where $\nu_n \xrightarrow{w} \mathcal{L}(\zeta, \zeta)$ with $\mathcal{L}(\zeta) = \mu$, and $\|\gamma_n\| = o(1/n)$. Then

$$M_0(\psi, t, \mu) = o(1/n) \quad \text{and} \quad S_0^2(\psi, t, \mu) = o(1/n). \tag{6.8}$$

2. *Suppose that for any $n \geq 1$ there exists a decomposition $\mu = n^{-1} \sum_{i=1}^{n} \mu_i^{(n)}$ such that $\mu_i^{(n)} \in H$ and $\max_{i \in I_n} \operatorname{diam}(D_{i,n}) \to 0$ with $D_{i,n} = \operatorname{support}(\mu_i^{(n)})$, $I_n \subset \mathcal{N}_n = \{1, 2, \ldots, n\}$ and $n^{-1} \operatorname{card} I_n \to 1$ as $n \to \infty$. If $\widehat{\zeta}_i^{(n)}$ are independent, $\mathcal{L}(\widehat{\zeta}_i^{(n)}) = \mu_i^{(n)}$ and $\xi_n = n^{-1} \sum_{i=1}^{n} \delta_{\widehat{\zeta}_i^{(n)}}$, then (6.8) holds.*

The proof of both theorems may be found in [9] (in a brief form) and [10].

Remark 6.2 1. *The natural sufficient condition for* (6.5) *and* (6.6) *is the independence of the $\zeta_i^{(n)}$ (then $V_n \equiv 0$). The second statement of Theorem* 6.3 *shows that a proper variant of the stratification technique gives smaller orders of both M_0 and S_0^2, than those in the standard case of i.i.d. random variables.*
2. *Let $\psi \in \mathbf{C}^3(H)$. To establish $T^t\psi \in \mathbf{C}^3(H)$ it is sufficient to require that conditions $A0 - A5$ are valid for the right-hand side G of the equation* (4.2) *together with the additional conditions on the third derivative of G, similar to A4 and A5. The analogous remark holds for the condition $T^t\psi \in \mathbf{C}^5(H)$.*

6.2 Bias and deviation

Denote for $\nu \in H$

$$s_1^2(\psi, \tau, \nu) = \int_{D^2} f_1^2(\tau, \nu; v, u) B_1(dv; u; \nu) \nu(du) \tag{6.9}$$

and for $k \geq 2$

$$s_k^2(\psi, \tau, \nu) = s_1^2(\psi, \tau, \nu) + (k-1)\sigma_k(\psi, \tau, \nu) \tag{6.10}$$

with

$$\sigma_k(\psi, \tau, \nu) = \int_{D^4} f_1(v_1, u_1) f_1(v_2, u_2) \mathfrak{B}_{k,2}(dv_1 dv_2; u_1, u_2; \nu) \nu^{\otimes 2}(du_1 du_2), \tag{6.11}$$

where $\mathfrak{B}_{k,2}$ are defined by (5.25), and $f_1(v, u) = f_1(\tau, \nu; v, u)$.

Consider $m_k(\nu) = m_k(\psi, \tau, \nu)$ as the functions, defined on H by the formulas (5.23) and (5.24).

Theorem 6.4 *Suppose that a) the (n,k)-particle process $\xi_n(t)$ satisfies the hypotheses of the second statement of Theorem* 5.3 *with $\lambda_n^{(k)} = n/k$;*
b) its initial value $\xi_n = \xi_n(0)$ satisfies the hypotheses of Theorem 6.2;

c) the balance equation (4.2) *and the function* $T^t\psi$ *satisfy the hypotheses of Theorem* 6.2. *Then*
1. *The bias of the estimator* $\psi(\xi_n(t))$ *for* $\psi(\mu_t(\mu))$ *has a form*

$$\mathbf{E}\psi(\xi_n(t)) - \psi(\mu_t(\mu)) = M_0(\psi,t,\mu) + M_k(\psi,t,\mu) + o(1/n), \tag{6.12}$$

where $M_0(\psi,t,\mu)$ *is defined by* (6.5),

$$M_k(\psi,t,\mu) = \frac{1}{n}\int\limits_0^t \Big(\alpha_k(\psi,\tau,\mu_{t-\tau}(\mu)) + m_k(\psi,\tau,\mu_{t-\tau}(\mu))\Big)\,d\tau, \tag{6.13}$$

$\alpha_k = \alpha_k(\psi,\tau,\nu)$ *with* $\nu \in H$ *is defined by*

$$\alpha_k(\psi,\tau,\nu) = \frac{k(k-1)}{2}\,\Phi_1\Big(\tau,\nu; B(\nu) - \widetilde{B}_k(\nu)\Big)\,, \tag{6.14}$$

and $m_k = m_k(\psi,\tau,\nu)$ — *by* (5.23) *and* (5.24).
2. *The mean square error has a form*

$$\mathbf{E}\Big(\psi(\xi_n(t)) - \psi(\mu_t(\mu))\Big)^2 = S_0^2(\psi,t,\mu) + S_k^2(\psi,t,\mu) + o(1/n), \tag{6.15}$$

where $S_0^2(\psi,t,\mu)$ *may be found in* (6.6),

$$S_k^2(\psi,t,\mu) = \frac{1}{n}\int\limits_0^t s_k^2(\psi,\tau,\mu_{t-\tau}(\mu))d\tau \tag{6.16}$$

and s_k^2 *is defined by* (6.9), (6.10) *and* (6.11).

Proof. 1. By Theorem 6.2, Lemma 6.1, (5.12) and Theorem 5.3 all we need is to check that $\mathbf{E}\,m_k(\psi,\tau,\xi_n(t-\tau)) \underset{n\to\infty}{\longrightarrow} m_k(\psi,\tau,\mu_{t-\tau}(\mu))$ and the analogous convergence holds for $\mathbf{E}\,\alpha_k(\psi,\tau,\xi_n(t-\tau))$. But these facts follow from Theorem 6.1, as $\alpha_k(\psi,\tau,\nu)$ and $m_k(\psi,\tau,\nu)$ are w-continuous in ν due to the weak continuity of $\widetilde{B}_k$, B_k and B.

2. Note that $S_k^2(\psi,t,\mu) = M_k(\psi^2,t,\mu) - \psi(\mu_t(\mu))\,M_k(\psi,t,\mu)$. Thus, due to Theorem 6.2, the second proposition would be proved if we check that

$$s_k^2(\psi,\tau,\mu_{t-\tau}(\mu)) = m_k(\psi^2,\tau,\mu_{t-\tau}(\mu)) - \psi(\mu_t(\mu))\,m_k(\psi,\tau,\mu_{t-\tau}(\mu))$$

and $\alpha_k(\psi^2,\tau,\mu_{t-\tau}(\mu)) = \psi(\mu_t(\mu))\,\alpha_k(\psi,\tau,\mu_{t-\tau}(\mu))$ up to order $O(1/n)$.

Equalities (5.23), (5.24) and (6.14) show that m_k is a sum of integrals with the integrands expressed through f_2, and α_k is an integral with the integrand expressed through f_1. Therefore it is sufficient to prove that

$$\begin{aligned} & f_2(\psi^2;\tau,\nu;v_1,u_1;v_2,u_2) \\ & \quad = \psi(\mu_t(\mu)) f_2(\psi;\tau,\nu;v_1,u_1;v_2,u_2) + 2 f_1(\psi;\tau,\nu;v_1,u_1) f_1(\psi;\tau,\nu;v_2,u_2) \end{aligned}$$

and $f_1(\psi^2;\tau,\nu;v,u) = \psi(\mu_t(\mu))\, f_1(\psi;\tau,\nu;v,u)$ with $\nu = \mu_{t-\tau}(\mu)$. But these equalities follow from the identity $\psi(\mu_t(\mu)) = T^\tau \psi(\mu_{t-\tau}(\mu))$ and the elementary derivative properties. □

Remark 6.3 1. *Theorem* 6.3 *gives us the conditions when* M_0 *and* S_0^2 *have the order* $o(1/n)$ *in* (6.12) *and* (6.15).
2. *If* $\lambda_n^{(k)} = n/k + O(1)$ *then* S_k^2 *is determined by* (6.16) *with* (6.9) *and* (6.10).

7 Boltzmann-like equations

7.1 Equations and algorithms

Consider a homogeneous Boltzmann-like equation, that is, the o.d.e.

$$\begin{aligned} & \frac{d\mu_t}{dt}(du) = G(\mu_t)(du) \\ & \quad = \int_{D^2} \sigma(u_1,u_2)\widetilde{T}(du;u_1,u_2)\mu_t^{\otimes 2}(du_1 du_2) - \mu_t(du)\int_D \sigma(u,u_2)\mu_t(du_2), \end{aligned} \tag{7.1}$$

where $\mu \stackrel{def}{=} \mu_t|_{t=0} \in H$, $\widetilde{T}(\,\cdot\,;u_1,u_2)$ is a distribution on $\mathcal{B}$, functions $\sigma(u_1,u_2)$ and $\widetilde{T}(A;u_1,u_2)$ with fixed $A \in \mathcal{B}$ are measurable in (u_1,u_2), and $0 \le \sigma(u_1,u_2) \le c < \infty$. As usual, we suppose that $c = 1$. We call $\widetilde{T}$ a *Boltzmann transformation function.*

If we define the new Boltzmann transformation function by

$$T(\,\cdot\,;u_1,u_2) = \sigma(u_1,u_2)\widetilde{T}(\,\cdot\,;u_1,u_2) + (1-\sigma(u_1,u_2))\delta_{u_1}, \tag{7.2}$$

then (7.1) may be rewritten as

$$\frac{d\mu_t}{dt} = G(\mu_t) = \int_{D^2} T(\,\cdot\,;u_1,u_2)\mu_t(du_1)\mu_t(du_2) - \mu_t. \tag{7.3}$$

Surely, $G(\nu) = B(\nu) - \nu$ with

$$B(\nu) \stackrel{def}{=} \int\limits_{D^2} T(\,\cdot\,; u_1, u_2)\nu(du_1)\nu(du_2). \tag{7.4}$$

Lemma 7.1 *Let $T(\,\cdot\,; u_1, u_2)$ be w-continuous in (u_1, u_2). Then the Boltzmann-like equation (7.3) satisfies conditions $A0 - A5$. Moreover, the analogous statement for any higher order derivative is valid too.*

