***U*-Statistics of Ornstein–Uhlenbeck Branching Particle System**

Radosław Adamczak · Piotr Miłoś

Received: 22 November 2011 / Revised: 21 February 2013 / Published online: 22 November 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract We consider a branching particle system consisting of particles moving according to the Ornstein–Uhlenbeck process in $\mathbb{R}^d$ and undergoing a binary, supercritical branching with a constant rate $\lambda > 0$. This system is known to fulfill a law of large numbers (under exponential scaling). Recently the question of the corresponding central limit theorem (CLT) has been addressed. It turns out that the normalization and the form of the limit in the CLT fall into three qualitatively different regimes, depending on the relation between the branching intensity and the parameters of the Ornstein–Uhlenbeck process. In the present paper, we extend those results to *U*-statistics of the system, proving a law of large numbers and CLT.

Keywords Supercritical branching particle systems · *U*-statistics · Central limit theorem

Mathematics Subject Classification (2010) Primary 60F05 · 60J80;
Secondary 60G20

1 Introduction

We consider a single particle located at time $t = 0$ at $x \in \mathbb{R}^d$, moving according to the Ornstein–Uhlenbeck process and branching after an exponential time independent of the spatial movement. The branching is binary and supercritical, with probability $p > 1/2$ the particle is replaced by two offspring, and with probability $1 - p$ it vanishes.

R. Adamczak · P. Miłoś (✉)
Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Banacha 2,
02-097 Warsaw, Poland
e-mail: pmilos@mimuw.edu.pl

R. Adamczak
e-mail: radamcz@mimuw.edu.pl

The offspring particles follow the same dynamics (independently of each other). We will refer to this system of particles as the OU branching process and denote it by $X = \{X_t\}_{t \geq 0}$.

We identify the system with the empirical process, i.e., X takes values in the space of Borel measures on $\mathbb{R}^d$ and for each Borel set A , $X_t(A)$ is the (random) number of particles at time t in A . We refer to [14] for the general construction of X as a measure-valued stochastic process.

It is well known (see, e.g., [16]) that the system satisfies the law of large numbers, i.e., for any bounded continuous function f , conditionally on the set of non-extinction

$$|X_t|^{-1} \langle X_t, f \rangle \rightarrow \langle \varphi, f \rangle, \quad a.s., \quad (1)$$

where $|X_t|$ is the number of particles at time t , $\{X_t(1), X_t(2), \dots, X_t(|X_t|)\}$ are their positions, $\langle X_t, f \rangle := \sum_{i=1}^{|X_t|} f(X_t(i))$ and φ is the invariant measure of the Ornstein–Uhlenbeck process.

In a recent article [1], we investigated second-order behavior of this system and proved central limit theorems (CLT) corresponding to (1). We found three qualitatively different regimes, depending on the relation between the branching intensity and the parameters of the Ornstein–Uhlenbeck process.

In the present article, we extend these results on the LLN and CLT to the case of U -statistics of the system of arbitrary order $n \geq 1$, i.e., to random variables of the form

$$U_t^n(f) := \sum_{\substack{i_1, i_2, \dots, i_n=1 \\ i_k \neq i_j \text{ if } k \neq j}}^{|X_t|} f(X_t(i_1), X_t(i_2), \dots, X_t(i_n)), \quad (2)$$

(note that $U_t^1(f) = \langle X_t, f \rangle$). Our investigation parallels the classical and well-developed theory of U -statistics of independent random variables; however, we would like to point out that in our context, additional interest in this type of functionals of the process X stems from the fact that they capture “average dependencies” between particles of the system. This will be seen from the form of the limit, which turns out to be more complicated than in the i.i.d. case. This is one of the motivations for studying U -statistics in a more general context than for independent random variables. Another one is that while being structurally the simplest generalization of additive functionals (which are U -statistics of degree one), they may be considered building blocks for other, more complicated statistics as they appear naturally in their Taylor expansions (see [7], where in the i.i.d. context such expansions are used in the analysis of some statistical estimators). In another context, they also appear in the study of tree-based expansions and propagation of chaos in interacting particle systems (see [9–11, 23]).

The organization of the paper is as follows. After introducing the basic notation and preliminary facts in Sect. 2, we describe the main results of the paper in Sect. 3. Next (Sect. 4), we restate the results in the special case of $n = 1$ (as proven in [1]) to serve as a starting point for the general case. Finally, in Sect. 5, we provide proofs for arbitrary n , postponing some of the technical details (which may obscure the main ideas of the proofs) to Sect. 6. We conclude with some remarks concerning the so-called non-degenerate case (Sect. 7).

2 Preliminaries

2.1 Notation

For a branching system $\{X_t\}_{t \geq 0}$, we denote by $|X_t|$ the number of particles at time t and by $X_t(i)$ —the position of the i th (in a certain ordering) particle at time t . We sometimes use $\mathbb{E}_x$ or $\mathbb{P}_x$ to denote the fact that we calculate the expectation for the system starting from a particle located at x . We use also $\mathbb{E}$ and $\mathbb{P}$ when this location is not relevant.

By $\rightarrow^d$, we denote the convergence in law. We use $\lesssim$ to denote the situation when an inequality holds with a constant $c > 0$, which is irrelevant to calculations, e.g., $f(x) \lesssim g(x)$ means that there exists a constant $c > 0$ such that $f(x) \leq cg(x)$.

By $x \circ y = \sum_{i=1}^d x_i y_i$, we denote the standard scalar product of $x, y \in \mathbb{R}^d$, by $\|\cdot\|$ the corresponding Euclidean norm. By $\otimes^n$, we denote the n -fold tensor product.

We use also $\langle f, \mu \rangle := \int_{\mathbb{R}^d} f(x) \mu(dx)$. We will write $X \sim \mu$ to describe the fact that a random variable X is distributed according to the measure μ , similarly $X \sim Y$ will mean that X and Y have the same law.

For a subset A of a linear space by $\text{span}(A)$, we denote the set of finite linear combinations of elements of A .

In the paper, we will use Feynman diagrams. A diagram γ on a set of vertices $\{1, 2, \dots, n\}$ is a graph on $\{1, 2, \dots, n\}$ consisting of a set of edges E_γ not having common endpoints and a set of unpaired vertices A_γ . We will use $r(\gamma)$ to denote the rank of the diagram, i.e., the number of edges. For properties and more information, we refer to [21, Definition 1.35].

In the paper, we will use the space

$$\mathcal{P} = \mathcal{P}(\mathbb{R}^d) := \{f : \mathbb{R}^d \mapsto \mathbb{R} : f \text{ is continuous and} \\ \exists_k \text{ such that } |f(x)|/\|x\|^k \rightarrow 0 \text{ as } \|x\| \rightarrow +\infty\}. \quad (3)$$

We endow this space with the following norm

$$\|f\|_{\mathcal{P}} = \|f\|_{\mathcal{P}(\mathbb{R}^d)} := \sup_{x \in \mathbb{R}^d} |n(x)f(x)|, \quad (4)$$

where $n(x) := \exp\left(-\sum_{i=1}^d |x_i|\right)$. We will also use

$$\mathcal{C}_c = \mathcal{C}_c(\mathbb{R}^d),$$

to denote the space of continuous compactly supported functions.

Given a function $f \in \mathcal{P}(\mathbb{R}^d)$, we will implicitly understand its derivatives (e.g., $\frac{\partial f}{\partial x_i}$) in the space of tempered distributions (see, e.g., [25, p. 173]).

By $f(a \cdot)$, we denote the function $x \mapsto f(ax)$.

Let μ_1, μ_2 be two probability measures on $\mathbb{R}$, and $\text{Lip}(1)$ be the space of 1-Lipschitz functions $\mathbb{R} \mapsto [-1, 1]$. We define

$$m(\mu_1, \mu_2) := \sup_{g \in \text{Lip}(1)} |\langle g, \mu_1 \rangle - \langle g, \mu_2 \rangle|. \quad (5)$$

It is well known that m is a distance metrizing the weak convergence (see, e.g., [12, Theorem 11.3.3]). One easily checks that when μ_1, μ_2 correspond to two random variables X_1, X_2 on the same probability space then we have

$$m(\mu_1, \mu_2) \leq \|X_1 - X_2\|_1 \leq \sqrt{\|X_1 - X_2\|_2}. \quad (6)$$

2.2 Basic Facts on the Galton–Watson Process

The number of particles $\{|X_t|\}_{t \geq 0}$ is the celebrated Galton–Watson process. We present basic properties of this process used in the paper. The main reference in this section is [2]. In our case, the expected total number of particles grows exponentially at the rate

$$\lambda_p := (2p - 1)\lambda. \quad (7)$$

The process becomes extinct with probability (see [2, Theorem I.5.1])

$$p_e = \frac{1 - p}{p}.$$

We will denote the extinction and non-extinction events by Ext and Ext^c , respectively. The process $V_t := e^{-\lambda_p t} |X_t|$ is a positive martingale. Therefore, it converges (see also [2, Theorem 1.6.1])

$$V_t \rightarrow V_\infty, \quad a.s. \quad \text{as } t \rightarrow +\infty. \quad (8)$$

We have the following simple fact (we refer to [1] for the proof).

Proposition 1 *We have $\{V_\infty = 0\} = Ext$ and conditioned on non-extinction V_∞ has the exponential distribution with parameter $\frac{2p-1}{p}$. We have $\mathbb{E}(V_\infty) = 1$ and $\text{Var}(V_\infty) = \frac{1}{2p-1}$. $\mathbb{E}e^{-4\lambda_p t} |X_t|^4$ is uniformly bounded, i.e., there exists $C > 0$ such that for any $t \geq 0$ we have $\mathbb{E}e^{-4\lambda_p t} |X_t|^4 \leq C$. Moreover, all moments are finite, i.e., for any $n \in \mathbb{N}$ and $t \geq 0$ we have $\mathbb{E}|X_t|^n < +\infty$.*

We will denote the variable V_∞ conditioned on non-extinction by W .

2.3 Basic Facts on the Ornstein–Uhlenbeck Process

We recall that the Ornstein–Uhlenbeck process with parameters $\sigma, \mu > 0$ is a time-homogeneous Markov process with the infinitesimal operator

$$L := \frac{1}{2} \sigma^2 \Delta - \mu x \circ \nabla. \quad (9)$$

The corresponding semigroup will be denoted by $\mathcal{T}$. The density of the invariant measure of the Ornstein–Uhlenbeck process is given by

$$\varphi(x) := \left(\frac{\mu}{\pi\sigma^2} \right)^{d/2} \exp \left(-\frac{\mu}{\sigma^2} \|x\|^2 \right), \quad (10)$$

i.e., it is a centered Gaussian distribution with the covariance matrix $\frac{\sigma^2}{2\mu} \text{Id}$. For simplicity, we will denote both this measure and its density with the same letter φ .

2.4 Basic Facts Concerning U -Statistics

We will now briefly recall basic notation and facts concerning U -statistics. A U -statistic of degree n based on an $\mathcal{X}$ -valued sample $X_1, \dots, X_N$ and a function $f: \mathcal{X}^n \rightarrow \mathbb{R}$ is a random variable of the form

$$\sum_{\substack{i_1, i_2, \dots, i_n=1 \\ i_k \neq i_j \text{ if } k \neq j}}^N f(X_{i_1}, \dots, X_{i_n}).$$

The function f is usually referred to as the kernel of the U -statistic. Without loss of generality, it can be assumed that f is symmetric, i.e., invariant under permutation of its arguments. We refer the reader to [7, 22] for more information on U -statistics of sequences of independent random variables.

In our case, we will consider U -statistics based on positions of particles from the branching system as defined by (2). We will be interested in weak convergence of properly normalized U -statistics when $t \rightarrow \infty$. Similarly as in the classical theory, the asymptotic behavior of U -statistics depends heavily on the so-called order of degeneracy of the kernel f , which we will briefly recall in Sect. 5.2.

A function f is called completely degenerate or canonical (with respect to some measure of reference φ , which in our case will be the stationary measure of the Ornstein–Uhlenbeck process) if

$$\int_{\mathcal{X}} f(x_1, \dots, x_n) \varphi(dx_k) = 0,$$

for all $x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n \in \mathcal{X}$. The complete degeneracy may be considered a centeredness condition, in the classical theory of U -statistics canonical kernels are counterparts of centered random variables from the theory of sums of independent random variables. Their importance stems from the fact that each U -statistic can be decomposed into a sum of canonical U -statistics of different degrees, a fact known as the Hoeffding decomposition (see Sect. 5.2). Thus, in the main part of the article, we prove results only for canonical U -statistics. Their counterparts for general U -statistics, which can be easily obtained via the Hoeffding decomposition, are stated in Sect. 7.

3 Main Results

This section is devoted to the presentation of our results. The proofs are deferred to Sect. 5.

We start with the following law of large numbers (throughout the article when dealing with U -statistics of order n we will identify $\underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_n$ with $\mathbb{R}^{nd}$).

Theorem 2 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $f : \mathbb{R}^{nd} \mapsto \mathbb{R}$ is a bounded continuous function. Then, on the set of non-extinction Ext^c there is the convergence*

$$\lim_{t \rightarrow +\infty} |X_t|^{-n} U_t^n(f) = \langle f, \varphi^{\otimes n} \rangle \text{ a.s.} \quad (11)$$

Moreover, when $f \in \mathcal{P}(\mathbb{R}^{nd})$, then the above convergence holds in probability.

Having formulated the law of large numbers, let us now pass to the corresponding CLTs. We recall that μ is the drift parameter in (9) and λ_p is the growth rate (7). As already mentioned in the introduction, the form of the limit theorems depends on the relation between λ_p and μ , more specifically, we distinguish three cases: $\lambda_p < 2\mu$, $\lambda_p = 2\mu$ and $\lambda_p > 2\mu$. We refer the reader to [1] (Introduction and Section 3) for a detailed discussion of this phenomenon as well as its heuristic explanation and interpretation. Here, we only stress that the situation for $\lambda_p > 2\mu$ differs substantially from the remaining two cases, as we obtain convergence in probability and the limit is not Gaussian even for $n = 1$. Intuitively, this is caused by large branching intensity which lets local correlations between particles prevail over the ergodic properties of the Ornstein–Uhlenbeck process.

3.1 Slow Branching Case: $\lambda_p < 2\mu$

Let Z be a Gaussian stochastic measure on $\mathbb{R}^{d+1}$ with intensity

$$\mu_1(dt dx) := (\delta_0(dt) + 2\lambda_p e^{\lambda_p t} dt) \varphi(x) dx,$$

defined according to [21, Definition 7.17], where δ_0 is the Dirac measure concentrated at 0. We denote the stochastic integral with respect to Z by I and the corresponding multiple stochastic integral by I_n [21, Section 7.2]. We assume that Z is defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

For $f \in \mathcal{P}(\mathbb{R}^{nd})$ we define (we recall that $\mathcal{T}$ is the semigroup of the Ornstein–Uhlenbeck process)

$$\begin{aligned} H(f)(s_1, x_1, s_2, x_2, \dots, s_n, x_n) &:= \left(\otimes_{i=1}^n \mathcal{T}_{s_i} \right) f(x_1, x_2, \dots, x_n), \\ s_i &\in \mathbb{R}_+, x_i \in \mathbb{R}^d. \end{aligned} \quad (12)$$

It will be useful to treat this function as a function of n variables of type $z_i := (s_i, x_i) \in \mathbb{R}_+ \times \mathbb{R}^d$. For a Feynman diagram γ labeled by $\{1, 2, \dots, n\}$ and $f \in \mathcal{P}(\mathbb{R}^{nd})$ we define

$$L(f, \gamma) := I_{|A_\gamma|} \left(\left(\prod_{(j,k) \in E_\gamma} \int \mu_2(dz_{j,k}) \right) H(f)(u_1, u_2, \dots, u_n) \right),$$

$$\mu_2(dsdx) := 2\lambda p e^{\lambda p s} \varphi(x) dsdx, \quad (13)$$

where $u_i = z_{j,k}$ if $(j, k) \in E_\gamma$ and $(i = j \text{ or } i = k)$ and $u_i = z_i$ if $i \in A_\gamma$. Less formally, for each pair (j, k) , we integrate over diagonal of coordinates j and k with respect to μ_2 . The function obtained in this way is integrated using the multiple stochastic integral $I_{|A_\gamma|}$. We define

$$L_1(f) := \sum_{\gamma} L(f, \gamma), \quad (14)$$

where the sum spans over all Feynman diagrams labeled by $\{1, 2, \dots, n\}$.