Proof. As for $A0$ (see Section 4), it is evident. It may be easily seen that

$$\partial^1 B(\mu; \nu) = \int\limits_{D^2} \Big(T(\,\cdot\,; u_1, u_2) + T(\,\cdot\,; u_2, u_1)\Big)\mu(du_1)\nu(du_2),$$

$$\partial^2 B(\mu; \nu_1, \nu_2) = \int\limits_{D^2} \Big(T(\,\cdot\,; u_1, u_2) + T(\,\cdot\,; u_2, u_1)\Big)\nu_1(du_1)\nu_2(du_2)$$

and $d^m B(\mu)/d\mu^m \equiv \mathbf{0}$ for $m > 2$. This means that $G(\mu) = B(\mu) - \mu$ has derivatives of any order and therefore conditions $A1$, $A2$ and $A4$ on strong properties of G are fulfilled. Moreover, G and its derivatives are w-continuous by the weak continuity of T. Therefore $A3$ and $A5$ are valid, as well as the analogous statements for higher derivatives. □

Remark 7.1 1. *Let T be defined by (7.2). If $\widetilde{T}$ is w-continuous and $\sigma \in \mathbf{C}(D^2)$, then T is w-continuous.*
2. *Lemma 7.1 gives us the mild conditions for the existence/uniqueness of the smooth solution of the Boltzmann-like equation (7.3), and tells us that all previous approximation results may be applied to it.*

From now on we consider weakly continuous transformation functions $T(\,\cdot\,; u_1, u_2)$, which are invariant under permutation of u_1 and u_2, and (n, k)-particle processes satisfying the hypotheses of Theorem 6.4. This means that for a smooth ψ both bias and mean square error of the estimate $\psi(\xi_n(t))$ have the order $O(1/n)$ and the form, defined by (6.5), (6.12) – (6.14), (5.23) and (5.24) for the bias, and by (6.15), (6.6), (6.16), (6.9) and (6.10) for the mean square error. Note that as the bias is rather small, the mean square error coincides with the variance of the estimate up to order $o(1/n)$.

By (6.15), the main term of the variance consists of two summands, the first related to the simulation of the initial distribution μ and the second —

to the simulation of the process itself. Thus the difference between (n,k)-particle processes, solving Boltzmann-like equations, appears in the second summand, but not in the first one, and, in terms of Theorem 6.4, we are interested in $M_k(\psi,t,\mu)$ and $S_k^2(\psi,t,\mu)$, but not in $M_0(\psi,t,\mu)$ and $S_0^2(\psi,t,\mu)$. Moreover (see Theorem 6.3), one may hope that with the help of a proper variant of the stratification he can get both terms $M_0(\psi,t,\mu)$ and $S_0^2(\psi,t,\mu)$ as $o(1/n)$, while $M_k(\psi,t,\mu)$ and $S_k^2(\psi,t,\mu)$ have the order $O(1/n)$.

Different examples of (n,k)-particle processes, solving the general balance equation (4.2) and its variants, were given in Section 5. Generally, all of them have the same form. First, we fix k and take $\lambda_n^{(k)} = n/k$ and $p_{i_1\ldots i_k} = 1/A_n^k$.

Then we choose distributions $B_k(\,\cdot\,;u_1,\ldots,u_k;\mu)$ on $\mathcal{B}^k$ with $u_i \in D$ and $\mu \in H$. These distributions must satisfy the simple invariance and continuity conditions (see Section 5) and are connected with $G(\mu) = B(\mu) - \mu$ by

$$B(dv;\mu) = \int\limits_{D^k} B_k\left(dv \times D^{k-1}; w_1,\ldots,w_k;\mu\right) \bigotimes_{i=1}^{k} \mu(dw_i). \tag{7.5}$$

Finally, for $\mathbf{u} = (u_1,\ldots,u_n)$, $\nu = n^{-1}\sum\limits_{i=1}^{n}\delta_{u_i}$ and pairwise different $1 \le i_1,\ldots,i_k \le n$ we take the transformation functions as

$$T_{i_1\ldots i_k}(\,\cdot\,;\mathbf{u}) = B_k(\,\cdot\,;u_{i_1},\ldots,u_{i_k};\nu)$$

and are ready to describe the jump mechanism of the process.

Let $\mathbf{u} = (u_1,\ldots,u_n)$ be the position of the process just before a jump. Then one randomly chooses k different particles with numbers $i_1,\ldots,i_k$, and these particles perform a jump according to $T_{i_1\ldots i_k}(\,\cdot\,;\mathbf{u})$, while all other particles stay still. Note that the choice of particles is uniform in the sense that any collection $(i_1,\ldots,i_k)$ is taken with the probability $1/A_n^k = (n-k)!/n!$.

Examples for distributions B_k were given in Section 5. Now we reformulate these examples for Boltzmann-like equations and describe them in a more algorithmic manner. Note that the w-continuity of distributions B_k, corresponding to these examples, follows from the hypotheses of Lemma 7.1.

Thus let us consider the following algorithms.

S-algorithm (*The simplest algorithm*).
The simplest algorithm corresponds to $k=1$ (therefore $\lambda_n^{(1)} = n$, $p_i = 1/n$) and to $B_1 = B$. For Boltzmann-like equations this means that

$$T_i(dv;\mathbf{u}) = \frac{1}{n^2}\sum_{l,m=1}^{n} T(dv;u_l,u_m),$$

where T is the Boltzmann transformation function. Thus the unique particle modifies its position through a jump of the process, this particle is chosen randomly over the entire set of particles, and to get a new position of the particle with coordinate u_i just before a jump, one must randomly choose two particles with coordinates u_l and u_m and simulate the distribution $T(\,\cdot\,; u_l, u_m)$. Briefly, particles l and m collide, but the particle i goes through a jump.

N-algorithm (*The Nanbu algorithm*).
Here we have $k = 1$ and

$$B_1(dv; u_1; \mu) = \int_D T(dv; u_1, u_2)\mu(du_2), \tag{7.6}$$

that corresponds to

$$T_i(dv; \mathbf{u}) = \frac{1}{n}\sum_{j=1}^{n} T(du; u_i, u_j). \tag{7.7}$$

Thus, having the particle number i with the coordinate u_i to get a new position, we randomly choose a particle number j with a coordinate u_j, and they collide by $T(\,\cdot\,; u_i, u_j)$. Up to the possibility for a particle to collide with itself (this occurs with the probability $1/n$), the described process corresponds to a variant of the continuous-time homogeneous Nanbu process.

If we take $k = 2$, then $\lambda_n^{(2)} = n/2$, two particles get their new positions through a process jump, and this pair of particles is chosen with the probability $1/A_n^2$. Dealing with Boltzmann-like equations, we call the $(n, 2)$-particle process a *Bird-like* (and the corresponding algorithm an **B**-algorithm) if

$$B_2(dv \times D; u_1, u_2; \mu) = T(dv; u_1, u_2).$$

Note that the invariance conditions on B_2 have a form $B_2(dv_1 dv_2; u_1, u_2; \mu) = B_2(dv_2 dv_1; u_2, u_1; \mu)$ for any u_1, u_2 and μ. Surely, one may deal with many Bird-like processes. Let us consider some of them.

B$_{in}$-algorithm (*Independent collisions*).
For this algorithm we take

$$\begin{aligned} &B_2(dv_1 dv_2; u_1, u_2; \mu) \\ &\quad = \int_D T(dv_1; u_1, w_1)\mu(dw_1) \otimes \int_D T(dv_2; u_2, w_2)\mu(dw_2) \end{aligned} \tag{7.8}$$

and therefore

$$T_{ij}(dv_1 dv_2; \mathbf{u}) = \frac{1}{n^2} \sum_{l,m=1}^{n} T(dv_1; u_i, u_l) \otimes T(dv_2; u_j, u_m).$$

Here each collision corresponds to a pair of independent Nanbu collisions.

$\mathbf{B}_d$-algorithm (*Double collisions*).
Now we suppose that B_2 does not depend on μ and

$$B_2(dv_1 dv_2; u_1, u_2) = T(dv_1; u_1, u_2) \otimes T(dv_2; u_2, u_1). \tag{7.9}$$

Surely, T_{ij} has the same form here and, algorithmically, we simulate two independent collisions of the same pair of particles.

Let $\varphi(u) = \varphi(u; u_1, u_2)$ be a continuous function such that for any u_1, u_2

$$\int_D T(dw; u_1, u_2)\delta_{\varphi(w;u_1,u_2)}(dv) = T(dv; u_2, u_1). \tag{7.10}$$

In other words, if $\mathcal{L}(\eta_1) = T(\,\cdot\,; u_1, u_2)$, then $\eta_2 = \varphi(\eta_1; u_1, u_2)$ has the same distribution (one can take $\varphi(u) \equiv u$ as an example of such a situation). Then we can consider one more Bird-like process.

$\mathbf{B}_\varphi$-algorithm (φ-*collisions*).
As dealing with double collisions, we take B_2 not depending on μ, and put

$$B_2(dv_1 dv_2; u_1, u_2) = T(dv_1; u_1, u_2) \otimes \delta_{\varphi(v_1;u_1,u_2)}(dv_2). \tag{7.11}$$

Then (7.10) guarantees us the identity (7.5), T_{ij} depends only on u_i, u_j and has the same form as B_2. Thus, after two particles with coordinates u_i and u_j are chosen for a collision, the first of them gets its new position v_i according to $T(\,\cdot\,; u_i, u_j)$, while the coordinate of the second particle turns to $\varphi(v_i; u_i, u_j)$.

Note that all **B**-algorithms correspond to the Nanbu one in the sense that

$$B_1(dv; u; \mu) = \int_D B_2(dv \times D; u, w; \mu)\mu(dw) \tag{7.12}$$

(see (5.18)). As for the **B**-algorithms mentioned above, they are similar to examples (5.27) – (5.29), valid for more general equations. Surely, one can consider algorithms with $k > 2$ defined through T in the analogous manner.

7.2 Algorithm comparison

Let us start with the analysis of mean square errors for the described above algorithms neglecting the term produced by the initial distribution. For $k=1$ we get by (6.9) and (6.15) that

$$\mathbf{E}\Big(\psi(\xi_n(t))-\psi(\mu_t(\mu))\Big)^2=\frac{1}{n}V(\psi,t,\mu)+o(1/n) \tag{7.13}$$

with $\psi\in\mathbf{C}^2(H)$, $\mu\in H$,

$$V(\psi,t,\mu)=\int\limits_0^t s_1^2(\psi,\tau,\mu_{t-\tau}(\mu))d\tau, \tag{7.14}$$

$$s_1^2(\psi,\tau,\nu)=\int\limits_{D^2} f_1^2(\tau,\nu;v,u)B_1(dv;u;\nu)\nu(du) \tag{7.15}$$

and

$$f_1(\tau,\nu;v,u)=\frac{d\psi(\mu_\tau(\nu))}{d\nu}(\delta_v-\delta_u).$$

Note that by definition $f_1(\tau,\nu;v,u)=-f_1(\tau,\nu;u,v)$ and

$$f_1(\tau,\nu;v,u)+f_1(\tau,\nu;u,z)=f_1(\tau,\nu;v,z).$$

If $k=2$, then we have by (6.16), (6.9), (6.10) and (6.11) that (7.13) is still valid here, (7.14) turns to

$$V(\psi,t,\mu)=\int\limits_0^t\Big(s_1^2(\psi,\tau,\mu_{t-\tau}(\mu))+\sigma(\psi,\tau,\mu_{t-\tau}(\mu))\Big)d\tau, \tag{7.16}$$

s_1^2 is defined by (7.15) with B_1 derived from B_2 by (7.12) and

$$\begin{aligned}&\sigma(\psi,\tau,\nu)\\&=\int\limits_{D^4} f_1(\tau,\nu;v_1,u_1)f_1(\tau,\nu;v_2,u_2)B_2(dv_1dv_2;u_1,u_2;\nu)\nu^{\otimes 2}(du_1du_2).\end{aligned} \tag{7.17}$$

Let us compare mean square errors for Nanbu and Bird-like algorithms. For this purpose, we denote the corresponding asymptotical variance coefficients (7.14) and (7.16) by $V_{\mathbf{N}}$, $V_{\mathbf{B}_{in}}$, $V_{\mathbf{B}_d}$ and $V_{\mathbf{B}_\varphi}$. Denotation $V_{\mathbf{B}}$ stands for (7.16) for any Bird-like algorithm. In the same manner we use denotations $\sigma_{\mathbf{B}}$, $\sigma_{\mathbf{B}_{in}}$ etc. Finally, we write $s_{\mathbf{N}}^2$ and $s_{\mathbf{S}}^2$ instead of s_1^2 for B_1 defined by (7.6) and $B_1=B$ correspondingly.

Lemma 7.2 *Let the hypotheses of Lemma* 7.1 *be fulfilled and the Boltzmann transformation function* $T(\,\cdot\,; u_1, u_2)$ *be invariant under permutation of* u_1 *and* u_2. *If* $\mu \in H$ *and* $\psi \in \mathbf{C}^2(H)$ *then for any* $t \geq 0$

$$V_{\mathbf{B}}(\psi, t, \mu) \leq 2V_{\mathbf{N}}(\psi, t, \mu), \qquad V_{\mathbf{N}}(\psi, t, \mu) \leq V_{\mathbf{B}_{in}}(\psi, t, \mu) \tag{7.18}$$

and $V_{\mathbf{B}_d}(\psi, t, \mu) \leq V_{\mathbf{B}_\varphi}(\psi, t, \mu)$ *for* $\varphi(u) \equiv u$.

Proof. As all Bird-like algorithms correspond to the Nanbu one (see (7.12)), functions (7.15) coincide with $s^2_{\mathbf{N}}$ for them, and their discrepances appear in (7.17). Let us fix τ and ν and delete them and ψ from denotation.

Consider independent ζ_1, ζ_2 with $\mathcal{L}(\zeta_1) = \mathcal{L}(\zeta_2) = \nu$. Suppose that the conditional distribution of η under the condition $\zeta_1 = u_1$, $\zeta_2 = u_2$ coincides with $T(\,\cdot\,; u_1, u_2)$ and under the same condition the mutual distribution of η_1, η_2 is equal to $B_2(\,\cdot\,; u_1, u_2; \nu)$. Then $\mathcal{L}(\eta \in \,\cdot \,|\, \zeta_1 = u_1) = B_1(\,\cdot\,; u_1; \nu)$ with B_1 defined by (7.6), and $\mathcal{L}(\eta_1, \zeta_1) = \mathcal{L}(\eta_2, \zeta_2) = \mathcal{L}(\eta, \zeta_1)$ due to (7.12) and the symmetrical features of B_2. Therefore $s^2_{\mathbf{N}} = \mathbf{E} f_1^2(\eta, \zeta_1)$, $\sigma_{\mathbf{B}} = \mathbf{E}\Big(f_1(\eta_1, \zeta_1) f_1(\eta_2, \zeta_2)\Big) \leq s^2_{\mathbf{N}}$ and

$$s_{\mathbf{N}} + \sigma_{\mathbf{B}} = \frac{1}{2}\mathbf{E}\Big(f_1(\eta_1, \zeta_1) + f_1(\eta_2, \zeta_2)\Big)^2$$

for any Bird-like algorithm. Thus the first inequality in (7.18) is proved.

Dealing with $\mathbf{B}_{in}$ we see from (7.8) that the random vectors (η_1, ζ_1) and (η_2, ζ_2) are independent and therefore $\sigma_{\mathbf{B}_{in}} \geq 0$. This proves the second inequality in (7.18). Finally, we get by the derivative properties that

$$\begin{aligned}
\sigma_{\mathbf{B}_\varphi} &= \mathbf{E}\Big(f_1(\eta_1, \zeta_1) f_1(\varphi(\eta_1), \zeta_2)\Big) \\
&= \mathbf{E}\Big(f_1(\eta_1, \zeta_1) f_1(\varphi(\eta_1), \zeta_1)\Big) + \mathbf{E}\Big(f_1(\eta_1, \zeta_1) f_1(\zeta_1, \zeta_2)\Big)
\end{aligned}$$

and, analogously, $\sigma_{\mathbf{B}_d} = \mathbf{E}\Big(f_1(\eta_1, \zeta_1) f_1(\eta_2, \zeta_1)\Big) + \mathbf{E}\Big(f_1(\eta_1, \zeta_1) f_1(\zeta_1, \zeta_2)\Big)$, where the φ is described in (7.10) for $\mathbf{B}_\varphi$ and the conditional distribution B_2 is defined by (7.9). It may be easily seen that

$$\begin{aligned}
\sigma_{\mathbf{B}_\varphi} - \sigma_{\mathbf{B}_d} &= \mathbf{E}\Big(f_1(\eta_1, \zeta_1) f_1(\varphi(\eta_1), \zeta_1)\Big) - \mathbf{E}\Big(f_1(\eta_1, \zeta_1) f_1(\eta_2, \zeta_1)\Big) \\
&= \int\limits_{D^2} \operatorname{cov}\Big(f_1(\eta_1, \zeta_1), f_1(\varphi(\eta_1), \zeta_1) \,|\, \zeta_1 = u_1, \zeta_2 = u_2\Big) \nu(du_1)\nu(du_2).
\end{aligned}$$

Surely, this conditional covariation is positive if $\varphi(u) \equiv u$. $\square$

As for bias, it is much more difficult to compare the algorithms from this viewpoint. Nevertheless we display here all formulas for future discussions.

We have (see Theorem 6.4 and Remark 6.3) that under denotation

$$\Phi_1(\tau,\nu;\lambda) = \frac{d\psi(\mu_\tau(\nu))}{d\nu}\,(\lambda) \quad \text{with } \nu\in H \text{ and } \lambda\in Z,$$

$$f_2(\tau,\nu;v_1,u_1;v_2,u_2) = \frac{d^2\psi(\mu_\tau(\nu))}{d\nu^2}\,(\delta_{v_1}-\delta_{u_1},\delta_{v_2}-\delta_{u_2}),$$

and

$$\widetilde{B}(dv;\nu) = \int\limits_D B_2(dv\times D;u,u;\nu)\nu(du),$$

one gets for both $(n,1)$-algorithms the identity

$$\mathbf{E}\psi(\xi(t)) - \psi(\mu_t) = \frac{1}{n}\int\limits_0^t m_1(\psi,\tau,\mu_{t-\tau})d\tau + o(1/n)$$

with $\mu_t = \mu_t(\mu)$ and

$$m_1(\psi,\tau,\nu) = \frac{1}{2}\int\limits_{D^2} f_2(\tau,\nu;v,u;v,u)B_1(dv;u;\nu)\nu(du). \tag{7.19}$$

Here $B_1 = B$ for the **S**-algorithm and B_1 is defined by (7.6) for the Nanbu one.

For the Bird-like algorithms, we get that up to order $o(1/n)$

$$\mathbf{E}\psi(\xi(t)) - \psi(\mu_t) = \frac{1}{n}\int\limits_0^t \Big(m_2(\psi,\tau,\mu_{t-\tau}) + \alpha_2(\psi,\tau,\mu_{t-\tau})\Big)d\tau, \tag{7.20}$$

where

$$\begin{aligned} m_2(\psi,\tau,\nu) &= m_1(\psi,\tau,\nu) \\ &\quad + \frac{1}{2}\int\limits_{D^4} f_2(\tau,\nu;v_1,u_1;v_2,u_2)B_2(dv_1dv_2;u_1,u_2;\nu)\nu^{\otimes 2}(du_1du_2) \end{aligned} \tag{7.21}$$

and $\alpha_2(\psi,\tau,\nu) = \Phi_1(\tau,\nu;B-\widetilde{B}_2)$ with

$$\widetilde{B}_2(\mu) = \int\limits_D B_2(\,\cdot\times D;u,u;\mu)\mu(du).$$

Note that m_1 is defined here by the Nanbu variant of (7.19), whereas B_2 and $\widetilde{B}_2$ depend on the concrete variant of the Bird-like algorithm. In particular, $\alpha_2 \equiv 0$ for the $\mathbf{B}_{in}$.

Now let us turn to the complexity of the algorithms under consideration. As before, we do not take into account the part of the mean square error corresponding to simulation of the initial distribution, supposing that the proper stratification technique is used to delete it.

Let t, μ and ψ be fixed, $\lambda_n^{(k)} = n/k$, and suppose that the conditions of Theorem 6.4 on the (n,k)-particle processes $\xi_n(t)$ are fulfilled. By Theorem 6.4, to obtain the identity $\mathbf{E}(\psi(\xi_n(t)) - \psi(\mu_t))^2 = \varepsilon > 0$, we must take the number $n = n(\varepsilon)$ of particles as

$$n(\varepsilon) = \frac{1}{\varepsilon}\int_0^t s_k^2(\psi, \tau, \mu_{t-\tau}(\mu))d\tau + o(1/\varepsilon). \tag{7.22}$$

Now, the average number of jumps up to time t is equal to nt/k, and it is natural to suppose that the main part of the jump simulation complexity corresponds to the simulation of the Boltzmann transformation function T. Therefore we define the *complexity* of the algorithm by

$$L(t) = \frac{\mathfrak{T}t}{k}\int_0^t s_k^2(\psi, \tau, \mu_{t-\tau}(\mu))d\tau,$$

where $\mathfrak{T}$ is the complexity of the unique jump simulation. Assuming that φ is a rather simple function from the computational viewpoint, we get that $\mathfrak{T}_{\mathbf{B}_{in}} \approx \mathfrak{T}_{\mathbf{B}_d} \approx 2\mathfrak{T}_{\mathbf{N}}$ and $\mathfrak{T}_{\mathbf{B}_\varphi} \approx \mathfrak{T}_{\mathbf{N}}$. Therefore $L_{\mathbf{B}_\varphi}(t) \le L_{\mathbf{N}}(t) \le L_{\mathbf{B}_{in}}(t)$ by Lemma 7.2, and, asymptotically, the Bird-like algorithm $\mathbf{B}_\varphi$ is not worse than the Nanbu one.