We are now ready to formulate our main result for processes with small branching rate. Recall (8) and that W is V_∞ conditioned on Ext^c .

Theorem 3 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $f \in \mathcal{P}(\mathbb{R}^{nd})$ is a canonical kernel and $\lambda_p < 2\mu$. Then conditionally on the set of non-extinction Ext^c there is the convergence*

$$\left(e^{-\lambda_p t} |X_t|, \frac{|X_t| - e^{\lambda_p t} V_\infty}{\sqrt{|X_t|}}, \frac{U_t^n(f)}{|X_t|^{n/2}} \right) \rightarrow^d (W, G_1, L_1(f)), \quad (15)$$

where $G_1 \sim \mathcal{N}(0, 1/(2p-1))$ and $W, G_1, L_1(f)$ are independent random variables.

3.2 Critical Branching Case: $\lambda_p = 2\mu$

Before we present the main results in the critical branching case, we need to introduce some additional notation.

Consider the orthonormal Hermite basis $\{h_i\}_{i \geq 0}$ for the measure $\gamma = \mathcal{N}(0, \frac{\sigma^2}{2\mu})$ (i.e, $h_0 = 1, h_i$ is a polynomial of degree i and $\int h_i h_j d\gamma = \delta_{ij}$). Then for any positive integer n , the set $\{h_{i_1} \otimes \dots \otimes h_{i_{nd}}\}_{i_1, \dots, i_{nd} \geq 0}$ of multivariate Hermite polynomials is an orthonormal basis in $L_2(\mathbb{R}^{nd}, \varphi^{\otimes n})$. For a function $f \in L_2(\mathbb{R}^{nd}, \varphi^{\otimes n})$ let $\tilde{f}_{i_1, \dots, i_{nd}}$ be the sequence of coefficients of f with respect to this basis, i.e., $f = \sum_{i_1, \dots, i_{nd} \geq 0} \tilde{f}_{i_1, \dots, i_{nd}} h_{i_1} \otimes \dots \otimes h_{i_{nd}}$.

Let $\mathcal{H}^n = \text{span}\{h_{i_1} \otimes \dots \otimes h_{i_{nd}} : \sum_{j=1}^d i_{kd+j} = 1, k = 0, \dots, n-1\}$ be the subspace of $L_2(\mathbb{R}^{nd}, \varphi^{\otimes n})$. In other words, $\mathcal{H}^1$ is the subspace of $L_2(\mathbb{R}^d, \varphi)$ spanned by Hermite polynomials of degree 1 on $\mathbb{R}^d$ and $\mathcal{H}^n = (\mathcal{H}^1)^{\otimes n}$.

Consider now a centered Gaussian process $(G_f)_{f \in L_2(\mathbb{R}^d, \varphi)}$ on $L_2(\mathbb{R}^d, \varphi)$, with the covariance structure given by

$$\begin{aligned} \text{Cov}(G_f, G_g) &= 2\lambda p (\tilde{f}_{1,0,\dots,0} \tilde{g}_{1,0,\dots,0} + \tilde{f}_{0,1,\dots,0} \tilde{g}_{0,1,\dots,0} + \dots + \tilde{f}_{0,\dots,0,1} \tilde{g}_{0,\dots,0,1}) \\ &= 2\lambda p \langle (Pf)(Pg), \varphi \rangle, \end{aligned} \quad (16)$$

where P is the orthogonal projection from $L_2(\mathbb{R}^d, \varphi)$ onto $\mathcal{H}^1$.

We will identify this process with a map $I: L_2(\mathbb{R}^d, \varphi) \rightarrow L_2(\Omega, \mathcal{F}, \mathbb{P})$, such that $I(f) = G_f$. One can easily check that I is a bounded linear operator. Moreover, $I = IP$. In fact I is the stochastic integral of Pf with respect to the random Gaussian measure on $\mathbb{R}^d$ with intensity $2\lambda p\varphi$ (however, we will not use this fact in the sequel).

Since $\mathcal{H}^n = (\mathcal{H}^1)^{\otimes n}$, there exists a unique linear operator $\tilde{L}_2: \mathcal{H}^n \rightarrow L_2(\Omega, \mathcal{F}, \mathbb{P})$ such that for any functions $f_1, \dots, f_n \in \mathcal{H}^1$,

$$\tilde{L}_2(f_1 \otimes \dots \otimes f_n) = I(f_1) \cdots I(f_n)$$

(we used here the fact that Gaussian variables have all moments finite).

Let now $P_n: L_2(\mathbb{R}^{nd}, \varphi^{\otimes n}) \rightarrow \mathcal{H}^n$ be the orthogonal projection onto $\mathcal{H}^n$. We have $P_n = P^{\otimes n}$ and using the fact that $\mathcal{H}^n$ are finite dimensional we obtain that the linear operator $L_2: L_2(\mathbb{R}^{nd}, \varphi^{\otimes n}) \rightarrow L_2(\Omega, \mathcal{F}, \mathbb{P})$ defined as

$$L_2(f) = \tilde{L}_2(P_n f) \quad (17)$$

is bounded. Note that for $f_1, \dots, f_n \in L_2(\mathbb{R}^d, \varphi)$ we have

$$L_2(f_1 \otimes \dots \otimes f_n) = G_{f_1} \cdots G_{f_n}.$$

We are now ready to formulate the theorem, which describes the asymptotic behavior of U -statistics in the critical case.

Theorem 4 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $f \in \mathcal{P}(\mathbb{R}^{nd})$ is a canonical kernel and $\lambda_p = 2\mu$. Then conditionally on the set of non-extinction Ext^c there is the convergence*

$$\left(e^{-\lambda_p t} |X_t|, \frac{|X_t| - e^{\lambda_p t} V_\infty}{\sqrt{|X_t|}}, \frac{U_t^n(f)}{(t|X_t|)^{n/2}} \right) \rightarrow^d (W, G, L_2(f)),$$

where $G \sim \mathcal{N}(0, 1/(2p-1))$ and $W, G, L_2(f)$ are independent random variables.

Remark 5 One can express $L_2(f)$ in terms of the Hermite expansion of the function f , which might give more insight into the structure of the limiting law. Indeed, define $\tilde{h}_i: \mathbb{R}^d \rightarrow \mathbb{R}, i = 1, \dots, d$, by $\tilde{h}_i(x_1, \dots, x_d) = h_1(x_i)$ (thus, $\tilde{h}_i = h_0^{\otimes(i-1)} \otimes h_1 \otimes h_0^{\otimes(d-i)}$). Then for $f \in L_2(\mathbb{R}^{dn}, \varphi^{\otimes n})$,

$$L_2(f) = (2\lambda p)^{n/2} \sum_{i_1, \dots, i_n \leq d} \langle f \tilde{h}_{i_1} \otimes \dots \otimes \tilde{h}_{i_n}, \varphi^{\otimes n} \rangle G_{i_1} \cdots G_{i_n},$$

where $G_i = (2\lambda p)^{-1/2} G_{\tilde{h}_i}$. In particular $(G_1, \dots, G_d)$ is a vector of independent standard Gaussian variables (this follows easily from the covariance structure of the process (G_f) and the fact that the functions $\tilde{h}_i$ form an orthonormal system).

3.3 Fast Branching Case: $\lambda_p > 2\mu$

In order to describe the limit, we introduce an $\mathbb{R}^d$ -valued process $\{H_t\}_{t \geq 0}$ by

$$H_t := e^{(-\lambda_p + \mu)t} \sum_{i=1}^{|X_t|} X_t(i). \quad (18)$$

The following two facts have been proved in [1, Propositions 3.9, 3.10].

Proposition 6 *H is a martingale with respect to the filtration of the OU branching system starting from $x \in \mathbb{R}^d$. Moreover for $\lambda_p > 2\mu$, we have $\sup_t \mathbb{E} \|H_t\|^2 < +\infty$, therefore there exists $H_\infty := \lim_{t \rightarrow +\infty} H_t$ (a.s. limit) and $H_\infty \in L_2$. When the OU branching system starts from 0, then martingales V_t and H_t are orthogonal.*

It is worthwhile to note that the distribution of H_∞ depends on the starting conditions.

Proposition 7 *Let $\{X_t\}_{t \geq 0}$ and $\{\tilde{X}_t\}_{t \geq 0}$ be two OU branching processes, the first one starting from 0 and the second one from x . Let us denote the limit of the corresponding martingales by $H_\infty, \tilde{H}_\infty$, respectively. Then*

$$\tilde{H}_\infty \sim H_\infty + xV_\infty,$$

where V_∞ is the limit given by (8) for the system X .

H_∞ is $\mathbb{R}^d$ -valued, we denote its coordinates by H_∞^i . Let $f \in \mathcal{P}(\mathbb{R}^{nd})$. We define

$$\tilde{L}_3(f) := \sum_{i_1, i_2, \dots, i_n=1}^d \left\langle \frac{\partial^n f}{\partial x_{1,i_1} \partial x_{2,i_2}, \dots, \partial x_{n,i_n}}, \varphi^{\otimes n} \right\rangle H_\infty^{i_1} H_\infty^{i_2} \dots H_\infty^{i_n},$$

where we adopted the convention that $x_{j,l}$ is the l th coordinate of the j th variable. By $L_3(f)$, we will denote $\tilde{L}_3(f)$ conditioned on Ext^c .

Theorem 8 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $f \in \mathcal{P}(\mathbb{R}^{nd})$ is a canonical kernel and $\lambda_p > 2\mu$. Then conditionally on the set of non-extinction Ext^c there is the convergence*

$$\left(e^{-\lambda_p t |X_t|}, \frac{|X_t| - e^{\lambda_p t} V_\infty}{\sqrt{|X_t|}}, e^{-n(\lambda_p - \mu)t} U_t^n(f) \right) \rightarrow^d (W, G_1, L_3(f)), \quad (19)$$

where $G_1 \sim \mathcal{N}(0, 1/(2p-1))$ and $(W, L_3(f)), G_1$ are independent. Moreover

$$\left(e^{-\lambda_p t |X_t|}, e^{-n(\lambda_p - \mu)t} U_t^n(f) \right) \rightarrow (V_\infty, \tilde{L}_3(f)) \text{ in probability.}$$

3.4 Remarks on the CLT for U -Statistics of i.i.d. Random Variables

For comparison purposes, we will now briefly recall known results on the CLT for U -statistics of independent random variables. U -statistics were introduced in the 1940s in the context of unbiased estimation by Halmos [19] and Hoeffding who obtained the CLT for non-degenerate (degenerate of order 0, see Sect. 5.2) kernels [20]. The full description of the CLT was obtained in [15, 24] (see also the article [18] where the CLT is proven for a related class of V -statistics). Similarly as in our case, the asymptotic behavior of U -statistics based on a function $f: \mathcal{X}^n \rightarrow \mathbb{R}$ and an i.i.d. $\mathcal{X}$ -valued sequence $X_1, X_2, \dots$ is governed by the order of degeneracy of the function f (see Sect. 5.2) with respect to the law of X_1 (call it P). The case of general f can be reduced to the canonical one, for which one has the weak convergence

$$N^{-n/2} \sum_{\substack{i_1, i_2, \dots, i_n=1 \\ i_k \neq i_j \text{ if } k \neq j}}^N f(X_{i_1}, \dots, X_{i_n}) \rightarrow^d J_n(f),$$

where J_n is the n -fold stochastic integral with respect to the so-called isonormal process on $\mathcal{X}$, i.e., the stochastic Gaussian measure with intensity P .

For the small branching rate case, the behavior of U -statistics in our case resembles the classical one as the limit is a sum of multiple stochastic integrals of different orders. In the remaining two cases, the behavior differs substantially. This can be regarded as a result of the lack of independence. Although asymptotically the particles' positions become less and less dependent, in short timescale, offspring of the same particle stay close one to another.

Let us finally mention some results for U -statistics in dependent situations, which have been obtained in the last years. In [6], the authors analyzed the behavior of U -statistics of stationary absolutely regular sequences and obtained the CLT in the non-degenerate case (with Gaussian limit). In [5], the authors considered α and φ mixing sequences and obtained a general CLT for canonical kernels. Interesting results for long-range dependent sequences have been also obtained in [8]. A more recent interesting work (already mentioned in the introduction) is [9–11, 23], where the authors consider U -statistics of interacting particle systems.

4 The Case of $n = 1$

In the special case of $n = 1$, the results presented in the previous section were proven in [1]. Although this case obviously follows immediately from the results for general n , it is actually a starting point in the proof of the general result (similarly as in the case of U -statistics of i.i.d. random variables). Therefore, for the reader's convenience, we will now restate this case in a simpler language of [1], not involving multiple stochastic integrals.

We start with the law of large numbers

Theorem 9 Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $f \in \mathcal{P}(\mathbb{R}^d)$. Then

$$\lim_{t \rightarrow +\infty} e^{-\lambda_p t} \langle X_t, f \rangle = \langle f, \varphi \rangle V_\infty \text{ in probability,}$$

or equivalently on the set of non-extinction, Ext^c , we have

$$\lim_{t \rightarrow +\infty} |X_t|^{-1} \langle X_t, f \rangle = \langle f, \varphi \rangle \text{ in probability.} \quad (20)$$

Moreover, if f is bounded then the almost sure convergence holds.

4.1 Small Branching Rate: $\lambda_p < 2\mu$

We denote $\tilde{f}(x) := f(x) - \langle f, \varphi \rangle$ and

$$\sigma_f^2 := \left\langle \varphi, \tilde{f}^2 \right\rangle + 2\lambda_p \int_0^{+\infty} \left\langle \varphi, \left(e^{(\lambda_p/2)s} \mathcal{T}_s \tilde{f} \right)^2 \right\rangle ds. \quad (21)$$

Let us also recall (8) and that W is V_∞ conditioned on Ext^c . In this case, the behavior of X is given by the following

Theorem 10 Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $\lambda_p < 2\mu$ and $f \in \mathcal{P}(\mathbb{R}^d)$. Then $\sigma_f^2 < +\infty$ and conditionally on the set of non-extinction Ext^c , there is the convergence

$$\left(e^{-\lambda_p t} |X_t|, \frac{|X_t| - e^{\lambda_p t} V_\infty}{\sqrt{|X_t|}}, \frac{\langle X_t, f \rangle - |X_t| \langle f, \varphi \rangle}{\sqrt{|X_t|}} \right) \rightarrow^d (W, G_1, G_2),$$

where $G_1 \sim \mathcal{N}(0, 1/(2p - 1))$, $G_2 \sim \mathcal{N}(0, \sigma_f^2)$ and W, G_1, G_2 are independent random variables.