As for the comparison of the $\mathbf{N}$ and $\mathbf{S}$ algorithms, the examples for linear equations show (see Section 8) that there is no universal advantage of one of them over the other.

7.3 Random number of colliding particles

Now let us demonstrate that the technique proposed in the previous sections may be applied to algorithms with a random number of colliding particles (all above results are valid for the fixed k). As before, we consider the Boltzmann-like equation (7.3), satisfying the conditions of Lemma 7.1.

Let T_i be defined by (7.7) and $0 < p_n < 1$. Consider a jump empirical process $\xi_n(t)$ with the infinitesimal characteristics $\lambda_n(\nu) \equiv n$ and

$$\pi_n(\nu, \mathcal{D}) = \int\limits_{Q^{-1}\mathcal{D}} \sum_{k=0}^{n} b(k; n, p_n) \frac{1}{A_n^k} \sum_{J_k} \bigotimes_{l=1}^{k} T_{i_l}(dv_{i_l}; \mathbf{u}) \bigotimes_{j \notin \{i_1,\ldots,i_k\}} \delta_{u_j}(dv_j),$$

where $\mathbf{u} \in D^n$, $\nu = Q\mathbf{u}$ and $b(k; n, p_n) = C_n^k (1-p_n)^{n-k} p_n^k$, $C_n^k = n!/k!(n-k)!$.

A jump of such a process may be described as follows: each of n particles gets a new position with the probability p_n independently of the others; the new coordinate of the i-th particle has the distribution T_i.

Theorem 7.1 *Let $np_n = 1 + \mathfrak{C}/n + o(1/n)$ and $\psi \in \mathbf{C}^2(H)$. Then the bias formulas* (7.19) - (7.21) *and the variance formulas* (7.13), (7.15) - (7.17) *hold with $\alpha_2(\psi, \tau, \nu) = \mathfrak{C}\Phi_1(\tau, \nu; G(\nu))$ and B_1, B_2 defined by* (7.6) *and* (7.8).

Proof. To demonstrate the assertion we must retrace the same way as when proving Theorem 6.4. Here we outline a sketch of this proof, starting with Theorem 4.1. First, the condition (4.8) is fulfilled as

$$\beta_n = \sup_{\nu \in H_n} n\mathbf{E}\|\theta_\nu^{(n)}\|^2 = \mathbf{E}\kappa_n^2/n,$$

where the random variable κ_n has the Binomial distribution with the parameters n and $p = p_n$. Thus $\mathbf{E}\kappa_n^2 = O(1)$ and $\beta_n = O(1/n)$.

Second, we must check the condition (4.7). Let ω_i and η_i, $1 \le i \le n$, be independent random variables with $\mathbf{P}(\omega_i = 1) = 1 - \mathbf{P}(\omega_i = 0) = p_n$ and $\mathcal{L}(\eta_i) = T_i(\,\cdot\,; \mathbf{u})$. For $\mathbf{u} = (u_1, \ldots, u_n)$ and $\nu = Q\mathbf{u}$ we get that

$$\mathcal{L}\left(\theta_\nu^{(n)}\right) = \mathcal{L}\left(\frac{1}{n}\sum_{i=1}^{n} \omega_i(\delta_{\eta_i} - \delta_{u_i})\right).$$

Therefore $G_n(\nu) = n\mathcal{E}(\theta_\nu^{(n)}) = np_n G(\nu)$ and $G_n(\nu) - G(\nu) = \mathfrak{C}G(\nu)/n$ up to order $o(1/n)$. Thus the assumptions of Theorem 4.1 are fulfilled. The rest is in a sense standard: the identity (3.9) being proved, we sequentially demonstrate assertions similar to the second proposition of Theorem 5.1, the first proposition of Theorem 5.3 and, finally, Theorem 6.4. $\square$

Remark 7.2 *Theorem* 7.1 *shows that the considered algorithm is asymptotically equivalent to the $\mathbf{B}_{in}$-algorithm from the viewpoint of its variance and complexity. Therefore its complexity is greater than that of the Nanbu one.*

7.4 Locally compact D

All previous results are valid for a compact space D. Here we extend them to locally compact spaces.

Let D be a metric complete separable locally compact space and let $\overline{D} = D \cup \{\infty\}$ stand for its one-point compactification. Denote by $\mathbf{C}_0(D)$ the set of all continuous functions vanishing at ∞, by $H = H(D)$ the space of all probability measures and by $H_0^1 = H_0^1(D)$ the space of all substochastical measures on $(D, \mathcal{B})$. The so-called *vague-convergence* ($\mu_m \xrightarrow{v} \mu$, if $\int g d\mu_m \to \int g d\mu$ for any $g \in \mathbf{C}_0(D)$) induces the *vague topology* $\mathcal{F}_v$ on H_0^1.

Consider a function $\psi : H \mapsto \mathbf{R}$ such that it may be extended to a v-continuous function $H_0^1 \mapsto \mathbf{R}$, which has m strong v-continuous derivatives, and denote a set of such functions by $\mathbf{C}_v^m(H)$. Evidently, if $g \in \mathbf{C}_0(D)$ then the linear function $\psi(\nu) = \int g d\nu$ belongs to $\mathbf{C}_v^m(H)$ for any m.

We deal with the defined on H Boltzmann-like equation (7.3), its invariant Boltzmann transformation function T, and the sequence of (n,k)-particle processes with $\lambda_n^{(k)} = n/k$, $p_{i_1 \dots i_k}(\mathbf{u}) = 1/A_n^k$ and

$$T_{i_1 \dots i_k}(\,\cdot\,; \mathbf{u}) = B_k(\,\cdot\,; u_{i_1}, \dots, u_{i_k}; Q\mathbf{u}).$$

As before we assume that $B_k(C_1 \times \dots \times C_k; u_1, \dots, u_k; \mu)$ is invariant under any simultaneous permutation of $C_1, \dots, C_k$ and $u_1, \dots, u_k$ for any $\mu \in H$. For the simplicity we suppose that $\widehat{B}_k = B$ (see (5.16)).

We call the Boltzmann transformation function *v-extendible*, if for any $\phi \in \mathbf{C}_0(D)$ the function $h(u_1, u_2) = \int_D \phi(v) T(dv; u_1, u_2)$ is continuous in (u_1, u_2), and $h_1(u_1) = \sup_{u_2} |h(u_1, u_2)| \underset{u_1 \to \infty}{\longrightarrow} 0$.

The (n,k)-particle process is said to be *v-extendible*, if there exists a mapping $B_k' : D^k \times H_0^1 \mapsto H_0^1$ such that $B_k'(\mathbf{u}; \nu) = B_k(\,\cdot\,; \mathbf{u}; \nu)$ for any $\mathbf{u} \in D^k$ and $\nu \in H$, the function

$$h(\mathbf{u}; \nu) = \int\limits_{D^k} \phi(v_1, \dots, v_k) B_k'(dv_1 \dots dv_k; \mathbf{u}; \nu)$$

is continuous in $\mathbf{u}$ and v-continuous in $\nu \in H_0^1$ for any $\phi \in \mathbf{C}_0(D^k)$ and

$$\sup_{\nu \in H_0^1,\, u_j : j \neq i} |h(\mathbf{u}; \nu)| \underset{u_i \to \infty}{\longrightarrow} 0, \quad 1 \le i \le k.$$

As B_k' is a sort of extension of B_k, we use the same denotation B_k for it.

Theorem 7.2 *Let $\psi \in \mathbf{C}_v^3(H)$ and suppose that $\xi_n = \xi_n(0)$ satisfy the hypotheses of Theorem 6.2. If the Boltzmann transformation function T and the sequence of (n,k)-particle processes $\xi_n(t)$ are v-extendible, then both conclusions of Theorem 6.4 are valid.*

Proof. Denote by $\overline{\mathcal{B}}$ the Borel σ-algebra of the one-point compactification $\overline{D}$, and by $H(\overline{D})$ the set of all probability measures on $(\overline{D}, \overline{\mathcal{B}})$.

Let us put $T(\,\cdot\,; u_1, \infty) = T(\,\cdot\,; \infty, u_2) = \delta_\infty$ for $u_1, u_2 \in \overline{D}$ and extend the Boltzmann-like equation (7.3) to an equation defined on $H(\overline{D})$ supposing that its right-hand side G has the same form (7.3) for $\mu \in H_0^1(D)$ and is determined by the gomeomorphism $(H_0^1, \mathcal{F}_v) \mapsto (H(\overline{D}), \mathcal{F}_w)$ for $\mu \in H(\overline{D})$. Then G satisfies conditions $A0 - A5$ with $H = H(\overline{D})$, and $\mu_t(\mu) \in H(D) = H(\overline{D})|_D$ for $\mu \in H(D)$.

We extend the (n,k)-particle process with the phase space $H(D^n)$ to a (n,k)-particle process with the phase space $H(\overline{D}^n)$. For this purpose we put

$$B_k(dv_1 \dots dv_k; \infty, u_2, \dots, u_k; \nu) = \bigotimes_{i=1}^{k} \delta_\infty(dv_i)$$

for $u_2, \dots, u_k \in \overline{D}$, $\nu \in H(D)$ and assume that $B_k(\,\cdot\,; \mathbf{u}; \nu)$ is determined by the gomeomorphism mentioned above, if $\nu \in H(\overline{D})$. Then the extended process $\overline{\xi}_n(t)$ has w-continuous on $H(\overline{D})$ infinitesimal characteristics and $\mathbf{P}_\nu(\overline{\xi}_n(t) \in H(D)) = 1$ for $\nu \in H(D)$.