4.2 Critical Branching Rate: $\lambda_p = 2\mu$

Recall the notation related to Hermite polynomials introduced in Sect. 3.2 and denote

$$\begin{aligned} \sigma_f^2 &:= \frac{\lambda_p \sigma^2}{\mu} \sum_{i=1}^d \left\langle f, \frac{\partial \varphi}{\partial x_i} \right\rangle^2 \\ &= 2\lambda_p (\tilde{f}_{1,0,\dots,0}^2 + \tilde{f}_{0,1,\dots,0}^2 + \dots + \tilde{f}_{0,\dots,0,1}^2) \\ &= 2\lambda_p \langle (Pf)^2, \varphi \rangle \end{aligned} \quad (22)$$

(where the first equality follows from the form of the Gaussian density and its relation to Hermite polynomials, whereas the second one from the definition of P).

Note that the same symbol σ_f^2 has already been used to denote the asymptotic variance in the small branching case. However, since these cases will always be treated separately, this should not lead to ambiguity.

Theorem 11 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $\lambda_p = 2\mu$ and $f \in \mathcal{P}(\mathbb{R}^d)$. Then $\sigma_f^2 < +\infty$ and conditionally on the set of non-extinction Ext^c there is the convergence*

$$\left(e^{-\lambda_p t} |X_t|, \frac{|X_t| - e^{\lambda_p t} V_\infty}{\sqrt{|X_t|}}, \frac{\langle X_t, f \rangle - |X_t| \langle f, \varphi \rangle}{t^{1/2} \sqrt{|X_t|}} \right) \rightarrow^d (W, G_1, G_2),$$

where $G_1 \sim \mathcal{N}(0, 1/(2p-1))$, $G_2 \sim \mathcal{N}(0, \sigma_f^2)$ and W, G_1, G_2 are independent random variables.

4.3 Fast Branching Rate: $\lambda_p > 2\mu$

In the following theorem, we use the notation introduced in Sect. 3.3.

Theorem 12 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Let us assume that $\lambda_p > 2\mu$ and $f \in \mathcal{P}(\mathbb{R}^d)$. Then conditionally on the set of non-extinction Ext^c , there is the convergence*

$$\left(e^{-\lambda_p t} |X_t|, \frac{|X_t| - e^{\lambda_p t} V_\infty}{\sqrt{|X_t|}}, \frac{\langle X_t, f \rangle - |X_t| \langle f, \varphi \rangle}{\exp((\lambda_p - \mu)t)} \right) \rightarrow^d (W, G, \langle \nabla f, \varphi \rangle \circ J), \quad (23)$$

where $G \sim \mathcal{N}(0, 1/(2p-1))$, $(W, J), G$ are independent and J is H_∞ conditioned on Ext^c . Moreover

$$\left(e^{-\lambda_p t} |X_t|, \frac{\langle X_t, f \rangle - |X_t| \langle f, \varphi \rangle}{\exp((\lambda_p - \mu)t)} \right) \rightarrow (V_\infty, \langle \nabla f, \varphi \rangle \circ H_\infty), \quad \text{in probability.}$$

5 Proofs of Main Results

5.1 Outline of the Proofs

We will now pass to the proofs of the results announced in Sect. 3. Their general structure is similar to the case of U -statistics of independent random variables, i.e., all the theorems will be proved first for linear combinations of tensor products and then via suitable approximations extended to the function space $\mathcal{P}(\mathbb{R}^{nd})$. Below we provide a brief outline of the proofs, common for all the cases considered in the paper.

1. Using the one-dimensional versions of the results, presented in Sect. 4, and the Cramér–Wold device, one proves convergence for functions f , which are linear combinations of tensor products. This class is shown to be dense in $\mathcal{P}$.

2. Using algebraic properties of the covariance, one obtains explicit formulas for the limit, which are well defined for any function $f \in \mathcal{P}$. Further, one shows that they depend on f in a continuous way.
3. One obtains a *uniform in t* bound on the distance between the laws of $U_t^n(f)$ and of $U_t^n(g)$ in terms of the distance between f and g in $\mathcal{P}$.
This is the most involved and technical step as it relies on the analysis of moments of U -statistics. It turns out that the formulas for moments can be expressed in terms of auxiliary branching processes indexed by combinatorial structures, more specifically by labeled trees of a special type (introduced in Sect. 6.3). Having this representation, one can then obtain moment bounds via combinatorial arguments.
4. Combining the above three steps, one can easily conclude the proofs by standard metric-theoretic arguments. By step 3, a general U -statistic based on a function f can be approximated (uniformly in t) by a U -statistics based on special functions f_n whose laws converge by step 1 as $t \rightarrow \infty$ to some limiting measure μ_n . By step 2, when the approximation becomes finer and finer ($n \rightarrow \infty$), one has $\mu_n \rightarrow \mu$ for some probability measure μ . Finally, it is easy to see that μ is the limiting measure for the original U -statistic.

The organization of the rest of the paper is as follows. First, we recall some basic facts about U - and V -statistics and Hoeffding projections, which we will need already at step 1. Then we present the proof of the law of large numbers and CLTs. In the latter proofs, we formulate and use the estimates related to step 3 without proving them. Only later in Sect. 6 do we introduce the necessary notation and combinatorial arguments which give those estimates.

We choose this way of presentation since it allows the readers to see the structure of the proofs without being distracted by rather heavy notation and quite lengthy technical arguments related to step 3.

From now on, we will often work conditionally on the set of non-extinction Ext^c , which will not be explicitly mentioned in the proofs (however, should be clear from the context).

5.2 Basic Facts on U - and V -Statistics

We will now briefly recall one of the standard tools of the theory of U -statistics, which we will use in the sequel, namely the Hoeffding decomposition.

Let us introduce for $I \subseteq \{1, \dots, n\}$ the Hoeffding projection of $f: \mathbb{R}^{nd} \rightarrow \mathbb{R}$ corresponding to I as the function $\Pi_I f: \mathbb{R}^{I|d} \rightarrow \mathbb{R}$, given by the formula

$$\Pi_I f((x_i)_{i \in I}) = \int_{\mathbb{R}^{nd}} \left(\prod_{i \notin I} \varphi(dy_i) \prod_{i \in I} (\delta_{x_i} - \varphi)(dy_i) \right) f(y_1, \dots, y_n). \quad (24)$$

One can easily see that for $|I| \geq 1$, $\Pi_I f$ is a canonical kernel. Moreover $\Pi_{\emptyset} f = \int_{\mathbb{R}^{nd}} f(x_1, \dots, x_n) \prod_{i=1}^n \varphi(dx_i)$.

Note that if f is symmetric (i.e., invariant with respect to permutations of arguments), $\Pi_I f$ depends only on the cardinality of I . In this case, we speak about the k th Hoeffding projection ($k = 0, \dots, n$), given by

$$\Pi_k f(x_1, \dots, x_k) = \left\langle (\delta_{x_1} - \varphi) \otimes \dots \otimes (\delta_{x_k} - \varphi) \otimes \varphi^{\otimes(n-k)}, f \right\rangle.$$

A symmetric kernel in n variables is called *degenerate of order $k - 1$* ($1 \leq k \leq n$) iff $k = \min\{i > 0: \Pi_i f \neq 0\}$. The order of degeneracy is responsible for the normalization and the form of the limit in the CLT for U -statistics, e.g., if the kernel is non-degenerate, i.e., $\Pi_1 f \neq 0$, then the corresponding U -statistic of an i.i.d. sequence behaves like a sum of independent random variables and converges to a Gaussian limit. The same phenomenon will be present also in our situation (see Sect. 7).

In the particular case $k = n$, the definition of the Hoeffding projection reads as

$$\Pi_n(f) := \left\langle \left(\bigotimes_{i=1}^n (\delta_{x_i} - \varphi) \right), f \right\rangle.$$

One easily checks that

$$f(x_1, \dots, x_n) = \sum_{I \subseteq \{1, \dots, n\}} \Pi_I f((x_i)_{i \in I}),$$

which gives us the aforementioned Hoeffding decomposition of U -statistics

$$U_t^n(f) = \sum_{I \subseteq \{1, \dots, n\}} \frac{(|X_t| - |I|)!}{(|X_t| - n)!} U_t^{|I|}(\Pi_I f),$$

which in the case of symmetric kernels simplifies to

$$U_t^n(f) = \sum_{k=0}^n \binom{n}{k} \frac{(|X_t| - k)!}{(|X_t| - n)!} U_t^k(\Pi_k f),$$

where we use the convention $U_t^0(a) = a$ for any constant a .

For technical reasons, we will also consider the notion of a V -statistic which is closely related to U -statistics, and is defined as

$$V_t^n(f) := \sum_{i_1, i_2, \dots, i_n=1}^{|X_t|} f(X_t(i_1), X_t(i_2), \dots, X_t(i_n)). \quad (25)$$

The corresponding Hoeffding decomposition is

$$V_t^n(f) = \sum_{I \subseteq \{1, \dots, n\}} |X_t|^{n-|I|} V_t^{|I|}(\Pi_I f), \quad (26)$$

where again we set $V_t^0(a) = a$ for any constant a .

In the proofs of our results, we will use a standard observation that a U -statistic can be written as a sum of V -statistics. More precisely, let $\mathcal{J}$ be the collection of partitions of $\{1, \dots, n\}$ i.e., of all sets $J = \{J_1, \dots, J_k\}$, where J_i 's are non-empty, pairwise disjoint and $\bigcup_i J_i = \{1, \dots, n\}$. For J as above let f_J be a function of $|J|$ variables

$x_1, \dots, x_{|J|}$, obtained by substituting x_i for all the arguments of f corresponding to the set J_i , e.g., for $n = 3$ and $J = \{\{1, 2\}, \{3\}\}$, $f_J(x_1, x_2) = f(x_1, x_1, x_2)$. An easy application of the inclusion–exclusion formula yields that

$$U_t^n(f) = \sum_{J \in \mathcal{J}} a_J V_t^{|J|}(f_J), \quad (27)$$

where a_J are some integers depending only on the partition J . Moreover one can easily check that if $J = \{\{1\}, \dots, \{n\}\}$, then $a_J = 1$, whereas if J consists of sets with at most two elements then $a_J = (-1)^k$ where k is the number of two-element sets in J . Let us also note that partitions consisting only of one- and two-element sets can be in a natural way identified with Feynman diagrams (defined in Sect. 2.1).

5.3 Proof of the Law of Large Numbers

Proof of Theorem 2 Consider the random probability measure $\mu_t = |X_t|^{-1} X_t$ (recall that formally we identify X_t with the corresponding counting measure). By Theorem 9 with probability one (conditionally on Ext^c), μ_t converges weakly to φ . Thus, by Theorem 3.2 in [3], $\mu_t^{\otimes n}$ converges weakly to $\varphi^{\otimes n}$.

Let f be bounded and continuous. We notice that $\langle f, \mu_t^{\otimes n} \rangle = |X_t|^{-n} V_t^n(f)$, which gives the almost sure convergence $|X_t|^{-n} V_t^n(f) \rightarrow \langle f, \varphi \rangle$. Now it is enough to note that the number of “off-diagonal” terms in the sum (25) defining $V_t^n(f)$ is of order $|X_t|^{n-1}$ and use the fact that $|X_t| \rightarrow \infty$ a.s. on Ext^c .

We note that in the proofs below we will use this fact only in the version for $f \in C(\mathbb{R}^{nd})$ which we have just proven. The proof of convergence in probability for $f \in \mathcal{P}(\mathbb{R}^{nd})$ follows directly from the CLT presented in Sect. 7. $\square$

5.4 Approximation

Before we proceed to the proofs of CLTs, we will demonstrate the simple fact that any function in $\mathcal{P}(\mathbb{R}^{nd})$ can be approximated by tensor functions.

Lemma 13 *Let $A := \{\otimes_{i=1}^n g_i : g_i \text{ bounded continuous}\}$ and $f \in \mathcal{P}(\mathbb{R}^{nd})$ be a canonical kernel. For every $m > 0$ there exists a sequence $\{f_k\} \subset \text{span}(A)$ such that each f_k is canonical and*

$$\|f_k(m \cdot) - f(m \cdot)\|_{\mathcal{P}} \rightarrow 0, \quad \text{as } k \rightarrow +\infty.$$

Proof First, we prove that $\text{span}(A)$ is dense in $\mathcal{P}(\mathbb{R}^{nd})$. Let us notice that given a function $f \in \mathcal{P}(\mathbb{R}^{nd})$ it suffices to approximate it uniformly on some box $[-M, M]^d$, $M > 0$. The box is a compact set and an approximation exists due to the Stone–Weierstrass theorem.

Now, let $f \in \mathcal{P}(\mathbb{R}^{nd})$. We may find a sequence $\{h_k\} \subset \text{span}(A)$ such that $h_k(m \cdot) \rightarrow f(m \cdot)$ in $\mathcal{P}$. Let us recall the Hoeffding projection (24) and denote $I = \{1, 2, \dots, n\}$. Now direct calculation (using exponential integrability of Gaussian variables) reveals that the sequence $f_k := \Pi_I h_k$ fulfills the conditions of the lemma. $\square$

5.5 Small Branching rate: Proof of Theorem 3

Let us first formulate two crucial facts, whose rather technical proofs we defer to Sect. 6.4. The first one corresponds to Step 2 in the outline of the proof presented in Sect. 5.1. Recall the definition of L_1 given in (14).

Proposition 14 *For any canonical $f \in \mathcal{P}(\mathbb{R}^{nd})$ we have $\mathbb{E}L_1(f)^2 < +\infty$. Moreover L_1 is a continuous function $L_1 : \text{Can} \mapsto L_2(\Omega, \mathcal{F}, \mathbb{P})$, where*

$$\text{Can} = \left\{ f \in \mathcal{P}(\mathbb{R}^{nd}) : f \text{ is a canonical kernel} \right\},$$

and Can is endowed with the norm $\|\cdot\|_{\mathcal{P}}$.

The other fact we will use allows for a uniform in t approximation of general canonical U -statistics by those, whose kernels are sums of tensor products. This corresponds to step 3 of the outline. Recall the distance m given by (5).

Proposition 15 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$ and $\lambda_p < 2\mu$. For any $n \geq 2$ there exists a function $l_n : \mathbb{R}_+ \mapsto \mathbb{R}_+$, fulfilling $\lim_{s \searrow 0} l_n(s) = 0$ and such that for any $f_1, f_2 \in \text{Can}$ and any $t > 1$ we have*

$$m(v_1, v_2) \leq l_n(\|f_1(2n\cdot) - f_2(2n\cdot)\|_{\mathcal{P}}),$$

where $v_1 \sim |X_t|^{-n/2} U_t^n(f_1)$, $v_2 \sim |X_t|^{-n/2} U_t^n(f_2)$ (the U -statistics are considered here conditionally on Ext^c).

We can now proceed with the proof of Theorem 3.

Proof of Theorem 3 For simplicity, we concentrate on the third coordinate. The joint convergence can be easily obtained by a straightforward modification of the arguments below (using the joint convergence in Theorem 10 for $n = 1$). In the whole proof, we work conditionally on the set of non-extinction Ext^c .

Let us consider bounded continuous functions $f_i^l : \mathbb{R}^d \mapsto \mathbb{R}$, $l = 1, \dots, m$, $i = 1, \dots, n$, which are centered with respect to φ and set $f_l := \otimes_{i=1}^n f_i^l$ and $f := \sum_{l=1}^m f_l$. In this case the U -statistic (2) writes as

$$U_t^n(f) = \sum_{l=1}^m \sum_{\substack{i_1, i_2, \dots, i_n=1, \\ i_j \neq i_k, \text{ for } j \neq k}}^{|X_t|} f_1^l(X_t(i_1)) f_2^l(X_t(i_2)) \dots f_n^l(X_t(i_n)).$$

Let γ be a Feynman diagram labeled by $\{1, 2, \dots, n\}$, with edges E_γ and unpaired vertices A_γ . Let

$$S_t(\gamma) := \sum_{\substack{i_1, i_2, \dots, i_n=1, \\ i_j=i_k \text{ if } (j,k) \in E_\gamma}}^{|X_t|} f(X_t(i_1), X_t(i_2), \dots, X_t(i_n)).$$

Decomposition (27) writes here as

$$U_t^n(f) = \sum_{\gamma} (-1)^{r(\gamma)} S_t(\gamma) + R_t. \quad (28)$$

where the sum spans over all Feynman diagrams labeled by $\{1, 2, \dots, n\}$ (note that when γ has no edges, then $S_t(\gamma) = V_t^n(f)$), and the remainder R is the sum of V -statistics corresponding to partitions of $\{1, \dots, n\}$ containing at least one set with more than two elements. First, we will prove that $|X_t|^{-(n/2)} R_t \rightarrow 0$.