Due to Theorem 6.4, we must prove that $\widehat{B}_k(\nu) = B(\nu)$ for any $\nu \in H(\overline{D})$. But this identity is valid on $H(D)$ by the assumption and may be extended to $H(\overline{D})$ as $H(D)$ is weakly dense in $H(\overline{D})$ and measures $\widehat{B}_k$, B are w-continuous on $H(\overline{D})$. All other reasonings are standard. $\square$

Remark 7.3 *Let T be defined by (7.2). If $\widetilde{T}$ is v-extendible, $\sigma \in \mathbf{C}(D^2)$ and $\sup_{u_1} \sigma(u_1, u_2) \underset{u_2 \to \infty}{\longrightarrow} \sup_{u_1, u_2} \sigma(u_1, u_2) < \infty$, then T is v-extendible.*

Note that if T is v-extendible, then both Bird-like and Nanbu-like processes of Section 7.1 satisfy the Theorem 7.2 conditions. To demonstrate this assertion, it is sufficient to define the corresponding measures $B_1(\,\cdot\,; u; \nu)$ and $B_2(\,\cdot\,; u_1, u_2; \nu)$ for $\nu \in H_0^1$ by (7.6), (7.8), (7.9) and (7.11).

The following statement shows that the well-known Boltzmann transformation function of rarefied gas dynamics is v-extendible. Denote by $\mathfrak{S}_{a,r}$ the uniform distribution on the sphere $S(a,r)$ of the radius r with the center $a \in \mathbf{R}^3$. For $u_1, u_2 \in \mathbf{R}^3$ we put $u_+ = (u_1 + u_2)/2$ and $u_- = (u_1 - u_2)/2$.

Lemma 7.3 *Let* $D = \mathbb{R}^3$*. Then* $T(\,\cdot\,; u_1, u_2) = \mathfrak{S}_{u_+,|u_-|}$ *is v-extendible.*

Proof. Let $\phi \in \mathbf{C}_0(D)$ and $h(u_1, u_2) = \int \phi(v) T(dv, u_1, u_2)$. Then, evidently, $h \in \mathbf{C}(D^2)$ and

$$h(u_1, u_2) = \int\limits_{\mathbf{R}^3} \phi\Big(u_+ + \omega |u_-|\Big) \mathfrak{S}_{0,1}(d\omega).$$

As $\phi \in \mathbf{C}_0(D)$, then for any $\varepsilon > 0$ there exists $R > 0$ such that $|\phi(u)| < \varepsilon$ for $|u| > R$. Therefore under denotation $M = \sup\limits_{|w| \le R} |\phi(w)|$,

$$|h_2(u_2)| \le M \sup_{u_1} \int\limits_{|w|<R} \mathfrak{S}_{u_+,|u_-|}(dw) + \varepsilon. \tag{7.23}$$

Now it is sufficient to prove that the first term on the right-hand side of (7.23) tends to zero as $|u_2| \to \infty$. It is easy to see that for any fixed u_2 the corresponding integral reaches its maximum value at $u_1 = \mathbf{0}$.

Let $|u_2| > 2R$ and $u_1 = \mathbf{0}$. Put $\mathcal{S} = S(u_+, |u_-|) \cap \{w : |w| \le R\}$ and denote by $\alpha = \alpha(\mathcal{S})$ the angle of this sector. Then

$$\cos\alpha = \frac{2|u_2|^2 - 4R^2}{2|u_2|^2} \underset{u_2 \to \infty}{\longrightarrow} 1,$$

and the proof is complete. □

Let $\varphi(u; u_1, u_2) = u_1 + u_2 - u$. Surely, such φ satisfies (7.10) with the Boltzmann transformation function of Lemma 7.3, and the corresponding $\mathbf{B}_\varphi$-algorithm is the standard one in the Monte Carlo simulation of the rarefied gas dynamics problems. Moreover, this transformation function satisfies the condition $T(\,\cdot\,; u, u) = \delta_u$. Therefore, due to Remark 5.3, if we take $\lambda_n^{(2)} = (n-1)/2$ instead of $\lambda_n^{(2)} = n/2$ in $\mathbf{B}_\varphi$-algorithm, then the addend α_2 in (7.20) turns to zero. Such a choice of $\lambda_n^{(2)}$ may be found in [13].

7.5 Example of the equilibrium convergence

Consider a Boltzmann-like equation

$$\frac{d\mu_t}{dt} = \int\limits_{D^2} T(\cdot\,; u_1, u_2)\mu_t(du_1)\mu_t(du_2) - \mu_t, \quad \mu_t|_{t=0} = \mu, \tag{7.24}$$

with $D = (0, 1]$, the metric $\rho(u, v) = \min(|u-v|, 1-|u-v|)$ and the Boltzmann transformation function $T(\,\cdot\,; u_1, u_2) = \delta_{u_1 \oplus u_2}$, where $\oplus$ stands for addition modulo 1. This equation is one of the simplest model kinetic equations and was studied in [2, 19]. Note that (D,ρ) corresponds to the circle of the unit length. As the mapping $(u_1,u_2)\mapsto u_1\oplus u_2$ is ρ-continuous, T is w-continuous.

Consider the linear function $\psi(\mu_t)=\int g d\mu_t$, $g\in \mathbf{C}(D)$. Our aim is to investigate the asymptotical $(t\to\infty)$ behaviour of main terms of the bias and the mean square error for the $\mathbf{N}$ and $\mathbf{B}_\varphi$ algorithms. As T is expressed through the Dirac measure, all $\mathbf{B}_\varphi$-algorithms coincide (and coincide with $\mathbf{B}_d$). Thus we can take $\varphi(u)\equiv u$. As above we suppose that the proper stratification is used when simulating the initial distribution $\mathcal{L}(\xi_n(0))$ and therefore $\xi_n(0)$ has no influence on the main terms of both errors. Denote

$$\mathcal{M}_\tau(dvdu_1du_2)=T(dv;u_1,u_2)\mu_\tau(du_1)\mu_\tau(du_2),$$

$$\mathcal{N}_\tau(dvdu_1du_2)=T(dv;u_1,u_1)\mu_\tau(du_1)\mu_\tau(du_2),$$

$\mathcal{M}_\tau^{(1)}(dvdu_1)=\mathcal{M}_\tau(dvdu_1\times D)$, $\mathcal{M}_\tau^{(2)}(dv)=\mathcal{M}_\tau(dv\times D^2)$, and $\mathcal{N}_\tau^{(2)}(dv)=\mathcal{N}_\tau(dv\times D^2)$. For the $\mathbf{N}$-algorithm (see (7.19) and (7.15)) we are interested in the asymptotical $(t\to\infty)$ behavior of functions

$$A_1(t)=\frac{1}{2}\int_0^t\int_{D^2} f_2(t-\tau,\mu_\tau;v,u_1,v;u_1,)\mathcal{M}_\tau^{(1)}(dvdu_1)\,d\tau$$

and

$$V_1(t)=\int_0^t\int_{D^2} f_1^2(t-\tau,\mu_\tau;v,u_1)\mathcal{M}_\tau^{(1)}(dvdu_1)\,d\tau. \tag{7.25}$$

Analogous functions of interest for the $\mathbf{B}_\varphi$-algorithm have the form $A_2(t)=A_1(t)+\mathcal{A}(t)+\mathfrak{A}(t)$ and $V_2(t)=V_1(t)+\mathfrak{V}(t)$ with

$$\mathcal{A}(t)=\int_0^t \Phi_1\Big(t-\tau,\mu_\tau;\mathcal{M}_\tau^{(2)}-\mathcal{N}_\tau^{(2)}\Big)\,d\tau,$$

$$\mathfrak{A}(t)=\frac{1}{2}\int_0^t\int_{D^3} f_2(t-\tau,\mu_\tau;v,u_1;v,u_2)\mathcal{M}_\tau(dvdu_1du_2)\,d\tau$$

and

$$\mathfrak{V}(t) = \int_0^t \int_{D^3} f_1(t-\tau, \mu_\tau; v, u_1) f_1(t-\tau, \mu_\tau; v, u_2) \mathcal{M}_\tau(dvdu_1du_2)\, d\tau \quad (7.26)$$

(see (7.20), (7.21), (7.16) and (7.17)). Let $\mathbf{D}\psi(\nu) = \int_0^1 g^2 d\nu - \psi^2(\nu)$.

Theorem 7.3 *Denote by* μ_∞ *the uniform distribution on* $(0,1]$ *and suppose that* $\|\mu - \mu_\infty\| = q < 1$. *Then* $\lim\limits_{t\to\infty} A_1(t) = \lim\limits_{t\to\infty} A_2(t) = 0$ *and*

$$\lim_{t\to\infty} V_1(t) = \mathbf{D}\psi(\mu_\infty), \qquad \lim_{t\to\infty} V_2(t) = \frac{3}{2}\,\mathbf{D}\psi(\mu_\infty). \quad (7.27)$$

Proof. Denote by $\nu_1 * \nu_2$ the convolution of signed measures ν_1 and ν_2 on the unit circle, ν^{*m} standing for m-times convolution of ν. Using peculiarities of the Boltzmann transformation function T and the property $\nu * \mu_\infty = \mu_\infty$, $\nu \in H$, we get by the Wild decomposition formula (e.g., [5, ch. 4]) that

$$\begin{aligned}\mu_t(\mu) &= e^{-t}\mu + e^{-t} \sum_{m=1}^{\infty} (1-e^{-t})^m \mu^{*(m+1)} \\ &= \mu_\infty + e^{-t} \sum_{m=0}^{\infty} (1-e^{-t})^m (\mu^{*(m+1)} - \mu_\infty). \end{aligned} \quad (7.28)$$

It follows from conditions on the initial data μ that $\|\mu^{*m} - \mu_\infty\| \le q^m$ and $\|\mu_t(\mu) - \mu_\infty\| \underset{t\to\infty}{\longrightarrow} 0$. Thus the uniform distribution μ_∞ is the equilibrium solution of (7.24). Denote for $l = 0, 1$

$$\mathfrak{B}_l(\tau, \mu) = \sum_{m=l}^{\infty} (1 - e^{-\tau})^m (\mu^{*(m+1)} - \mu_\infty).$$

Each of the following functions may be depicted as the sum of two terms: $\mu_\tau(\mu) = \mu_\infty + e^{-\tau}\mathfrak{B}_0(\tau, \mu)$,

$$f_1(\tau, \nu; v, u) = e^{-\tau}(g(v) - g(u)) + e^{-\tau}\psi\Big(\partial^1 \mathfrak{B}_1(\tau, \nu; \delta_v - \delta_u)\Big) \quad (7.29)$$

and

$$f_2(\tau, \nu; v_1, u_1; v_2, u_2) = 0 + e^{-\tau}\psi\Big(\partial^2 \mathfrak{B}_1(\tau, \nu; \delta_{v_1} - \delta_{u_1}; \delta_{v_2} - \delta_{u_2})\Big).$$

Later on we prove that, asymptotically, the second terms of these sums tend to zero. If we trust these assumptions, then we have that $A_1(t)\to 0$, $\mathfrak{A}(t)\to 0$,

$$\mathcal{A}(t) \to \mathcal{A}(\infty) = \int\limits_D g(v)\Big(\mathcal{M}_\infty^{(2)}(dv) - \mathcal{N}_\infty^{(2)}(dv)\Big), \tag{7.30}$$

$$V_1(t) \to V_1(\infty) = \frac{1}{2}\int\limits_{D^2} \Big(g(v) - g(u_1)\Big)^2 \mathcal{M}_\infty^{(1)}(dv du_1) \tag{7.31}$$

and

$$\mathfrak{V}(t) \to \mathfrak{V}(\infty) = \frac{1}{2}\int\limits_{D^3} \Big(g(v) - g(u_1)\Big)\Big(g(v) - g(u_2)\Big)\mathcal{M}_\infty(dv du_1 du_2), \tag{7.32}$$

where measures $\mathcal{M}_\infty$, $\mathcal{N}_\infty$, $\mathcal{M}_\infty^{(1)}$, etc, are defined in the same manner as $\mathcal{M}_\tau$, $\mathcal{N}_\tau$, $\mathcal{M}_\tau^{(1)}$, etc, with μ_τ replaced by μ_∞.

Let us calculate the right-hand side expressions in (7.30) – (7.32). If random variables α_1 and α_2 are independent and uniformly distributed on D, then $\beta \stackrel{def}{=} \alpha_1 \oplus \alpha_2$ and $\alpha_1^{\oplus 2}$ are also uniformly distributed on D and random variables β, α_1 are independent. Therefore $\mathcal{A}(\infty) = \mathbf{E}(g(\beta) - g(\alpha_1^{\oplus 2}) = 0$, $V_1(\infty) = 0.5\,\mathbf{E}\Big(g(\beta) - g(\alpha_1)\Big)^2 = \mathbf{D}g(\alpha_1)$ and

$$\mathfrak{V}(\infty) = \frac{1}{2}\,\mathbf{E}\Big(g(\beta) - g(\alpha_1)\Big)\Big(g(\beta) - g(\alpha_2)\Big) = \frac{1}{2}\,\mathbf{D}g(\alpha_1).$$

Note that these expressions correspond to the theorem conclusion. Now we must prove the correctness of limit transitions (7.30) – (7.32). We demonstrate this correctness dealing with $V_1(t)$, other cases may be treated in the same way. First, we prove that for any τ

$$\|\mathfrak{B}_0(\tau,\mu)\| \le C_0 \tag{7.33}$$

and

$$\sup_{t\ge\tau,\,\kappa\in Z_2} \|\partial^1 \mathfrak{B}_1(t-\tau,\mu_\tau;\kappa)\| \le C_1 e^{-\tau} \tag{7.34}$$

with $Z_2 = \{\nu \in E : \nu(D) = 0, \|\nu\| \le 2\}$ and certain $C_0, C_1 > 0$. Surely,

$$\|\mathfrak{B}_0(\tau,\mu)\| \le \sum_{m=0}^{\infty} (1-e^{-\tau})^m q^{m+1} = \frac{q}{1-(1-e^{-\tau})q} \le \frac{q}{1-q}$$

and (7.33) is proved. As $\mu_\infty * \kappa = \mathbf{0}$ for $\kappa \in Z_2$ and

$$\partial^1 \mathfrak{B}_1(t-\tau, \mu_\tau; \kappa) = \sum_{m=1}^{\infty} (m+1)\left(1-e^{-(t-\tau)}\right)^m \mu_\tau^{*m} * \kappa$$

$$= \sum_{m=1}^{\infty} (m+1)\left(1-e^{-(t-\tau)}\right)^m \left(\mu_\infty + e^{-\tau}\mathfrak{B}_0(\tau,\mu)\right)^{*m} * \kappa$$

$$= \sum_{m=1}^{\infty} (m+1)\left(1-e^{-(t-\tau)}\right)^m e^{-m\tau}\left(\mathfrak{B}_0(\tau,\mu)\right)^{*m} * \kappa,$$

we obtain for $\tau \le t$ that

$$\begin{aligned}
&\|\partial^1 \mathfrak{B}_1(t-\tau, \mu_\tau; \kappa)\| \\
&\quad \le \|\kappa\| \sum_{m=1}^{\infty} (m+1)\left(1-e^{-(t-\tau)}\right)^m e^{-m\tau}\left(\sum_{l=1}^{\infty}\left(1-e^{-\tau}\right)^l q^{l+1}\right)^m \\
&\quad \le \|\kappa\| \sum_{m=1}^{\infty} (m+1)\,\omega^m(\tau) \le 2\,\|\kappa\|\,\omega(\tau)\left(1-\omega(\tau)\right)^{-2},
\end{aligned}$$

where $\omega(\tau) = e^{-\tau}q/(e^{-\tau}q + 1 - q) < 1$. Therefore

$$\|\partial^1 \mathfrak{B}_1(t-\tau, \mu_\tau; \kappa)\| \le \|\kappa\| \frac{e^{-\tau}q(e^{-\tau}q+1-q)}{(1-q)^2}$$

and the inequality (7.34) is proved too.

Denote $\mathfrak{G}(\tau,\mu;du) = e^{-\tau}(2\mu_\infty(du) + \mathfrak{B}_0(\tau,\mu;du))$ and $\Delta_1(\psi,\tau,\nu,v,u) = \psi\left(\partial^1\mathfrak{B}_1(\tau,\nu;\delta_v-\delta_u)\right)$. By (7.25), (7.26) and (7.29) we obtain that

$$\begin{aligned}
V_1(t) - V_1(\infty) = &\int_0^t \int_{D^2} e^{-2(t-\tau)}\left(g(u_1\oplus u_2) - g(u_1)\right)^2 \mathfrak{G}(\tau,\mu;du_1)\mu_\infty(du_2) \\
&+2\int_0^t \int_{D^2} e^{-2(t-\tau)}\left(g(v)-g(u_1)\right)\Delta_1(\psi, t-\tau, \mu_\tau, v, u_1)\mathcal{M}_\tau^{(1)}(dv du_1)\,d\tau \\
&+\int_0^t \int_{D^2} e^{-2(t-\tau)}\Delta_1(\psi^2, t-\tau, \mu_\tau, v, u_1)\mathcal{M}_\tau^{(1)}(dv du_1)\,d\tau.
\end{aligned}$$

Therefore by (7.33) and (7.34) we get that

$$|V_1(t) - V_1(\infty)| \\ \leq 4\Gamma^2 \int\limits_0^t e^{-2(t-\tau)}\Big(2e^{-\tau}C_0 + e^{-2\tau}C_0^2 + 2C_1e^{-\tau} + C_1^2e^{-2\tau}\Big)\,d\tau \underset{t\to\infty}{\longrightarrow} 0$$

with $|g(v)| \leq \Gamma$. Convergences $\mathfrak{V}(t) \to \mathfrak{V}(\infty)$ and $\mathcal{A}(t) \to \mathcal{A}(\infty)$ may be proved similarly. The inequality $\sup\limits_{t\geq\tau,\,\kappa_1,\kappa_2\in Z_2} \|\partial^2\mathfrak{B}_1(t-\tau,\mu_\tau;\kappa_1,\kappa_2)\| \leq C_2e^{-\tau}$ is demonstrated in the same way as (7.34) and therefore the convergence $A_1(t)\to A_1(\infty)$ is also valid as well as the convergence $\mathfrak{A}(t)\to\mathfrak{A}(\infty)$. $\square$

Remark 7.4 *This example shows that the variance coefficient for a Bird-like algorithm $\mathbf{B}_\varphi$ may be greater than that for a Nanbu one. On the other hand, $\mathbf{B}_\varphi$ is better from the complexity viewpoint.*

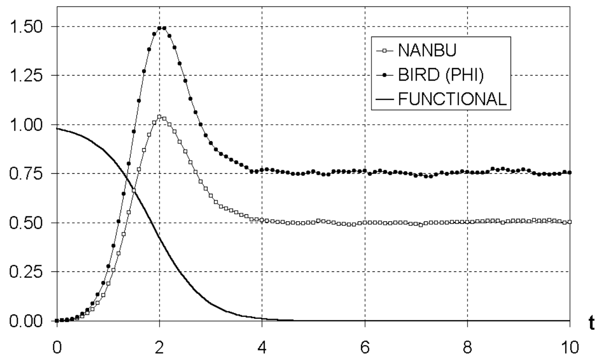

Figure 1: Variance coefficients convergence.

Simulation experiments confirm these results. If we take μ as the uniform distribution on $(0, 0.1]$ and $g(u) = \cos(2\pi u)$, then $\mathbf{D}\psi(\mu_\infty) = 0.5$. The Figure 1, where estimated asymptotical variance coefficients for $\mathbf{N}$ and $\mathbf{B}_\varphi$ algorithms (with a stratification) are presented, shows that (7.27) really holds. The third graph depicts estimated values of $\psi(\mu_t)$. It is remarkable that all tree functions reach their equilibrium at the same time (about $t = 4$). For the simulation, $n = 100$ was taken and 10000 independent trajectories of n-particle processes were used to obtain estimates.

8 Linear equations

Consider the linear balance equation for measures, i.e. the equation

$$\frac{d\mu_t}{dt} = \int_D T(\,\cdot\,;u)\mu_t(du) - \mu_t = B(\mu_t) - \mu_t, \quad \mu_t|_{t=0} = \mu \in H. \tag{8.1}$$

As above we suppose that the *linear transformation function* T is w-continuous and $T(\,\cdot\,;u) \in H$ for any $u \in D$. The equation (8.1) is the particular case of a Boltzmann-like equation, and (due to Lemma 7.1) we see that all general results on the stochastic solution for balance equations are valid here.

We suppose that the functional of interest $\psi(\mu_t)$ is also linear, that is, $\psi(\mu_t) = \int_D g d\mu_t$, $g \in \mathbf{C}(D)$. If we define a linear operator $B : E \mapsto E$ by the same formula $B\nu = B(\nu) = \int_D T(\,\cdot\,;u)\nu(du)$, then the solution $\mu_t = \mu_t(\mu)$ of (8.1) may by expressed through a series:

$$\mu_t(\mu) = \sum_{k=0}^{\infty} \frac{t^k}{k!}\, e^{-t}\; B^k\mu. \tag{8.2}$$

Note that one may consider the equation (8.1) as defined on E. Then (8.2) is valid for $\mu \in E$. If we put $\mu_t(\,\cdot\,;v) \stackrel{def}{=} \mu_t(\delta_v)$ (this measure-valued function may be called the *fundamental solution* of (8.1)) then, evidently, $\mu_t(\nu) = \int_D \mu_t(\,\cdot\,;v)\nu(dv)$, $\nu \in E$.