To this end, let us consider a partition $J = \{A_r\}_{1 \leq r \leq m_1} \cup \{B_r\}_{1 \leq r \leq m_2} \cup \{C_r\}_{1 \leq r \leq m_3}$ of the set $\{1, 2, \dots, n\}$, in which $|A_r| \geq 3$, $|B_r| = 2$ and $|C_r| = 1$. Assume that $m_1 \geq 1$ and recall the definition of f_J used in (27). For any $l = 1, \dots, m$, we have

$$\begin{aligned} \frac{1}{|X_t|^{n/2}} V_t^{|J|}(f_J^l) &= \prod_{r \leq m_1} \left(|X_t|^{-|A_r|/2} \left\langle X_t, \prod_{k \in A_r} f_k^l \right\rangle \right) \prod_{r \leq m_2} \left(|X_t|^{-1} \left\langle X_t, \prod_{k \in B_r} f_k^l \right\rangle \right) \\ &\quad \times \prod_{r \leq m_3} \left(|X_t|^{-1/2} \left\langle X_t, \prod_{k \in C_r} f_k^l \right\rangle \right). \end{aligned}$$

By the first part of Theorem 9, the first product on the right-hand side converges almost surely to 0 and the second one converges to a finite limit. Each factor of the third product, by Theorem 10, converges (in law) to a Gaussian random variable. We conclude that $|X_t|^{-(n/2)} V_t^{|J|}(f_J^l) \xrightarrow{d} 0$. Thus, only the first summand of (28) is relevant for the asymptotics of $|X_t|^{-(n/2)} U_t^n(f)$.

Consider now

$$\begin{aligned} |X_t|^{-(n/2)} S_t(\gamma) &= \sum_{l=1}^m \prod_{(j,k) \in E_\gamma} \left(|X_t|^{-1} \sum_{i=1}^{|X_t|} f_j^l(X_t(i)) f_k^l(X_t(i)) \right) \\ &\quad \times \prod_{r \in A_\gamma} \left(|X_t|^{-1/2} \sum_{i=1}^{|X_t|} f_r^l(X_t(i)) \right). \end{aligned}$$

Let us denote $Z_{f_j^l}(t) := |X_t|^{-1/2} \sum_{i=1}^{|X_t|} f_j^l(X_t(i))$. By Theorem 10 and the Cramér-Wold device, we get that

$$(Z_{f_j^l}(t))_{1 \leq j \leq n, 1 \leq l \leq m} \xrightarrow{d} (G_{f_j^l})_{1 \leq j \leq n, 1 \leq l \leq m},$$

where $(G_{f_j^{l_1}})_{1 \leq j \leq n, 1 \leq l \leq m}$ is a centered Gaussian vector with the covariances

$$\text{Cov}(G_{f_j^{l_1}}, G_{f_k^{l_2}}) = \left\langle \varphi, f_j^{l_1} f_k^{l_2} \right\rangle + 2\lambda p \int_0^{+\infty} \left\langle \varphi, \left(e^{(\lambda p/2)s} \mathcal{T}_s f_j^{l_1} \right) \left(e^{(\lambda p/2)s} \mathcal{T}_s f_k^{l_2} \right) \right\rangle ds. \quad (29)$$

Let $D := \mathbb{R}_+ \times \mathbb{R}^d$ and note that

$$\begin{aligned} \text{Cov}(G_{f_j^{l_1}}, G_{f_k^{l_2}}) &= \int_D \mathcal{T}_t f_j^{l_1}(x) \mathcal{T}_t f_k^{l_2}(x) \mu_1(dt dx) \\ &= \int_D H(f_j^{l_1}) H(f_k^{l_2}) d\mu_1 \\ &= \text{Cov}(I_1(H(f_j^{l_1})), I_1(H(f_k^{l_2}))). \end{aligned}$$

(recall that I_1 is the Gaussian stochastic integral with respect to the random Gaussian measure with intensity μ_1). Thus, without loss of generality, we can assume that $G_{f_j^l} = I_1(H(f_j^l))$ for all l, j .

On the other hand by Theorem 9, one easily obtains

$$|X_t|^{-1} \sum_{i=1}^{|X_t|} f_j^l(X_t(i)) f_k^l(X_t(i)) \rightarrow \left\langle f_j^l f_k^l, \varphi \right\rangle, \quad a.s.$$

We conclude that

$$\left\{ |X_t|^{-(n/2)} S_t(\gamma) \right\}_\gamma \rightarrow^d \left\{ \sum_{l=1}^m \prod_{(j,k) \in E_\gamma} \left\langle f_j^l f_k^l, \varphi \right\rangle \prod_{r \in A_\gamma} G_{f_r^l} \right\}_\gamma =: \{w(\gamma)\}_\gamma.$$

Thus, by decomposition (28) and the considerations above, we obtain

$$|X_t|^{-(n/2)t} U_t^n(f) \rightarrow^d \sum_\gamma (-1)^{r(\gamma)} w(\gamma) =: L.$$

We will now show that L is equal to $L_1(f)$ given by (14). By linearity of $L_1(f)$ it is enough to consider the case of $m = 1$. We will therefore drop the superscript and write f_i instead of f_i^l . We recall (12) and denote $P(f_i, f_j) := \int_D H(f_i \otimes f_j)(z, z) \mu_2(dz)$, where $D = \mathbb{R}_+ \times \mathbb{R}^d$. By (29) and the definition of μ_2 given in (13)

$$\begin{aligned} L &= \sum_\gamma (-1)^{r(\gamma)} \prod_{(j,k) \in E_\gamma} \left\langle f_j f_k, \varphi \right\rangle \prod_{r \in A_\gamma} G_{f_r} \\ &= \sum_\gamma (-1)^{r(\gamma)} \prod_{(j,k) \in E_\gamma} (\mathbb{E} G_{f_j} G_{f_k} - P(f_j, f_k)) \prod_{r \in A_\gamma} G_{f_r}. \end{aligned}$$

We now adopt the notation that $\eta \subset \gamma$ when $E_\eta \subset E_\gamma$ and write

$$\begin{aligned} L &= \sum_{\gamma} (-1)^{r(\gamma)} \sum_{\eta \subset \gamma} (-1)^{r(\eta)} \prod_{(j,k) \in E_{\gamma} \setminus E_{\eta}} \mathbb{E} G_{f_j} G_{f_k} \prod_{(j,k) \in E_{\eta}} P(f_j, f_k) \prod_{r \in A_{\gamma}} G_{f_r} \\ &= \sum_{\eta} \prod_{(j,k) \in E_{\eta}} P(f_j, f_k) \sum_{\gamma \supset \eta} (-1)^{r(\gamma) - r(\eta)} \prod_{(j,k) \in E_{\gamma} \setminus E_{\eta}} \mathbb{E} G_{f_j} G_{f_k} \prod_{r \in A_{\gamma}} G_{f_r}. \end{aligned}$$

Let us notice that the inner sum can be written as

$$\sum_{\sigma} (-1)^{r(\sigma)} \prod_{(j,k) \in E_{\sigma}} \mathbb{E} G_{f_j} G_{f_k} \prod_{r \in A_{\sigma}} G_{f_r},$$

where σ runs over all Feynman diagrams on the set of vertices A_{η} . Thus, by [21, Theorem 3.4 and Theorem 7.26] this equals $I_{|A_{\eta}|} (H(\otimes_{i \in A_{\eta}} f_i))$ and in consequence

$$L = \sum_{\eta} \prod_{(j,k) \in E_{\eta}} P(f_j, f_k) I_{|A_{\eta}|} (H(\otimes_{i \in A_{\eta}} f_i)).$$

It is easy to see that in the case of $f = \otimes_{i=1}^n f_i$, the expression above is equivalent to (14). Thus, for each f , which is a finite sum of tensors, we have $L = L_1(f)$ and in consequence $|X_t|^{-n/2} U_t^n(f) \xrightarrow{d} L_1(f)$.

Let us now consider a general canonical function $f \in \mathcal{P}$. We put $h(x) := f(2nx)$. By Lemma 13 we may find a sequence of canonical functions $\{f_k\}_k \subset \text{span}(A)$ such that $f_k(2n \cdot) \rightarrow h$ in $\mathcal{P}$. Now by Proposition 15, we may approximate $|X_t|^{-(n/2)} U_t^n(f)$ with $|X_t|^{-(n/2)} U_t^n(f_k)$ uniformly in $t > 1$. This together with Proposition 14 and standard metric-theoretic considerations concludes the proof. $\square$

5.6 Critical Branching Rate: Proof of Theorem 4

For the critical case, we will need the following counterpart of Proposition 15, which will be proved in Sect. 6.5.

Proposition 16 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$ and $\lambda_p = 2\mu$. For any $n \geq 2$ there exists a function $l_n : \mathbb{R}_+ \mapsto \mathbb{R}_+$, fulfilling $\lim_{s \rightarrow 0} l_n(s) = 0$ and such that for any canonical $f_1, f_2 \in \mathcal{P}(\mathbb{R}^{nd})$ and any $t > 1$ we have*

$$m(v_1, v_2) \leq l_n(\|f_1(2n \cdot) - f_2(2n \cdot)\|_{\mathcal{P}}),$$

where $v_1 \sim (t|X_t|)^{-n/2} U_t^n(f_1)$, $v_2 \sim (t|X_t|)^{-n/2} U_t^n(f_2)$ (the U -statistics are considered here conditionally on Ext^c).

Proof of Theorem 4 As in the subcritical case, we will focus on the third coordinate.

The proof is slightly easier than the one of Theorem 3 as, because of larger normalization, the notion of U -statistics and V -statistics coincide in the limit. Indeed, let

us consider bounded continuous functions $f_1^l, f_2^l, \dots, f_n^l$, $l = 1, \dots, m$, which are centered with respect to φ and denote $f := \sum_{l=1}^m \otimes_{i=1}^n f_i^l$. By (28) we have

$$U_t^n(f) - V_t^n(f) = \sum_{\gamma \text{ with at least one edge}} (-1)^{r(\gamma)} S_t(\gamma) + R_t,$$

simply by the fact that the Feynman diagram without edges corresponds to $V_t^n(f)$. Analogously as in the proof of Theorem 3, we have $(t|X_t|)^{-n/2} R_t \xrightarrow{d} 0$. Let us now fix some diagram γ with at least one edge. Without loss of generality, we assume that $E_\gamma = \{(1, 2), (2, 3), \dots, (2k-1, 2k)\}$ for $k \geq 1$. We have

$$(t|X_t|)^{-n/2} S(\gamma) = \sum_{l=1}^m \prod_{i \leq k} \left((|X_t|t)^{-1} \left\langle X_t, f_{2i}^l f_{2i+1}^l \right\rangle \right) \prod_{i > 2k} \left((|X_t|t)^{-1/2} \left\langle X_t, f_i^l \right\rangle \right).$$

By Theorem 11 each of the factors in the second product converges in distribution, whereas by the first part of Theorem 2 each factor in the first product converges almost surely to 0, in consequence $(t|X_t|)^{-n/2} S(\gamma)$ converges in probability to 0, which shows that $(t|X_t|)^{-n/2} U_t^n(f)$ and $(t|X_t|)^{-n/2} V_t^n(f)$ are asymptotically equivalent.

Let us denote $Z_{f_j^l}(t) := (t|X_t|)^{-1/2} \sum_{i=1}^{|X_t|} f_j^l(X_t(i))$. By Theorem 11 and the Cramér–Wold device, we get that

$$(Z_{f_j^l}(t))_{1 \leq i \leq n, l \leq m} \xrightarrow{d} (G_{f_j^l})_{1 \leq i \leq n, l \leq m},$$

where $(G_{f_j^l})_{i \leq n, l \leq m}$ is centered Gaussian with the covariances given by (16). Thus,

$$(t|X_t|)^{-n/2} V_t^n(f) \xrightarrow{d} \sum_{l=1}^m \prod_{j=1}^n G_{f_j^l} = L_2(f),$$

where L_2 is defined by (17).

Now we pass to general canonical functions $f \in \mathcal{P}$. By Lemma 13, we can approximate f by canonical f_k from $\text{span}(A)$ in such a way that $\|f - f_k\|_{\mathcal{P}} \leq \|f(2n \cdot) - f_k(2n \cdot)\|_{\mathcal{P}} \rightarrow 0$. Thus, by Proposition 16, the law of $(t|X_t|)^{-1/2} U_t^n(f_k)$ converges to the one of $(t|X_t|)^{-1/2} U_t^n(f)$ as $k \rightarrow \infty$ uniformly in $t > 1$. Moreover by the fact that L_2 is bounded on $L_2(\mathbb{R}^{nd}, \varphi^{\otimes n})$ and there exists $C < \infty$ such that $\|\cdot\|_{L_2(\mathbb{R}^{dn}, \varphi^{\otimes n})} \leq C \|\cdot\|_{\mathcal{P}}$ (which follows easily from exponential integrability of Gaussian variables), we obtain $L_2(f_k) \rightarrow L_2(f)$ in the space $L_2(\Omega, \mathcal{F}, \mathbb{P})$. The proof may now be concluded by standard metric-theoretic arguments. $\square$

5.7 Large Branching Rate: Proof of Theorem 8

As in the previous two cases, we start with a fact, which allows to approximate general U -statistics, by those with simpler kernels. It is slightly different than the corresponding statements in the small and critical branching case, which is related to a different

type of convergence and a deterministic normalization which we have for large branching. The proof is deferred to Sect. 6.6.

Proposition 17 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system with $\lambda_p > 2\mu$. There exist constants $C, c > 0$ such that for any canonical $f \in \mathcal{P}(\mathbb{R}^{nd})$ we have*

$$\mathbb{E}_x \left(e^{-n(\lambda_p - \mu)t} |U_t^n(f)| \right) \leq C \exp \{c \|x\|\} \|f(2n \cdot)\|_{\mathcal{P}}.$$

Proof of Theorem 8 Again we concentrate on the third coordinate. The joint convergence can be easily obtained by a modification of the arguments below (using the joint convergence in Theorem 12 for $n = 1$).

First, note that U -statistics and V -statistics are asymptotically equivalent. The argument is analogous to the one presented in the proof of Theorem 4, since under the assumption $\lambda_p > 2\mu$ we have

$$\frac{|X_t|}{\exp(2(\lambda_p - \mu)t)} \rightarrow 0 \text{ a.s.}$$

as $t \rightarrow \infty$ and consequently we can disregard the sum over all multi-indices $(i_1, \dots, i_n)$ in which the coordinates are not pairwise distinct.