As $\mu_t(\mu)$ is linear in μ, then $\partial^1\mu_t(\mu;\nu) = \mu_t(\nu)$ and $\partial^k\mu_t \equiv \mathbf{0}$ for $k > 1$. Moreover, derivatives f_1, Φ_1 and f_2 have a rather simple form:

$$f_1(\tau,\nu;v,u) = \frac{d\psi(\mu_\tau(\nu))}{d\nu}\,(\delta_v - \delta_u) = \psi(\mu_\tau(\delta_v)) - \psi(\mu_\tau(\delta_u)), \tag{8.3}$$

$$\Phi_1(\tau,\nu;\lambda) = \frac{d\psi(\mu_\tau(\nu))}{d\nu}\,(\lambda) = \int_D g\; d\mu_\tau(\lambda) = \psi(\mu_\tau(\lambda))$$

and $f_2 \equiv 0$. Denote $\psi(\tau,v) = \psi(\mu_\tau(\delta_v))$.

It is well-known that (8.1) corresponds to the jump Markov process $\zeta(t)$ with $\mathcal{L}(\zeta(0)) = \mu$, the jump law T and exponentially distributed (with the unit scale parameter) time intervals between jumps. Then $\mathcal{L}(\zeta(t)) = \mu_t(\mu)$. The process $\zeta(t)$ is called a *basic* process.

The traditional procedure for the estimation of $\psi(\mu_t)$ is evident: one simulates n independent copies $\zeta_i(t)$ of the basic process $\zeta(t)$ and takes

$$\widehat{\psi}_n(t) = \frac{1}{n}\sum_{i=1}^{n} g(\zeta_i(t)) \tag{8.4}$$

to estimate $\psi(\mu_t)$. Surely, $\mathbf{E}\widehat{\psi}_n(t) = \psi(\mu_t)$ and $\mathbf{D}\widehat{\psi}_n(t) = n^{-1}\,\mathbf{D}g(\zeta(t))$.

This procedure may be considered from the other viewpoint. If we gather up processes $\zeta_i(t)$ in the "n-dimensional" process $\eta^{(n)}(t) = (\zeta_1(t), \ldots, \zeta_n(t))$, then $\eta^{(n)}(t)$ is also a jump Markov process. Its jump rate is equal to n, and the jump law may be described as follows. If $\mathbf{u} = (u_1, \ldots, u_n)$ is the phase coordinate of the process before a jump, we randomly choose one of the particles, and supposing this particle has a number i, it jumps to a new position due to the distribution $T(\,\cdot\,; u_i)$. Surely, $\mathcal{L}(\eta^{(n)}(0)) = \mu^{\otimes n}$.

This process is the $(n,1)$-process with $B_1(\,\cdot\,; u, \mu) = T(\,\cdot\,; u)$, $\lambda_n^{(1)} = n$, and $p_i(\mathbf{u}) = 1/n$. Moreover, (8.4) may be rewritten as $\psi(\xi_n(t))$, where $\xi_n(t)$ is the measure-valued empirical process, corresponding to $\eta^{(n)}(t)$. This means that the usual Monte Carlo algorithm for the solution of the linear equation (8.1) may be represented in terms of the (n,k)-particle processes, considered in the previous sections. We call this process a *traditional* one and use the term **T**-*algorithm* for the corresponding simulation procedure.

On the other hand, as the whole problem is linear, all general expressions for the bias and variance become rather simple, and one can calculate them more or less explicitly. Therefore two problems are to be solved. First, it is interesting to investigate the role of simulation of an initial distribution here: what profit would we have using the proper stratification instead of the usual independent simulation?

The second problem concerns the transition mechanism of a process. Dealing with Boltzmann-like equations, we see that Bird-like algorithms $\mathbf{B}_\varphi$ (that is, special $(n,2)$-particle processes with a simple φ) have the advantage over the Nanbu-like $(n,1)$-particle processes in the sense of their complexity. But the traditional **T**-algorithm for a linear equation has the same structure as the **N**-algorithm for a Boltzmann-like equation, and there is hope to construct in the linear case an *artificial* $(n,2)$-particle process with better properties than that of the traditional one.

Let us start with the variance term, corresponding to the initial distribution. If $\xi_n(0)$ appears from the simulation of n independent random variables $\zeta_i(0)$ then, by (6.5) and the linearity of the problem, the corresponding term $M_0(\psi, t, \mu)$ of the bias is equal to zero and (by (6.4), (6.6) and (8.3)) the variance term has a form

$$S_0^2(\psi, t, \mu) = \frac{V_0(\psi, t, \mu)}{n} = \frac{1}{n}\left(\int_D \psi^2(t, u)\mu(du) - \psi^2(\mu_t(\mu))\right). \tag{8.5}$$

As $\psi(t, u) = \mathbf{E}(g(\zeta(t)) \,|\, \zeta(0) = u)$, (8.5) has a simple interpretation in terms of the basic process $\zeta(t)$. Surely, $S_0^2(\psi,t,\mu)=n^{-1}\mathbf{D}\Big(\mathbf{E}\Big(g(\zeta(t))\,|\,\zeta(0)\Big)\Big)$ and by the standard variance decomposition formula,

$$\mathbf{D}g(\zeta(t))=\mathbf{D}\Big(\mathbf{E}\Big(g(\zeta(t))\,|\,\zeta(0)\Big)\Big)+\mathbf{E}\Big(\mathbf{D}\Big(g(\zeta(t))\,|\,\zeta(0)\Big)\Big),$$

where $\mathbf{D}(\eta\,|\,\zeta)$ stands for $\mathbf{E}(\eta^2\,|\,\zeta)-(\mathbf{E}(\eta\,|\,\zeta))^2$. Therefore, dealing with the traditional $\mathbf{T}$-algorithm and using the proper stratification, we diminish (up to order $o(n^{-1})$) the variance of the estimate (8.4) from $n^{-1}\mathbf{D}g(\zeta(t))$ to

$$n^{-1}\,\mathbf{E}\Big(\mathbf{D}\Big(g(\zeta(t))\,|\,\zeta(0)\Big)\Big)=\frac{1}{n}\Bigg(\int\limits_D g^2 d\mu_t-\int\limits_D \psi^2(t,u)\mu(du)\Bigg). \tag{8.6}$$

This expression corresponds to (6.16), i.e. to the part of the variance, connected with the transition mechanism of the $\mathbf{T}$-process. Note that (8.5) does not depend on this mechanism and is the same for all (n,k)-particle processes.

Now let us analyze a pair of such processes. The first is an $(n,1)$-particle process with $\lambda_n^{(1)}=n$, $p_i(\mathbf{u})=1/n$ and $B_1(\,\cdot\,;u;\nu)=B(\nu)=\int T(\,\cdot\,;v)\nu(dv)$, that is, the *simplest* process. Its transformation function has a form

$$T_i(\,\cdot\,;\mathbf{u})=\frac{1}{n}\sum_{k=1}^{n}T(\,\cdot\,;u_k),\quad \mathbf{u}=(u_1,\ldots,u_n).$$

The jump simulation may seem to be a bit non-usual here: having n particles with coordinates $(u_1,\ldots,u_n)$ before a jump, we randomly choose two particles with coordinates u_i,u_j (i may coincide with j), and the first of them moves to a new position due to $T(\,\cdot\,;u_j)$. As above we use the abbreviation $\mathbf{S}$ for the simplest processes and their algorithms.

Denote by $V_\mathbf{S}$ the variance term (7.13) for our process, ignoring, as before, the other term of the variance, corresponding to the initial value $\xi_n(0)$. Then

$$V_\mathbf{S}(\psi,t,\mu)=\int\limits_0^t s_\mathbf{S}^2(\psi,\tau,\mu_{t-\tau}(\mu))d\tau,$$

$$s_\mathbf{S}^2(\psi,\tau,\nu)=\int\limits_{D^3}\Big(\psi(\tau,v)-\psi(\tau,u)\Big)^2 T(dv;u_1)\nu(du_1)\nu(du).$$

By the same way we get for the traditional process that

$$s_\mathbf{T}^2(\psi,\tau,\nu)=\int\limits_{D^2}\Big(\psi(\tau,v)-\psi(\tau,u)\Big)^2 T(dv;u)\nu(du)$$

and (in terms of the basic process $\zeta(t)$)

$$
\begin{aligned}
&V_{\mathbf{S}}(\psi, t, \mu) - V_{\mathbf{T}}(\psi, t, \mu) \\
&\quad = 2\int\limits_0^t \operatorname{cov}\Big(\psi(t-\tau, \zeta(\tau)), \int\limits_D \psi(t-\tau, v) T(dv; \zeta(\tau))\Big) d\tau.
\end{aligned}
$$

Surely, the sign of this difference depends on the concrete T, g and μ. For example, as $\psi(0, u) = g(u)$, we have for a small t that

$$
V_{\mathbf{S}}(\psi, t, \mu) - V_{\mathbf{T}}(\psi, t, \mu) = 2t \operatorname{cov}\Big(g(\zeta(0)), \int\limits_D g(v) T(dv; \zeta(0))\Big) + o(t),
$$

where the main term may be both positive or negative. Therefore there is no universal advantage of $\mathbf{T}$-algorithms over $\mathbf{S}$-algorithms and vice versa.

Surely, one can use different $(n, 2)$-particle processes to solve the linear equation (8.1). For this purpose we take $\lambda_n^{(2)} = n/2$, $p_{ij} = 1/n(n-1)$ and consider distributions $B_2(dv_1 dv_2; u_1, u_2; \nu) = B_2(dv_2 dv_1; u_2, u_1; \nu)$, w-continuous in ν, u_1, u_2 and satisfying (7.5). As earlier we suppose that these $(n, 2)$-particle processes are connected with the $\mathbf{T}$-process in the sense that

$$
B_2(dv_1 \times D; u_1, u_2; \nu) = T(dv_1; u_1). \tag{8.7}
$$

Note that (8.7) induces (7.5).

Suppose that B_2 does not depend on ν. To produce (8.7) we assume that there exist a random variable ω and a function φ such that $\mathcal{L}(\varphi(\omega, u)) = T(\,\cdot\,; u)$ for any u. Then, taking ω, ω' with $\mathcal{L}(\omega) = \mathcal{L}(\omega')$, we put

$$
B_2(\,\cdot\,; u_1, u_2) = \mathcal{L}(\varphi(\omega, u_1), \varphi(\omega', u_2))
$$

and (8.7) holds. The invariance conditions are also valid here. As for the continuity conditions, they may be easily formulated in terms of function φ. Surely, one can take $\omega' \equiv \omega$, ω' independent from ω and so on.