Let us consider bounded continuous functions $f_1^l, f_2^l, \dots, f_n^l: \mathbb{R}^d \rightarrow \mathbb{R}$, $l = 1, \dots, m$, which are centered with respect to φ and denote $f := \sum_{l=1}^m \otimes_{i=1}^n f_i^l$. By Theorem 12 for $n = 1$ we have

$$\begin{aligned} e^{-n(\lambda - \mu)t} V_t^n(f) &= \sum_{l=1}^m \prod_{i=1}^n \left(e^{-(\lambda_p - \mu)t} V_t^1(f_i^l) \right) \rightarrow \sum_{l=1}^m \prod_{i=1}^n \left(\langle \nabla f_i^l, \varphi \rangle \circ H_\infty \right) \\ &= L_3(f), \end{aligned}$$

in probability. Before our final step, we recall that the convergence in probability can be metrized by $d(X, Y) := \mathbb{E} \left(\frac{|X - Y|}{|X - Y| + 1} \right) \leq \|X - Y\|_1$. Let us now consider a function $f \in \mathcal{P}$. By Lemma 13 we may find a sequence of canonical functions $\{f_k\} \subset \text{span}(A)$ such that $f_k(2n \cdot) \rightarrow f(2n \cdot)$ in $\mathcal{P}$. Now by Proposition 17, we may approximate $e^{-n(\lambda_p - \mu)t} U_t^n(f)$ with $e^{-n(\lambda_p - \mu)t} U_t^n(f_k)$ uniformly in t in the sense of the metric d . Moreover, one can easily show that $\lim_{k \rightarrow +\infty} d(\tilde{L}_3(f_k), \tilde{L}_3(f)) = 0$. This concludes the proof. $\square$

6 Proofs of Technical Lemmas

We will now provide the proofs of the technical facts formulated in Sect. 5. The proofs are quite technical and require several preparatory steps. In what follows, we first recall some additional properties of the Ornstein–Uhlenbeck process, and then we introduce certain auxiliary combinatorial structures which will play a prominent role in the proofs.

6.1 Auxiliary Facts on the Ornstein–Uhlenbeck Process

The semigroup of the Ornstein–Uhlenbeck process can be represented by

$$\mathcal{T}_t f(x) = (g_t * f)(x_t), \quad x_t := e^{-\mu t} x, \quad (30)$$

where

$$g_t(x) = \left(\frac{\mu}{\pi \sigma_t^2} \right)^{d/2} \exp \left\{ -\frac{\mu}{\sigma_t^2} \|x\|^2 \right\}, \quad \sigma_t^2 := \sigma^2 (1 - e^{-2\mu t}). \quad (31)$$

Let us recall (10). We denote $\text{ou}(t) := \sqrt{1 - e^{-2\mu t}}$ and let $G \sim \varphi$. Then (30) can be written as

$$\begin{aligned} \mathcal{T}_t f(x) &= \int_{\mathbb{R}^d} f(x_t - y) g_t(y) dy \\ &= \int_{\mathbb{R}^d} f(xe^{-\mu t} + \text{ou}(t)y) \varphi(y) dy \\ &= \mathbb{E} f(xe^{-\mu t} + \text{ou}(t)G). \end{aligned} \quad (32)$$

We also denote

$$\mathcal{T}_s^\lambda := e^{\lambda s} \mathcal{T}_s.$$

6.2 Smoothing Things Out

It is well known that the Ornstein–Uhlenbeck semigroup increases the smoothness of a function. We will now introduce some simple auxiliary lemmas which quantify this statement and give bounds on the $\|\cdot\|_{\mathcal{P}}$ norms of derivatives of certain functions obtained from f by applying the Ornstein–Uhlenbeck semigroup on a subset of coordinates. Such bounds will be useful, since they will allow us to pass in the analysis to smooth functions.

Let $f \in \mathcal{P}(\mathbb{R}^{nd})$ and $I \subset \{1, 2, \dots, n\}$ with $|I| = k$. We define

$$\hat{f}_I(x_1, x_2, \dots, x_n) := \int_{\mathbb{R}^{kd}} f(z_1, z_2, \dots, z_n) \prod_{i \in I} g_1(x_i e^{-\mu} - y_i) dy_i, \quad (33)$$

where $z_i = y_i$ if $i \in I$, $z_i = x_i$ otherwise and g_1 is given by (31). We have

Lemma 18 *Let $f \in \mathcal{P}(\mathbb{R}^{nd})$ and $l \in \mathbb{N}$. Then for any $I \subset \{1, 2, \dots, n\}$ the function $\hat{f}_I$ is smooth with respect to coordinates in I . For any multi-index $\Lambda = (i_1, \dots, i_l) \subset \{1, \dots, nd\}^l$ such that $\{[i_j/d] : j = 1, \dots, l\} \subset I$ we have*

$$\left\| \frac{\partial^{|\Lambda|} \hat{f}_I}{\partial x_\Lambda} \right\|_{\mathcal{P}} \leq C \|f\|_{\mathcal{P}}, \quad (34)$$

where $C > 0$ does not depend on f and depends only on the parameters of the system (that is σ, μ, d, l, n). Moreover, when f is canonical, so is $\hat{f}_I$.

Proof Let us fix some I and Λ . Using (33) we get

$$K := \frac{\partial^{|\Lambda|} \hat{f}_I}{\partial x_\Lambda} = \int_{\mathbb{R}^{kd}} f(z_1, z_2, \dots, z_n) \frac{\partial^{|\Lambda|}}{\partial x_\Lambda} \prod_{i \in I} g_1(x_i e^{-\mu} - y_i) dy_i.$$

Therefore, by (4), the properties of the Gaussian density g_1 and easy calculations we arrive at

$$\begin{aligned} |K| &\leq \|f\|_{\mathcal{P}} \int_{\mathbb{R}^{kd}} n(z_1, z_2, \dots, z_n)^{-1} \left| \frac{\partial^{|\Lambda|}}{\partial x_\Lambda} \prod_{i \in I} g_1(x_i e^{-\mu} - y_i) \right| dy_i \\ &\leq C_\Lambda n(x_1, x_2, \dots, x_n)^{-1}, \end{aligned}$$

for some constant C_Λ . We recall that function n is defined after (3).

To prove (34), it is enough to take the maximum over all admissible pairs I, Λ .

Let us now assume that f is canonical. We would like to check that for any $j \in \{1, 2, \dots, n\}$ we have

$$\int_{\mathbb{R}^{nd}} \hat{f}_I(x_1, x_2, \dots, x_n) \varphi(x_j) dx_j = 0.$$

There are two cases, the first when $j \notin I$. Then we have

$$\begin{aligned} &\int_{\mathbb{R}^d} \hat{f}_I(x_1, x_2, \dots, x_n) \varphi(x_j) dx_j \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^{kd}} f(z_1, z_2, \dots, z_n) \prod_{i \in I} g_1(x_i e^{-\mu} - y_i) dy_i \varphi(z_j) dz_j \\ &= \int_{\mathbb{R}^{kd}} \left(\int_{\mathbb{R}^d} f(z_1, z_2, \dots, z_n) \varphi(z_j) dz_j \right) \prod_{i \in I} g_1(x_i e^{-\mu} - y_i) dy_i = 0. \end{aligned}$$

The second case is when $j \in I$. Then

$$\begin{aligned} &\int_{\mathbb{R}^d} \hat{f}_I(x_1, x_2, \dots, x_n) \varphi(x_j) dx_j \\ &= \int_{\mathbb{R}^{kd}} f(z_1, z_2, \dots, z_n) \prod_{i \in I \setminus \{j\}}^k g_1(x_i e^{-\mu} - y_i) dy_i \left(\int_{\mathbb{R}^d} g_1(x_j e^{-\mu} - y_i) \varphi(x_j) dx_j \right) dy_j \\ &= \int_{\mathbb{R}^{kd}} f(z_1, z_2, \dots, z_n) \prod_{i \in I \setminus \{j\}}^k g_1(x_i e^{-\mu} - y_i) dy_i \varphi(y_j) dy_j, \end{aligned}$$

where the second equality holds by the fact that φ is the invariant measure of the Ornstein–Uhlenbeck process. Now the proof reduces to the first case. $\square$

We will also need the following simple identity. We consider $\{x_i\}_{i=1,2,\dots,n}$, $\{\tilde{x}_i\}_{i=1,2,\dots,n}$. By induction one easily checks that the following lemma holds (we slightly abuse the notation here, e.g., $\frac{\partial}{\partial y_i}$ denotes the derivative in direction $\tilde{x}_i - x_i$ and $\int_a^b$ the integral over the segment $[a, b] \subset \mathbb{R}^d$).

Lemma 19 *Let f be a smooth function, then*

$$\begin{aligned} & \sum_{(\epsilon_1, \epsilon_2, \dots, \epsilon_n) \in \{0,1\}^n} (-1)^{\sum_{i=1}^n \epsilon_i} f(\tilde{x}_1 + \epsilon_1(x_1 - \tilde{x}_1), \tilde{x}_2 + \epsilon_2(x_2 - \tilde{x}_2), \dots, \tilde{x}_n + \epsilon_n(x_n - \tilde{x}_n)) \\ &= \int_{x_1}^{\tilde{x}_1} \int_{x_2}^{\tilde{x}_2} \dots \int_{x_n}^{\tilde{x}_n} \frac{\partial^n}{\partial y_1 \partial y_2 \dots \partial y_n} f(y_1, y_2, \dots, y_n) dy_n dy_{n-1} \dots dy_1. \end{aligned}$$

6.3 Bookkeeping of Trees

We will now introduce the “bookkeeping of trees” technique (for similar considerations see, e.g., [13, Section 2] or [4]), which via some combinatorics and introduction of auxiliary branching processes will allow us to pass from equations on the Laplace transform in the case of $n = 1$ to estimates of moments of V -statistics and consequently U -statistics, which will be crucial for proving Propositions 15, 16 and 17.

Our starting point is classical. We will use the equation on the Laplace transform of the branching process to obtain, via integration, recursive formulas for moments of V -statistics generated by tensors.

Recall thus (25). Let $f_1, f_2, \dots, f_n \in \mathcal{C}_c(\mathbb{R}^d)$ and $f_i \geq 0$. We would like to calculate

$$\mathbb{E}_x V_t^n (\otimes_{i=1}^n f_i) = \mathbb{E}_x \left(\prod_{i=1}^n \langle X_t, f_i \rangle \right). \quad (35)$$

Let $\Lambda \subset \{1, 2, \dots, n\}$, slightly abusing notation for $\alpha_i \geq 0, i \in \Lambda$ we denote

$$w_\Lambda(x, t, \alpha) := \mathbb{E}_x \left(\exp \left\{ - \sum_{i \in \Lambda} \alpha_i \langle X_t, f_i \rangle \right\} \right), \quad w_\Lambda(x, t) := \frac{\partial^{|\Lambda|}}{\partial \alpha_\Lambda} w(x, t, \alpha) \Big|_{\alpha=0}.$$

Note that this differentiation is valid by Proposition 1 and properties of the Laplace transform (e.g., [17, Chapter XIII.2]). By the calculations from Section 4.1. in [1] we know that

$$\begin{aligned}
 & w_{\Lambda}(x, t, \alpha) \\
 &= \mathcal{T}_t e^{-\sum_{i \in \Lambda} \alpha_i f_i}(x) + \lambda \int_0^t \mathcal{T}_{t-s} \left[p w_{\Lambda}^2(\cdot, s, \alpha) - w_{\Lambda}(\cdot, s, \alpha) + (1-p) \right] (x) ds.
 \end{aligned} \tag{36}$$

It is easy to check that

$$v_{\Lambda}(x, t) := \mathbb{E}_x \left(\prod_{i \in \Lambda} \langle X_t, f_i \rangle \right) = (-1)^{|\Lambda|} w_{\Lambda}(x, t). \tag{37}$$

Assume that $|\Lambda| > 0$. We denote by $P_1(\Lambda)$ all pairs (Λ_1, Λ_2) such that $\Lambda_1 \cup \Lambda_2 = \Lambda$ and $\Lambda_1 \cap \Lambda_2 = \emptyset$, and by $P_2(\Lambda) \subset P_1(\Lambda)$ pairs with an additional restriction that $\Lambda_1 \neq \emptyset$ and $\Lambda_2 \neq \emptyset$. Using (36) we easily check that

$$\begin{aligned}
 & \frac{\partial^{|\Lambda|}}{\partial \alpha_{\Lambda}} w_{\Lambda}(x, t, \alpha) = (-1)^{|\Lambda|} \mathcal{T}_t \left(\prod_{i \in \Lambda} f_i e^{-\sum_{i \in \Lambda} \alpha_i f_i} \right) (x) + \lambda \int_0^t \mathcal{T}_{t-s} \\
 & \times \left[p \sum_{(\Lambda_1, \Lambda_2) \in P_1(\Lambda)} \frac{\partial^{|\Lambda_1|}}{\partial \alpha_{\Lambda_1}} w_{\Lambda}(\cdot, s, \alpha) \frac{\partial^{|\Lambda_2|}}{\partial \alpha_{\Lambda_2}} w_{\Lambda}(\cdot, s, \alpha) - \frac{\partial^{|\Lambda|}}{\partial \alpha_{\Lambda}} w_{\Lambda}(\cdot, s, \alpha) \right] (x) ds.
 \end{aligned}$$

We evaluate it at $\alpha = 0$, (let us notice that $\frac{\partial^{|\Lambda_1|}}{\partial \alpha_{\Lambda_1}} w_{\Lambda}(x, s, 0) = \frac{\partial^{|\Lambda_1|}}{\partial \alpha_{\Lambda_1}} w_{\Lambda_1}(x, s, 0) = w_{\Lambda_1}(x, s)$), multiply both sides by $(-1)^{|\Lambda|}$ and use the definition of $v_{\Lambda}(x, t)$ to get

$$\begin{aligned}
 & v_{\Lambda}(x, t) \\
 &= \mathcal{T}_t \left(\prod_{i \in \Lambda} f_i \right) (x) + \lambda \int_0^t \mathcal{T}_{t-s} \left[p \sum_{(\Lambda_1, \Lambda_2) \in P_1(\Lambda)} v_{\Lambda_1}(\cdot, s) v_{\Lambda_2}(\cdot, s) - v_{\Lambda}(\cdot, s) \right] (x) ds \\
 &= \mathcal{T}_t \left(\prod_{i \in \Lambda} f_i \right) (x) + \lambda \int_0^t \mathcal{T}_{t-s} \left[p \sum_{(\Lambda_1, \Lambda_2) \in P_2(\Lambda)} v_{\Lambda_1}(\cdot, s) v_{\Lambda_2}(\cdot, s) + (2p-1) v_{\Lambda}(\cdot, s) \right] (x) ds.
 \end{aligned}$$

This can be easily transformed to (recall that $\mathcal{T}_s^{\lambda_p} f(x) = e^{\lambda_p s} \mathcal{T}_s f(x)$, $\lambda_p = \lambda(2p-1)$)

$$v_{\Lambda}(x, t) = \mathcal{T}_t^{\lambda_p} \left(\prod_{i \in \Lambda} f_i \right) (x) + p\lambda \int_0^t \mathcal{T}_{t-s}^{\lambda_p} \left[\sum_{(\Lambda_1, \Lambda_2) \in P_2(\Lambda)} v_{\Lambda_1}(\cdot, s) v_{\Lambda_2}(\cdot, s) \right] (x) ds. \tag{38}$$

The last formula is much easier to handle if written in terms of auxiliary branching processes. Firstly, we introduce the following notation. For $n \in \mathbb{N} \setminus \{0\}$ we denote by $\mathbb{T}_n$ the set of rooted trees described below. The root has a single offspring. All inner vertices (we exclude the root and the leaves) have exactly two offspring. For $\tau \in \mathbb{T}_n$,

by $l(\tau)$ we denote the set of its leaves. Each leaf $l \in l(\tau)$ is assigned a label, denoted by $\text{lab}(l)$, which is a non-empty subset of $\{1, 2, \dots, n\}$. The labels fulfill two conditions:

$$\bigcup_{l \in l(\tau)} \text{lab}(l) = \{1, 2, \dots, n\}, \quad \forall_{l_1, l_2 \in l(\tau)} (l_1 \neq l_2 \implies \text{lab}(l_1) \cap \text{lab}(l_2) = \emptyset).$$

In other words, the labels form a partition of $\{1, 2, \dots, n\}$. For a given $\tau \in \mathbb{T}_n$ let $i(\tau)$ denote the set of inner vertices (we exclude the root and the leaves), clearly $|i(\tau)| = |l(\tau)| - 1$ (as usual $|\cdot|$ denotes the cardinality). Let us identify the vertices of τ with $\{0, 1, 2, \dots, |\tau| - 1\}$ in such a way, that for any vertex i its parent, denoted by $p(i)$, is smaller. Obviously, this implies that 0 is the root and that the inner vertices have numbers in the set $\{1, 2, \dots, |\tau| - 1\}$. We denote also

$$s(\tau) := \{l \in l(\tau) : |\text{lab}(l)| = 1\}, \quad m(\tau) := \{l \in l(\tau) : |\text{lab}(l)| > 1\}, \quad (39)$$

leaves with singleton and non-singleton label sets, respectively.

Given $\tau \in \mathbb{T}_n$ and $t \in \mathbb{R}_+$ and $\{t_i\}_{i \in i(\tau)}$, we consider an Ornstein–Uhlenbeck branching walk on τ as follows. The initial particle is placed at time 0 at location x , it evolves up to the time $t - t_1$ according to the Ornstein–Uhlenbeck process and splits into two offspring, the first one is associated with the left branch of vertex 1 in tree τ and the second one with the right branch. Further each of them evolves independently until time $t - t_i$, where i is the first vertex in the corresponding subtree, when it splits and so on. At time t , the particles are stopped and their positions are denoted by $\{Y_i\}_{i \in l(\tau)}$ (the number of particles at the end is equal to the number of leaves). The construction makes sense provided that $t_i \leq t$ and $t_i \leq t_{p(i)}$ for all $i \in i(\tau)$ (which we implicitly assume). We define

$$\text{OU} \left(\otimes_{a=1}^n f_a, \tau, t, \{t_i\}_{i \in i(\tau)}, x \right) := \mathbb{E}_x \left(\prod_{a=1}^n f_a(Y_{j(a)}) \right), \quad (40)$$

where $j(a) = l \in l(\tau)$ is the unique leaf such that $a \in \text{lab}(l)$. We also define

$$S(\tau, t, x) := (p\lambda)^{|i(\tau)|} e^{\lambda p t} \left(\prod_{i \in i(\tau)} \int_0^t dt_i \right) \left(\prod_{i \in i(\tau)} e^{\lambda p t_i} 1_{t_i \leq t_{p(i)}} \right) \\ \text{OU} \left(\otimes_{a=1}^n f_a, \tau, t, \{t_i\}_{i \in i(\tau)}, x \right), \quad (41)$$

where we set $t_0 = t$. The reason to study the above objects becomes apparent by the following statement.

Proposition 20 *Let $\Lambda_n = \{1, 2, \dots, n\}$. The following identity holds*

$$v_{\Lambda_n}(x, t) = \sum_{\tau \in \mathbb{T}_n} S(\tau, t, x). \quad (42)$$

Proof The claim is a consequence of the identity

$$v_{\Lambda}(x, t) = \sum_{\tau \in \mathbb{T}(\Lambda)} S(\tau, t, x), \quad (43)$$

where $\Lambda \subset \{1, 2, \dots, n\}$ and $\mathbb{T}(\Lambda)$ is the set of trees, as $\mathbb{T}_{|\Lambda|}$, with the exception that the labels are in the set Λ . This identity in turn will follow by induction with respect to the cardinality of Λ . For $\Lambda = \{i\}$ Eq. (38) reads as $v_{\Lambda}(x, t) = e^{\lambda p t} \mathcal{T}_t f_i(x)$ (note that $P_2(\Lambda) = \emptyset$). The space $\mathbb{T}(\Lambda)$ contains only one tree, denoted by τ_s , consisting of the root and a single leaf labeled by $\{i\}$. We have $i(\tau) = \emptyset$ and $S(\tau_s, t, x) = e^{\lambda p t} \mathcal{T}_t f_i(x)$ so (43) follows.

Let now $|\Lambda| = k > 1$ and assume that (43) holds for all sets of cardinality at most $k - 1$. Apply again (38). Similarly as before, the first term corresponds to τ_s . Let p be the transition density of the Ornstein–Uhlenbeck semigroup. By induction the second term of (38) can be written as

$$\begin{aligned} & e^{\lambda p t} p \lambda \int_0^t e^{-\lambda p t_1} \mathcal{T}_{t-t_1} \left[\sum_{(\Lambda_1, \Lambda_2) \in P_2(\Lambda)} \sum_{\tau_1 \in \mathbb{T}(\Lambda_1)} \sum_{\tau_2 \in \mathbb{T}(\Lambda_2)} S(\tau_1, t_1, \cdot) S(\tau_2, t_1, \cdot) \right] (x) dt_1 \\ &= e^{\lambda p t} p \lambda \int_0^t \int_{\mathbb{R}^d} e^{-\lambda p t_1} p(t - t_1, x, y) \\ & \quad \times \left[\sum_{(\Lambda_1, \Lambda_2) \in P_2(\Lambda)} \sum_{\tau_1 \in \mathbb{T}(\Lambda_1)} \sum_{\tau_2 \in \mathbb{T}(\Lambda_2)} S(\tau_1, t_1, \cdot) S(\tau_2, t_1, \cdot) \right] (x) dy dt_1. \end{aligned}$$

By (41) the above expression equals

$$\begin{aligned} & \sum_{(\Lambda_1, \Lambda_2) \in P_2(\Lambda)} \sum_{\tau_1 \in \mathbb{T}(\Lambda_1)} \sum_{\tau_2 \in \mathbb{T}(\Lambda_2)} e^{\lambda p t} p \lambda \int_0^t \int_{\mathbb{R}^d} e^{-\lambda p t_1} p(t - t_1, x, y) \\ & \quad \left[(p \lambda)^{|i(\tau_1)|} e^{\lambda p t_1} \left(\prod_{i \in i(\tau_1)} \int_0^{t_1} dt_i \right) \left(\prod_{i \in i(\tau_1)} e^{\lambda p t_i} 1_{t_i \leq t_{p(i)}} \right) \right. \\ & \quad \text{OU} \left(\otimes_{i \in \Lambda_1} f_i, \tau_1, t_1, \{t_i\}_{i \in i(\tau_1)}, y \right) \\ & \quad (p \lambda)^{|i(\tau_2)|} e^{\lambda p t_1} \left(\prod_{i \in i(\tau_2)} \int_0^{t_1} dt_i \right) \left(\prod_{i \in i(\tau_2)} e^{\lambda p t_i} 1_{t_i \leq t_{p(i)}} \right) \\ & \quad \left. \text{OU} \left(\otimes_{i \in \Lambda_2} f_i, \tau_2, t_1, \{t_i\}_{i \in i(\tau_2)}, y \right) \right] dy dt_1. \quad (44) \end{aligned}$$

Now we create a new tree τ by setting τ_1 and τ_2 to be descendants of the vertex born from the root at time $t - t_1$ (thus this vertex is assigned the split time t_1). We

keep labels and the remaining split times unchanged. Consider the branching random walk on τ with the initial position of the first particle equal to x . Note that by the branching property the evolution of this process on subtrees τ_1 and τ_2 is conditionally independent given the evolution of the first particle up to time $t - t_1$. Thus, by the Markov property of the Ornstein–Uhlenbeck process, we can identify the branching random walk on τ_1 and τ_2 with the branching random walk on τ . More precisely we have

$$\int_{\mathbb{R}^d} p(t - t_1, x, y) \text{OU}(\otimes_{i \in \Lambda_1} f_i, \tau_1, t_1, \{t_i\}_{i \in i(\tau_1)}, y) \\ \times \text{OU}(\otimes_{i \in \Lambda_2} f_i, \tau_2, t_1, \{t_i\}_{i \in i(\tau_2)}, y) dy = \text{OU}(\otimes_{i \in \Lambda} f_i, \tau, t, \{t_i\}_{i \in i(\tau)}, x).$$

Using the Fubini theorem together with the equality $|i(\tau)| = |i(\tau_1)| + |i(\tau_2)| + 1$ we see that the summand corresponding to τ_1, τ_2 in (44) equals $S(\tau, t, x)$. It is also easy to check that the described correspondence is a bijection from the set of pairs (τ_1, τ_2) [as in (44)] to $\mathbb{T}_n \setminus \{\tau_s\}$ and therefore the expression (44) is equal to $\sum_{\mathbb{T}_n \setminus \{\tau_s\}} S(\tau, t, x)$, which ends the proof. $\square$

The calculations will be more tractable when we derive an explicit formula for $\{Y_i\}_{i \in l(\tau)}$. Let us recall the notation introduced in (32) and consider a family of independent random variables $\{G_i\}_{i \in \tau}$, such that $G_i \sim \varphi$ for $i \neq 0$ and $G_0 \sim \delta_x$. Recall also that $\text{ou}(t) = \sqrt{1 - e^{-2\mu t}}$. The following proposition follows easily from the construction of the branching walk on τ and (32).

Proposition 21 *Let $\{Y_i\}_{i \in l(\tau)}$ be positions of particles at time t of the Ornstein–Uhlenbeck process on tree τ with labels $\{t_i\}_{i \in i(\tau)}$. We have*

$$\{Y_i\}_{i \in l(\tau)} \sim \{Z_i\}_{i \in l(\tau)},$$

where for any $i \in l(\tau)$ we put

$$Z_i := \sum_{l \in P(i)} \text{ou}(t_{p(l)} - t_l) G_l e^{-\mu t_l} + \text{ou}(t_{p(i)}) G_i,$$

where $P(i) := \{\text{predecessors of } i\}$, by convention we set $t_0 = t$ and $\text{ou}(t_{p(0)} - t_0) = 1$.

We are now ready to prove an extended version of Proposition 20. This result will be instrumental in proving bounds needed to implement step 3 of the outline presented in Sect. 5.1.

Proposition 22 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$ and $f \in \mathcal{P}(\mathbb{R}^{nd})$. Then*

$$\mathbb{E}_x \sum_{i_1, i_2, \dots, i_n=1}^{|X_t|} f(X_t(i_1), X_t(i_2), \dots, X_t(i_n)) = \sum_{\tau \in \mathbb{T}_n} S(\tau, t, x), \quad (45)$$

where in (41) we extend the definition of OU in (40) by putting

$$\text{OU}(f, \tau, t, \{t_i\}_{i \in i(\tau)}, x) := \mathbb{E}_x f(Y_{j(1)}, Y_{j(2)}, \dots, Y_{j(n)}).$$

Moreover all the quantities above are finite.

Proof (Sketch) Using Proposition 20, Proposition 1, (35) and Lebesgue's monotone convergence theorem one may prove that (45) is valid for $f \equiv C$, $C > 0$. Using standard methods, we may drop the positivity assumption in (35) and (42). Therefore, by the Stone–Weierstrass theorem, linearity and Lebesgue's dominated convergence theorem, (45) is valid for any $f \in \mathcal{C}_c(\mathbb{R}^{nd})$. Let now $f \in \mathcal{P}(\mathbb{R}^{nd})$, $f \geq 0$. We notice that for any $\tau \in \mathbb{T}_n$ the expression $\text{OU}(f, \tau, t, \{t_i\}_{i \in i(\tau)}, x)$ is finite, which follows easily from Proposition 21. Further, one can find a sequence $\{f_k\}$ such that $f_k \in \mathcal{C}_c(\mathbb{R}^{nd})$, $f_k \geq 0$ and $f_k \nearrow f$ (pointwise). Appealing to Lebesgue's monotone convergence theorem yields that (45) still holds (and both sides are finite). To conclude, once more we remove the positivity condition. $\square$

As a simple corollary we obtain

Corollary 23 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system, then for any $n \geq 1$ there exists C_n such that*

$$\mathbb{E}|X_t|^n \leq C_n e^{n\lambda_p t}.$$

Proof We apply the above proposition with $f = 1$. Using the definition (41) and the inequality $|i(\tau)| \leq n - 1$ for $\tau \in \mathbb{T}_n$, it is easy to check that for any $t \in \mathbb{T}_n$ we have $S(\tau, t, x) \leq C_\tau e^{n\lambda_p t}$, for a certain constant depending only on τ and p, λ . $\square$

Let us recall the notation of (39). The following proposition will be crucial in proving moment estimates for V - and U -statistics.

Proposition 24 *For any $n > 0$ there exist $C, c > 0$, such that for any $\tau \in \mathbb{T}_n$, any split times $\{t_i\}_{i \in i(\tau)}$ and any canonical $f \in \mathcal{P}(\mathbb{R}^{nd})$ we have*

$$\text{OU}(f, \tau, t, \{t_i\}_{i \in i(\tau)}, x) \leq C \|f\|_{\mathcal{P}} \exp\{c \|x\|\} \exp\left(-\mu \sum_{i \in s(\tau)} t_{p(i)}\right). \quad (46)$$

Proof Let $k \leq n$. Without loss of generality we may assume that $I := \{1, 2, \dots, k\}$ are single numbers (i.e., $j(i) \in s(\tau)$) and $\{k + 1, \dots, n\}$ are multiple ones. Let us also assume for a moment that for $i \in \{1, 2, \dots, k\}$ we have $t_{p(j(i))} \geq 1$. Let Z_i and G_i be as in Proposition 21. We have $\mathbb{E}f(Y_{j(1)}, Y_{j(2)}, \dots, Y_{j(n)}) = \mathbb{E}f(Z_{j(1)}, Z_{j(2)}, \dots, Z_{j(n)})$. For $i \leq k$ we define

$$\tilde{Z}_i := \sum_{l \in P(i)} \text{ou}(t_{p(l)} - t_l) G_l e^{-\mu(t_l - 1)} + \text{ou}(t_{p(i)} - 1) G_i.$$

Moreover, let $\hat{f} := \hat{f}_I$ be given by (33). By the semigroup property of $\mathcal{T}$ we have

$$\mathbb{E}f(Z_{j(1)}, Z_{j(2)}, \dots, Z_{j(n)}) = \mathbb{E}\hat{f}(\tilde{Z}_{j(1)}, \dots, \tilde{Z}_{j(k)}, Z_{j(k+1)}, \dots, Z_{j(n)}) =: A.$$

By Lemma 18, $\hat{f}$ is smooth with respect to coordinates from I and canonical. Combining this with Lemma 19 we obtain

$$\begin{aligned} A &= \mathbb{E} \sum_{(\epsilon_1, \dots, \epsilon_k) \in \{0,1\}^k} (-1)^{\sum_{i=1}^n \epsilon_i} \hat{f}(\tilde{Z}_{j(1)} + \epsilon_1(G_{j(1)} - \tilde{Z}_{j(1)}), \dots, \tilde{Z}_{j(k)} \\ &\quad + \epsilon_k(G_{j(k)} - \tilde{Z}_{j(k)}), Z_{j(k+1)}, \dots, Z_{j(n)}) \\ &= \mathbb{E} \int_{G_{j(1)}}^{\tilde{Z}_{j(1)}} \dots \int_{G_{j(k)}}^{\tilde{Z}_{j(k)}} \frac{\partial^k}{\partial y_1 \dots \partial y_k} \hat{f}(y_1, \dots, y_k, Z_{j(k+1)}, \dots, Z_{j(n)}) dy_k \dots dy_1. \end{aligned}$$

From now on, we restrict to the case $d = 1$. The proof for general d proceeds along the same lines but it is notationally more cumbersome. Using Lemma 18 and applying the Schwarz inequality multiple times we have

$$\begin{aligned} |A| &\leq C \|f\|_{\mathcal{P}} \mathbb{E} \left| \left(\prod_{i=1}^k \int_{G_{j(i)}}^{\tilde{Z}_{j(i)}} \exp\{|y_i|\} dy_i \right) \prod_{i=k+1}^n \exp\{|Z_{j(i)}|\} \right| \\ &\leq C \|f\|_{\mathcal{P}} \prod_{i=1}^k \left\| \int_{G_{j(i)}}^{\tilde{Z}_{j(i)}} \exp\{|y_i|\} dy_i \right\|_{2^i} \prod_{i=k+1}^n \|\exp\{|Z_{j(i)}|\}\|_{2^i}. \end{aligned}$$

Note that by the definition of $\tilde{Z}_i$ we have $\tilde{Z}_i - G_i = H_i e^{-\mu(t_{p(i)}-1)} + (\text{ou}(t_{p(i)}-1) - 1) G_i$, where H_i is independent of G_i and $H_i \sim \mathcal{N}(x_i, \sigma_i^2)$ with $\sigma_i \leq \sigma/\sqrt{2\mu}$ and $\|x_i\| \leq \|x\|$. Thus, $\tilde{Z}_i - G_i$ is a Gaussian variable with the mean bounded by $C \|x_i\| e^{-\mu t_{p(i)}}$ and the standard deviation of order $e^{-\mu t_{p(i)}}$. In particular $\|\tilde{Z}_i - G_i\|_l \leq C_l \exp(C_l \|x\| - \mu t_{p(i)})$. Since

$$\int_{G_{j(i)}}^{\tilde{Z}_{j(i)}} \exp\{|y_i|\} dy_i \leq (e^{|G_{j(i)}|} + e^{|\tilde{Z}_{j(i)}|}) |\tilde{Z}_{j(i)} - G_{j(i)}|,$$

the proof can be concluded by yet another application of the Schwarz inequality and standard facts on exponential integrability of Gaussian variables.