Algorithmically, a jump of such a process may be described in the following way. If $(u_1, \ldots, u_n)$ is a position of the process just before a jump, we choose particles with coordinates u_i and u_j $(i \neq j)$ with the probability $1/n(n-1)$ and they get their new coordinates u_1' and u_2' by $u_1' \leftarrow \varphi(\omega, u_1)$ and $u_2' \leftarrow \varphi(\omega', u_2)$. This algorithm may be called an $\mathbf{B}_\varphi$-*algorithm*, as it corresponds to the artificial *binary* collisions, acting through φ.

Denote $\Delta_\varphi(\tau,\omega,u) = \psi(\tau,\varphi(\omega,u)) - \psi(\tau,u)$. Similarly to (7.16) and (7.17) we have for $\mathbf{B}_\varphi$-algorithms that

$$V_{\mathbf{B}_\varphi}(\psi,t,\mu) = \int\limits_0^t \Big(s^2_{\mathbf{T}}(\psi,\tau,\mu_{t-\tau}(\mu)) + \sigma_{\mathbf{B}_\varphi}(\psi,\tau,\mu_{t-\tau}(\mu))\Big)d\tau,$$

where $s^2_{\mathbf{T}}$ and $\sigma_{\mathbf{B}_\varphi}$ may be expressed as

$$s^2_{\mathbf{T}}(\psi,\tau,\nu) = \mathbf{E}\int\limits_D \Delta^2_\varphi(\tau,\omega,u)\nu(du),$$

$$\sigma_{\mathbf{B}_\varphi}(\psi,\tau,\nu) = \mathbf{E}\int\limits_{D^2} \Delta_\varphi(\tau,\omega,u_1)\Delta_\varphi(\tau,\omega',u_2)\nu(du_1)\nu(du_2).$$

Therefore

$$V_{\mathbf{B}_\varphi}(\psi,t,\mu) = \frac{1}{2}\int\limits_0^t\int\limits_{D^2} \mathbf{E}\Big(\Delta_\varphi(\tau,\omega,u_1) + \Delta_\varphi(\tau,\omega',u_2)\Big)^2 \mu^{\otimes 2}_{t-\tau}(du_1du_2)d\tau$$

and is not greater than $2V_{\mathbf{T}}(\psi,t,\mu)$. Evidently, if ω,ω' are independent, then

$$V_{\mathbf{B}_\varphi}(\psi,t,\mu) - V_{\mathbf{T}}(\psi,t,\mu) = \int\limits_0^t \left(\int\limits_D \mathbf{E}\Delta_\varphi(\tau,\omega,u)\mu_{t-\tau}(du)\right)^2 d\tau \geq 0. \quad (8.8)$$

If we put $\omega \equiv \omega'$ and suppose that the computational complexity of $\varphi(\omega,u)$ for fixed ω and u is negligible in comparison with the simulation complexity of $T(\,\cdot\,;u)$, then $\mathbf{B}_\varphi$-algorithm has an advantage over the traditional one. The simple example of such a situation holds if $T(\,\cdot\,;u)$ is the uniform distribution on the unit sphere with the center $u \in \mathbf{R}^d$. Then $\varphi(\omega,u) = u+\omega$, where $\mathcal{L}(\omega) = T(\,\cdot\,;\mathbf{0})$. Note that one can put $\omega' = -\omega$ instead of $\omega' = \omega$ here.

For the other example of such a kind we take $D = \mathbf{R}$ and $T(\,\cdot\,;u) = \mathbf{N}(u,1)$, that is, the normal distribution with the parameters u and 1 (the mean and the variance). Then, as it may be easily checked, the fundamental solution of (8.1) has the form

$$\mu_t(\,\cdot\,;v) = \sum_{k=0}^{\infty} \frac{t^k}{k!}\,e^{-t}\,\mathbf{N}(v,k), \quad \mathbf{N}(v,0) \equiv \delta_v\,.$$

If we put $\mu = \mathbf{N}(0,1)$ and $g(u) = u^2$ then

$$\mu_t(\mu) = \sum_{k=0}^{\infty} \frac{t^k}{k!} e^{-t}\, \mathbf{N}(0, k+1),$$

$\psi(\mu_t(\mu)) = t+1$ and $\psi(t,v) = t + v^2$. Therefore $V_0(\psi,t,\mu) + V_{\mathbf{T}}(\psi,t,\mu) = \mathbf{D}g(\zeta(t)) = 2t^2 + 7t + 2$, while (8.5) gives us $V_0(\psi,t,\mu) = 2$. Thus the role of stratification is rather small for big t. For the artificial collisions, we note that $\varphi(u,\omega) = u+\omega$ with $\mathcal{L}(\omega) = \mathbf{N}(0,1)$ here, and therefore binary collisions may be simulated as $u_1' \leftarrow u_1 + \omega$, $u_2' \leftarrow u_2 + \omega$ (or $u_2' \leftarrow u_2 - \omega$). Surely, the calculation of u_2' is negligible in comparison with the simulation of the normal distribution.

The identity (8.8) shows that $V_{\mathbf{B}_\varphi} = V_{\mathbf{T}} + 3t$ for both variants of u_2', $V_{\mathbf{B}_\varphi}/V_{\mathbf{T}} \to 1$ as $t \to \infty$ and therefore for big t the traditional algorithm has at most twice greater complexity than the binary one. As for the simplest algorithm, $V_{\mathbf{S}} = O(t^3)$ here.

Note that these conclusions are not strictly correct as some technical assumptions do not hold here: $\mathbf{R}$ is not a compact and g is not bounded. Nevertheless simulation shows that all final formulas are valid.

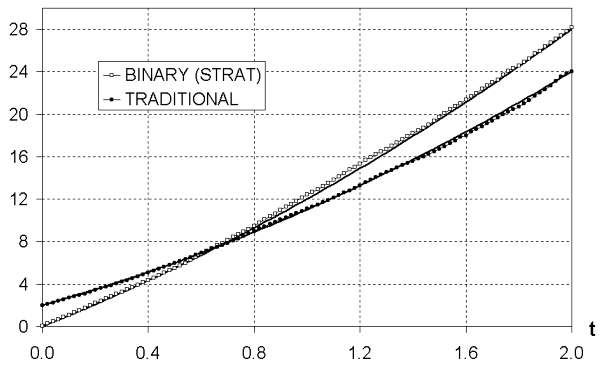

Figure 2: Theoretical and estimated variance coefficients

Four graphs of the Figure 2 depict the variance coefficients of the traditional (without a stratification) and the stratified binary algorithms, theoretical (thin lines) and calculated by simulation (points). For the simulation, $n = 100$ was taken and 10000 independent trajectories of n-particle processes were used to obtain estimates of the variance coefficients. The graphs show the excellent coincidence of theoretical and estimated characteristics.

References

[1] Berresford, G.C. (1976) A class of nonlinear partial differential equations and the associated Markov processes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **33**, 237–251.

[2] Bobylev, A. and Rjasanow, S. (1996) Analytical and numerical approach to one model kinetic equation, *Preprint* N° 60, IPM, Moscow (in Russian).

[3] Cartan, H. (1967) *Calcul différentiel. Formes différentielles,* Hermann, Paris.

[4] Clement, Ph. et al. (1987) *One-Parameter Semigroups,* North-Holland, Amsterdam–New York–Oxford–Tokyo.

[5] Ermakov, S.M., Nekrutkin, V.V. and Sipin, A.S. (1989) *Random Processes for Classical Equations of Mathematical Physics*, Kluwer Academic Publishers, Dordrecht-Boston-London.

[6] Ethier, S.N. and Kurtz, T.G. (1986) *Markov processes. Characterization and convergence,* John Wiley & Sons, New York.

[7] Golyandina, N. (1996) Markov processes for solving differential equations in measure spaces, In: *Mathematical methods in stochastic simulation and experimental design, Proc. 2nd St.Petersburg Workshop on simulation, St. Petersburg, June 18–21, 1996*, St. Petersburg University Press, 75–80.

[8] Golyandina, N. (1997) Rate of convergence of Markov processes to the solution of homogeneous Boltzmann type equation, *Workshop on Numerical Methods for Kinetic Equations, Book of Abstracts*, Berlin, 17.

[9] Golyandina, N. (1998) Stratification for functionals on measure spaces. In *Proc. 3-d St.Petersburg Workshop on simulation, St. Petersburg, June 28 – July 3, 1998*, St. Petersburg University Press, 25–30.

[10] Golyandina, N. and Nekrutkin, V. (1999) Estimation errors for functionals on measure spaces, In: *Stochastic simulation methods*, Eds.: N. Balakrishnan, S. Ermakov and V. Melas, Birkhauser, Boston (submitted); *Preprint* N° 1999-07, St. Petersburg Math. Soc., St. Petersburg, www.pdmi.ras.ru/spbmo.

[11] Graham, C. and Méléard, S. (1997) Stochastic particle approximation for generalized Boltzmann models and convergence estimates. *Ann. Prob.*, **25**, 115–132.

[12] Ivanov, A., Nekrutkin, V. and Tur, N. (1998) Asymptotical expansion in propagation of chaos for Boltzmann's type equations, *Proc. St. Petersburg Math. Soc.*, St. Peterburg, **5**, 93–111 (in Russian).

[13] Kac, M. (1956) Foundation of kinetic theory, *Proc. 3-d Berkly Symp. on Math. Stat. and Prob.*, Univ. Calif., **3**, 171–197.

[14] Méléard, S. (1996) Asymptotic behavior of some interacting particle systems; McKean-Vlasov and Boltzmann models. *LNM*, **1627**, 42–95.

[15] Méléard, S. (1998) Stochastic approximations of the solutions of a full Boltzmann equation with small initial data, *ESAIM: Probability and Statistics*, **2**, 23–40.

[16] Nekrutkin, V. and Skorina, I. (1994) Optimal simulation of simplest dynamic systems, In: *Mathematical Methods and Tools in Computer Simulation*, Intern. Workshop, May 24–26, 1994, St. Petersburg State University, St. Petersburg, Preprint N° MM94-01, 46–47.

[17] Nekrutkin, V. and Tur, N. (1987) On the stochastic approach to systems of nonlinear differential equations, *Sov. Math. Ref. J.*, Dep. N° 5565-B87 (in Russian).

[18] Nekrutkin, V. and Tur, N. (1997) Asymptotic expansions and estimators with small bias for Nanbu processes, *Monte Carlo Methods and Appl.*, **3**, N° 1, 1–35.

[19] Rjasanow, S. and Wagner, W. (1996) Numerical study of a stochastic weighted particle method for a model kinetic equation, *J. Comp. Physics*, **128**, 351–362.