Finally, if some i 's do not fulfill $t_{p(i)} \geq 1$, we repeat the above proof with $s(\tau)$ replaced by the set s' of indices from $s(\tau)$ for which additionally $t_{p(i)} \geq 1$. In this way, we obtain (46) with $\sum_{i \in s'} t_{p(i)}$ instead of $\sum_{i \in s(\tau)} t_{p(i)}$. In our setting

$$\sum_{i \in s(\tau)} t_{p(i)} - \sum_{i \in s'} t_{p(i)} = \sum_{i \in s(\tau) \setminus s'} t_{p(i)} \leq |s(\tau) \setminus s'| \leq n,$$

hence (46) still holds (with a worse constant C). $\square$

6.4 Small Branching Rate: Proofs of Propositions 14 and 15

Proof of Proposition 14 The sum (14) is finite hence it is enough to prove our claim for one $L(f, \gamma)$. Without loss of generality let us assume that $E_\gamma = \{(1, 2), (3, 4), (2k - 1, 2k)\}$ and $A_\gamma = \{2k + 1, \dots, n\}$ (we recall notation in Sect. 2.1). Using the same notation as in (13) we write

$$\begin{aligned} J(z_{2k+1}, \dots, z_n) &:= \left(\prod_{(j,k) \in E_\gamma} \int \mu_2(dz_{j,k}) \right) H(f)(u_1, u_2, \dots, u_n) \\ &= \int_D \int_D \dots \int_D H(f)(z_1, z_1, z_2, z_2, \dots, z_k, z_k, z_{2k+1}, \dots, z_n) \\ &\quad \times \mu_2(dz_1) \mu_2(dz_2) \dots \mu_2(dz_k). \end{aligned}$$

where $D := \mathbb{R}_+ \times \mathbb{R}^d$. We have $L(f, \gamma) = I_{n-2k}(J(x_{2k+1}, \dots, x_n))$. By the properties of the multiple stochastic integral [21, Theorem 7.26] we know that $\mathbb{E}L(f, \gamma)^2 \lesssim \left(\prod_{i \in \{2k+1, \dots, n\}} \int_D \mu_1(dz_i) \right) |J(z_{2k+1}, \dots, z_n)|^2$. Therefore, we need to estimate

$$\begin{aligned} &\int_D \int_D \dots \int_D H(f)(z_1^1, z_1^1, z_2^1, z_2^1, \dots, z_k^1, z_k^1, z_{2k+1}, \dots, z_n) \\ &\quad H(f)(z_1^2, z_1^2, z_2^2, z_2^2, \dots, z_k^2, z_k^2, z_{2k+1}, \dots, z_n) \\ &\quad \mu_2(dz_1^1) \mu_2(dz_1^2) \dots \mu_2(dz_k^1) \mu_2(dz_k^2) \mu_2(dz_{2k+1}^1) \dots \\ &\quad \mu_2(dz_{2k+1}^2) \mu_1(dz_{2k+1}) \dots \mu_1(dz_n) = (*). \end{aligned} \tag{47}$$

We will now estimate $H(f)(z_1, z_2, \dots, z_n)$. Let $Y_1, Y_2, \dots, Y_n$ be i.i.d., $Y_i \sim \varphi$. We define $Y_i(t) = x_i e^{-\mu t} + \text{ou}(t)Y_i$. Recall our notation $z_i = (s_i, x_i)$ and let $I = \{i \in \{1, \dots, n\} : s_i \geq 1\}$, $I^c = \{1, \dots, n\} \setminus I$. Let $\hat{f} := \hat{f}_I$ be defined according to (33).

Using (32), Lemmas 18, 19, the assumption that f is canonical and the semigroup property, we can rewrite (12) as

$$\begin{aligned} H(f)(z_1, z_2, \dots, z_n) &= \mathbb{E}f(Y_1(s_1), \dots, Y_n(s_n)) \\ &= \mathbb{E}\hat{f}((Y_i(s_i - 1))_{i \in I}, (Y_i(s_i))_{i \in I^c}) \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \sum_{(\epsilon_i)_{i \in I} \in \{0,1\}^I} (-1)^{\sum_{i \in I} \epsilon_i} \hat{f} \left(\left(Y_i(s_i-1) + \epsilon_i(Y_i - Y_i(s_i-1)) \right)_{i \in I}, (Y_i(s_i))_{i \in I^c} \right) \\
&= \mathbb{E} \left(\prod_{i \in I} \int_{Y_i}^{Y_i(s_i-1)} dy_i \right) \frac{\partial^{|I|}}{\partial y_I} \hat{f} \left((y_i)_{i \in I}, (Y_i(s_i))_{i \in I^c} \right).
\end{aligned}$$

By Lemma 18, we have (in order to simplify the notation, we calculate for $d = 1$, the general case is an easy modification)

$$\begin{aligned}
|H(f)(z_1, z_2, \dots, z_n)| &\lesssim \|f\|_{\mathcal{P}} \mathbb{E} \prod_{i \in I} \left| \int_{Y_i}^{Y_i(s_i-1)} e^{|y_i|} dy_i \right| \prod_{i \in I^c} e^{|Y_i(s_i)|} \\
&\leq \|f\|_{\mathcal{P}} \prod_{i \in I} \mathbb{E} \left| \int_{Y_i}^{Y_i(s_i-1)} (e^{y_i} + e^{-y_i}) dy_i \right| \prod_{i \in I^c} \mathbb{E} e^{|Y_i(s_i)|} \\
&\leq \|f\|_{\mathcal{P}} \left(\prod_{i \in I} (\mathbb{E} |\exp(Y_i(s_i-1)) - \exp(Y_i)| + \mathbb{E} |\exp(-Y_i(s_i-1)) - \exp(-Y_i)|) \right) \\
&\quad \times \mathbb{E} \prod_{i \in I^c} e^{|Y_i(s_i)|}.
\end{aligned}$$

For any $x, y \in \mathbb{R}$ we have $\max(\exp(x), \exp(y)) \leq \exp(x) + \exp(y)$. Therefore, by the mean value theorem, we get

$$\begin{aligned}
\mathbb{E} |\exp(Y_i(s_i-1)) - \exp(Y_i)| &\leq \mathbb{E} |Y_i(s_i-1) - Y_i| \exp(Y_i(s_i-1)) \\
&\quad + \mathbb{E} |Y_i(s_i-1) - Y_i| \exp(Y_i).
\end{aligned}$$

Using the Schwarz inequality and performing easy calculations, we get

$$\begin{aligned}
&\mathbb{E} |\exp(Y_i(s_i-1)) - \exp(Y_i)| \\
&\leq \sqrt{\mathbb{E}(Y_i(s_i-1) - Y_i)^2} \left(\sqrt{\mathbb{E} \exp(2Y_i(s_i-1))} + \sqrt{\mathbb{E} \exp(2Y_i)} \right) \lesssim \exp(2|x_i|) e^{-\mu s_i}.
\end{aligned}$$

Similarly

$$\mathbb{E} |\exp(-Y_i(s_i-1)) - \exp(-Y_i)| \lesssim \exp(2|x_i|) e^{-\mu s_i}.$$

Since we also have $\mathbb{E} \exp(|Y_i(s_i)|) \lesssim \exp(|x_i|)$, we have thus proved that

$$|H(f)(z_1, z_2, \dots, z_n)| \lesssim \|f\|_{\mathcal{P}} e^{2 \sum_{i=1}^n |x_i|} \prod_{i \in I} e^{-\mu s_i}. \quad (48)$$

We use the above inequality to estimate (47). To this end let us denote $D_0 = [0, 1) \times \mathbb{R}^d$, $D_1 = [1, \infty) \times \mathbb{R}^d$. To simplify the notation let us introduce the following

convention. For subsets $I_1, I_2 \subset \{1, \dots, k\}$, $I_3 \subset \{2k+1, \dots, n\}$ and $i \in \{1, \dots, n\}$ we will write $I_j(i) = 1$ if $i \in I_j$ and $I_j(i) = 0$ otherwise. Let us also denote

$$F(\mathbf{z}) = H(f)(z_1^1, z_1^1, z_2^1, z_2^1, \dots, z_k^1, z_k^1, z_{2k+1}^1, \dots, z_n^1) \\ \times H(f)(z_1^2, z_1^2, z_2^2, z_2^2, \dots, z_k^2, z_k^2, z_{2k+1}^2, \dots, z_n^2)$$

for $\mathbf{z} = (z_1^1, \dots, z_k^1, z_1^2, \dots, z_k^2, z_{2k+1}^1, \dots, z_n^1)$.

With this notation, we can estimate (47) as follows

$$(*) \leq \sum_{I_1, I_2 \subset \{1, \dots, k\}} \sum_{I_3 \subset \{2k+1, \dots, n\}} \left(\prod_{i=1}^k \int_{D_{I_1(i)}} \mu_2(dz_i^1) \right) \left(\prod_{i=1}^k \int_{D_{I_2(i)}} \mu_2(dz_i^2) \right) \\ \times \left(\prod_{i=2k+1}^n \int_{D_{I_3(i)}} \mu_1(dz_i) \right) F(\mathbf{z}) \\ =: \sum_{I_1, I_2 \subset \{1, \dots, k\}} \sum_{I_3 \subset \{2k+1, \dots, n\}} A(I_1, I_2, I_3).$$

Now, using (48) in combination with the Fubini theorem, the definition of the measures μ_i (given in Sect. 3.1) and our assumption $\lambda_p < 2\mu$, we get

$$A(I_1, I_2, I_3) \lesssim \|f\|_{\mathcal{P}}^2 \left(\prod_{i=1}^k \int_{D_{I_1(i)}} e^{4|x_i^1| - 2I_1(i)\mu s_i^1} \mu_2(dz_i^1) \right) \\ \times \left(\prod_{i=1}^k \int_{D_{I_2(i)}} e^{4|x_i^2| - 2I_2(i)\mu s_i^2} \mu_2(dz_i^2) \right) \\ \times \left(\prod_{i=2k+1}^n \int_{D_{I_3(i)}} e^{4|x_i| - 2I_3(i)\mu s_i} \mu_1(dz_i) \right) \lesssim \|f\|_{\mathcal{P}}^2.$$

To conclude the proof, we use the fact that $f \mapsto L(f, \gamma)$ is linear and $\|\cdot\|_{\mathcal{P}}$ is a norm. $\square$

Our next goal is to prove Proposition 15, which is the last remaining ingredient used in the proof of Theorem 3. This is where we will use for the first time the bookkeeping of trees technique introduced in Sect. 6.3. We will proceed in three steps. First we will obtain L_2 bounds on V -statistics with deterministic normalization (Proposition 25), then we will pass to L_1 bound of U -statistics with random normalization, restricted to the subset of the probability space, where $|X_t|$ is large (Corollary 26). Finally, we will obtain bounds on the distance between the distribution of two U -statistics (with

random normalization) in terms of the distance in $\mathcal{P}$ of the generating kernels (proof of Proposition 15).

Proposition 25 *Let $\{X_t\}_{t \geq 0}$ be the OU branching particle system with $\lambda_p < 2\mu$. There exist $C, c > 0$ such that for any canonical kernel $f \in \mathcal{P}(\mathbb{R}^{nd})$ we have*

$$\mathbb{E}_x \left(e^{-(n/2)\lambda_p t} V_t^n(f) \right)^2 \leq C \exp \{c \|x\|\} \|f\|_{\mathcal{P}}^2.$$

Proof We need to estimate

$$\mathbb{E}_x e^{-n\lambda_p t} \sum_{i_1, i_2, \dots, i_{2n} = 1}^{|X_t|} f(X_t(i_1), \dots, X_t(i_n)) f(X_t(i_{n+1}), \dots, X_t(i_{2n})).$$

Obviously the function $f \otimes f$ is canonical. Moreover, it is easy to check, that $\|f \otimes f\|_{\mathcal{P}(\mathbb{R}^{2nd})} \leq \|f\|_{\mathcal{P}(\mathbb{R}^{nd})}^2$.

By Proposition 22, it suffices to show that for each $\tau \in \mathbb{T}_{2n}$ there exist $C, c > 0$ such that for the function $f \otimes f$ and any $t > 0$ we have $e^{-n\lambda_p t} S(\tau, t, x) \leq C \exp \{c \|x\|\} \|f\|_{\mathcal{P}}^2$.

Let us fix $\tau \in \mathbb{T}_{2n}$ and denote by $P_1(\tau)$ and $P_2(\tau)$ the sets of inner vertices of τ with, respectively, one and two children in $s(\tau)$ [we recall (39)]. Set also $P_3(\tau) := i(\tau) \setminus (P_1(\tau) \cup P_2(\tau))$.

By the definition of $S(\tau, t, x)$, Proposition 24 and the assumption $\lambda_p < 2\mu$, we get

$$\begin{aligned} & e^{-n\lambda_p t} S(\tau, t, x) \\ & \lesssim e^{-(n-1)\lambda_p t} \left(\prod_{i \in i(\tau)} \int_0^t dt_i \right) \left(\prod_{i \in i(\tau)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \right) \text{OU}(f, \tau, t, \{t_i\}_{i \in i(\tau)}, x) \\ & \lesssim \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) e^{-(n-1)\lambda_p t} \times \left(\prod_{i \in i(\tau)} \int_0^t dt_i \right) \left(\prod_{i \in i(\tau)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \prod_{i \in s(\tau)} e^{-\mu t_{p(i)}} \right) \\ & \leq \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) \exp((-n-1) + |P_1(\tau)|/2 + |P_3(\tau)|)\lambda_p t) \\ & \quad \times \left(\prod_{i \in P_1(\tau)} \int_0^t e^{(\lambda_p - \mu)t_i - (\lambda_p t)/2} dt_i \right) \left(\prod_{i \in P_2(\tau)} \int_0^t e^{(\lambda_p - 2\mu)t_i} dt_i \right) \\ & \quad \times \left(\prod_{i \in P_3(\tau)} \int_0^t e^{\lambda_p(t_i - t)} dt_i \right) \\ & \lesssim \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) \exp((-n-1) + |P_1(\tau)|/2 + |P_3(\tau)|)\lambda_p t). \end{aligned}$$

To end the proof, it is thus sufficient to show that $|P_1(\tau)| + 2|P_3(\tau)| \leq 2n - 2$.

Note that $|P_1(\tau)| + 2|P_2(\tau)| = |s(\tau)|$ and $\sum_{i=1}^3 |P_i(\tau)| = |i(\tau)| = |l(\tau)| - 1$. Thus, $|P_1(\tau)| + 2|P_3(\tau)| = 2|l(\tau)| - 2 - |s(\tau)| = |l(\tau)| + |m(\tau)| - 2 \leq 2n - 2$ (recall that $m(\tau)$ denotes the set of leaves with multiple labels). This ends the proof. $\square$

The next corollary is a technical step toward the proof of Proposition 15. Since we would like to normalize the U -statistic by the random quantity $|X_t|^{n/2}$, we need to restrict the range of integration in the moment bound to the set on which $|X_t|$ is relatively large. It will not be an obstacle in the proof of Proposition 15, since the probability that $|X_t|$ is small will be negligible (on the set of non-extinction), which will allow us to pass from restricted L_1 estimates to bounds on the distance between distributions.

Corollary 26 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system with $\lambda_p < 2\mu$. There exist constants $C, c > 0$ such that for any canonical $f \in \mathcal{P}(\mathbb{R}^{nd})$ and $r \in (0, 1)$ we have*

$$\mathbb{E}_x \left(|X_t|^{-n/2} |U_t^n(f)| 1_{\{|X_t| \geq r e^{\lambda_p t}\}} \right) \leq C \exp \{c \|x\|\} r^{-n/2} \|f(2n \cdot)\|_{\mathcal{P}}.$$

Proof Let $\mathcal{J}$ be the collection of partitions of $\{1, \dots, n\}$, i.e., of all sets $J = \{J_1, \dots, J_k\}$, where J_i 's are non-empty, pairwise disjoint and $\bigcup_i J_i = \{1, \dots, n\}$. Using (27) and the notation introduced there, we have

$$|X_t|^{-n/2} U_t^n(f) = |X_t|^{-n/2} \sum_{J \in \mathcal{J}} a_J V_t^{|J|}(f_J), \quad (49)$$

where a_J are some integers depending only on the partition J . Since the cardinality of $\mathcal{J}$ depends only on n , it is enough to show that for each $J \in \mathcal{J}$ and some constants $C, c > 0$ we have

$$\mathbb{E}_x \left(|X_t|^{-n/2} |V_t^{|J|}(f_J)| 1_{\{|X_t| \geq r e^{\lambda_p t}\}} \right) \leq C \exp \{c \|x\|\} r^{-n/2} \|f(2n \cdot)\|_{\mathcal{P}}.$$

Let us thus consider $J = \{J_1, \dots, J_k\}$ and let us assume that among the sets J_i there are exactly l sets of cardinality 1, say $J_1, \dots, J_l$. We would like to use Proposition 25. To this end, we have to express $V_t^k(f_J)$ as a sum of V -statistics with canonical kernels. This can be easily done by means of Hoeffding's decomposition (26). Since f_J is already degenerate with respect to variables $x_1, \dots, x_l$, we get

$$|X_t|^{-n/2} V_t^k(f_J) = \sum_{I \subseteq \{l+1, \dots, k\}} |X_t|^{|I| - n/2} V_t^{|I^c|}(\Pi_{I^c} f_J), \quad (50)$$

where $I^c := \{1, \dots, k\} \setminus I$. Let us notice that $n \geq 2k - l$, so $n - k \geq k - l \geq |I|$, which gives

$$n - 2|I| \geq k - |I| = |I^c|. \quad (51)$$

We consider a single term of (50). We have

$$\begin{aligned}\mathbb{E}_x \left(|X_t|^{-n/2} |V_t^{I^c}| (\Pi_{I^c} f_J) | 1_{\{|X_t| \geq r e^{\lambda_p t}\}} \right) &\leq r^{-n/2+|I|} \\ \mathbb{E}_x \left(e^{-\lambda_p t(n-2|I|)/2} |V_t^{I^c}| (\Pi_{I^c} f_J) \right) &\leq r^{-n/2+|I|} \\ \mathbb{E}_x \left(e^{-\lambda_p t|I^c|/2} |V_t^{I^c}| (\Pi_{I^c} f_J) \right) &\leq C \exp \{c \|x\|\} r^{-n/2+|I|} \|\Pi_{I^c} f_J\|_{\mathcal{P}(\mathbb{R}^{|I^c|d})},\end{aligned}$$

for $I \neq \{1, \dots, k\}$, where in the third inequality we used Proposition 25. One can check that for any $n \geq 2$ there exists $C > 0$ such that for any I, J we have $\|\Pi_I f_J\|_{\mathcal{P}(\mathbb{R}^{|I|d})} \leq C \|f_J\|_{\mathcal{P}(\mathbb{R}^{|J|d})}$ and $\|f_J\|_{\mathcal{P}(\mathbb{R}^{|J|d})} \leq \|f(2n \cdot)\|_{\mathcal{P}(\mathbb{R}^{nd})}$. Therefore, it remains to bound the contribution from $I = \{1, \dots, k\}$ (in the case $l = 0$). But in this case $I^c = \emptyset$, so $|V_t^{I^c}|(\Pi_{I^c} f_J) = |\Pi_{I^c} f_J| = |\langle \varphi^{\otimes k}, f_J \rangle| \leq C \|f_J\|_{\mathcal{P}}$ and $\exp(-\lambda_p t(n-2|I|)) \leq 1$, which easily gives the desired estimate. $\square$

Proof of Proposition 15 Let us fix $g \in \text{Lip}(1)$. We consider

$$\begin{aligned}L &:= |\langle g, v_1 \rangle - \langle g, v_2 \rangle| \\ &= \mathbb{P}(Ext^c)^{-1} \left| \mathbb{E}_x 1_{Ext^c} g \left(|X_t|^{-n/2} U_t^n(f_1) \right) - \mathbb{E}_x 1_{Ext^c} g \left(|X_t|^{-n/2} U_t^n(f_2) \right) \right| \\ &\leq \mathbb{P}(Ext^c)^{-1} \mathbb{E}_x \left(1_{Ext^c} \left| g \left(|X_t|^{-n/2} U_t^n(f_1) \right) - g \left(|X_t|^{-n/2} U_t^n(f_2) \right) \right| \right).\end{aligned}$$

Let $h(x) := f_1(2nx) - f_2(2nx)$, take $r := \|h\|_{\mathcal{P}}^{1/n}$ and assume that $r < 1$. Then by Corollary 26 we get

$$\begin{aligned}\mathbb{E}_x \left(1_{Ext^c} \left| g \left(|X_t|^{-n/2} U_t^n(f_1) \right) - g \left(|X_t|^{-n/2} U_t^n(f_2) \right) \right| 1_{\{|X_t| \geq r e^{\lambda_p t}\}} \right) \\ \leq C \exp \{c \|x\|\} \|h\|_{\mathcal{P}}^{1/2}.\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbb{E}_x \left(1_{Ext^c} \left| g \left(|X_t|^{-n/2} U_t^n(f_1) \right) - g \left(|X_t|^{-n/2} U_t^n(f_2) \right) \right| 1_{\{|X_t| < r e^{\lambda_p t}\}} \right) \\ \leq 2 \|g\|_{\infty} \mathbb{P}(Ext^c \cap \{|X_t| e^{-\lambda_p t} < r\}).\end{aligned}$$

Since on Ext^c , we have $|X_t| \geq 1$ and $|X_t| e^{-\lambda_p t}$ converges to an absolutely continuous random variable,

$$\lim_{r \rightarrow 0+} \sup_t \mathbb{P}(Ext^c \cap \{|X_t| e^{-\lambda_p t} < r\}) = 0,$$

which ends the proof. $\square$

6.5 Critical Branching Rate: Proof of Proposition 16

As the proofs in this section follow closely the line of those in the subcritical case, we present only outlines, emphasizing differences.

Proposition 27 *Let $\{X_t\}_{t \geq 0}$ be the OU branching particle system with $\lambda_p = 2\mu$. There exist $C, c > 0$ such that for any canonical kernel $f \in \mathcal{P}(\mathbb{R}^{nd})$ and $t > 1$ we have*

$$\mathbb{E}_x \left(t^{-n/2} e^{-(n/2)\lambda_p t} V_t^n(f) \right)^2 \leq C \exp \{c \|x\|\} \|f\|_{\mathcal{P}}^2.$$

Proof We will use similar ideas as in the proof of Proposition 25 as well as the notation introduced therein. Consider any $\tau \in \mathbb{T}_{2n}$. By the definition of $S(\tau, t, x)$, Proposition 24 and the assumption $\lambda_p = 2\mu$, we obtain

$$\begin{aligned} & t^{-n} e^{-n\lambda_p t} S(\tau, t, x) \\ & \leq C_1 \|f\|_{\mathcal{P}}^2 \exp \{c \|x\|\} t^{-n} e^{-(n-1)\lambda_p t} \\ & \quad \times \left(\prod_{i \in i(\tau)} \int_0^t dt_i \right) \left(\prod_{i \in i(\tau)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \prod_{i \in s(\tau)} e^{-\mu t_{p(i)}} \right) \\ & \leq C_2 \|f\|_{\mathcal{P}}^2 \exp(c \|x\|) t^{-n} \exp \left(-(n-1) + |P_1(\tau)|/2 + |P_3(\tau)| \right) \lambda_p t \\ & \quad \times \left(\prod_{i \in P_1(\tau)} \int_0^t e^{(\lambda_p - \mu)t_i - (\lambda_p t)/2} dt_i \right) \left(\prod_{i \in P_2(\tau)} \int_0^t e^{(\lambda_p - 2\mu)t_i} dt_i \right) \\ & \quad \times \left(\prod_{i \in P_3(\tau)} \int_0^t e^{\lambda_p(t_i - t)} dt_i \right) \\ & \leq C_3 \|f\|_{\mathcal{P}}^2 \exp(c \|x\|), \end{aligned}$$

where we used the fact that $|P_2(\tau)| \leq n$ and the estimate $|P_1(\tau)| + 2|P_3(\tau)| \leq 2n - 2$ obtained in the proof of Proposition 25. $\square$

Now we can repeat the proof of Corollary 26 using Proposition 27 instead of Proposition 25 and obtain the following corollary, whose role is analogous to the one played by Corollary 26 in the slow branching case.

Corollary 28 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system with $\lambda_p = 2\mu$. There exist constants C, c such that for any canonical $f \in \mathcal{P}(\mathbb{R}^{nd})$ and $r \in (0, 1)$ we have for $t \geq 1$,*

$$\mathbb{E}_x \left((t|X_t|)^{-n/2} |U_t^n(f)| 1_{\{|X_t| \geq r e^{\lambda_p t}\}} \right) \leq C \exp \{c \|x\|\} r^{-n/2} \|f(2n \cdot)\|_{\mathcal{P}}.$$

Using the above corollary, we can now obtain Proposition 16 in an analogous way as we derived Proposition 15 from Corollary 26.

6.6 Supercritical Branching Case: Proof of Proposition 17

The proofs in this section diverge slightly from those in the critical and subcritical cases, and hence, we present more details.

Proposition 29 *Let $\{X_t\}_{t \geq 0}$ be the OU branching particle system with $\lambda_p > 2\mu$. There exist $C, c > 0$ such that for any canonical kernel $f \in \mathcal{P}(\mathbb{R}^{nd})$ we have*

$$\mathbb{E}_x \left(e^{-n(\lambda_p - \mu)t} V_t^n(f) \right)^2 \leq C \exp \{c \|x\|\} \|f\|_{\mathcal{P}}^2.$$

Proof As in the previous cases, consider any $\tau \in \mathbb{T}_{2n}$. We use the same notation as in the proof of Proposition 25. We have

$$\begin{aligned} e^{-2n(\lambda_p - \mu)t} S(\tau, t, x) &\leq C_1 e^{-2n(\lambda_p - \mu)t + \lambda_p t} \\ &\quad \times \left(\prod_{i \in i(\tau)} \int_0^t dt_i \right) \left(\prod_{i \in i(\tau)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \right) \text{OU}(f, \tau, t, \{t_i\}_{i \in i(\tau)}, x) \\ &\leq C_2 \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) e^{-2n(\lambda_p - \mu)t + \lambda_p t} \\ &\quad \times \left(\prod_{i \in i(\tau)} \int_0^t dt_i \right) \left(\prod_{i \in i(\tau)} e^{\lambda_p t_i} 1_{t_i \leq t_{p(i)}} \prod_{i \in s(\tau)} e^{-\mu t_{p(i)}} \right) \\ &\leq C_2 \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) e^{-2n(\lambda_p - \mu)t + \lambda_p t} \\ &\quad \times \left(\prod_{i \in P_1(\tau)} \int_0^t e^{(\lambda_p - \mu)t_i} dt_i \right) \left(\prod_{i \in P_2(\tau)} \int_0^t e^{(\lambda_p - 2\mu)t_i} dt_i \right) \left(\prod_{i \in P_3(\tau)} \int_0^t e^{\lambda_p t_i} dt_i \right) \\ &\leq C_3 \|f\|_{\mathcal{P}}^2 \exp(c\|x\|) \times \exp \left(-2n(\lambda_p - \mu)t + \lambda_p t + |P_1(\tau)|(\lambda_p - \mu)t \right. \\ &\quad \left. + |P_2(\tau)|(\lambda_p - 2\mu)t + |P_3(\tau)|\lambda_p t \right). \end{aligned}$$

Thus, it is enough to prove that

$$\lambda_p + |P_1|(\lambda_p - \mu) + |P_2|(\lambda_p - 2\mu) + |P_3|\lambda_p \leq 2n(\lambda_p - \mu), \quad (52)$$

where for simplicity we write P_i instead of $P_i(\tau)$ (in the rest of the proof we will use the same convention with other characteristics of τ). Using the equality $|s| = |P_1| + 2|P_2|$, we may rewrite (52) as

$$\lambda_p + |s|(\lambda_p - \mu) - |P_2|\lambda_p + |P_3|\lambda_p \leq 2n(\lambda_p - \mu),$$

so by the inequalities $2n \geq |s|$ and $\lambda_p - \mu > \lambda_p/2$ it is enough to prove that

$$2 + |s| - 2|P_2| + 2|P_3| \leq 2n.$$

But $|P_3| = |i| - |P_2| - |P_1|$ and so

$$\begin{aligned} 2 + |s| - 2|P_2| + 2|P_3| &= 2 + |s| + 2|i| - 4|P_2| - 2|P_1| = 2 - |s| + 2|i| \\ &= -|s| + 2|l| = 2|m| + |s| \leq 2n, \end{aligned}$$

which ends the proof. $\square$

Proof of Proposition 17 Using the notation from the proof of Corollary 26, we get

$$\begin{aligned} &\mathbb{E}_x e^{-n(\lambda_p - \mu)t} |V_t^k(f_J)| \\ &\leq \mathbb{E}_x \sum_{I \subseteq \{l+1, \dots, k\}} \frac{|X_t|^{|I|}}{e^{\lambda_p |I|t}} e^{-n(\lambda_p - \mu)t + \lambda_p |I|t} |V_t^{|I^c|}(\Pi_{I^c} f_J)| \\ &\leq \sum_{I \subseteq \{l+1, \dots, k\}} \left(\mathbb{E} \left(\frac{|X_t|^{|I|}}{e^{\lambda_p |I|t}} \right)^2 \right)^{1/2} \left(\mathbb{E}_x \left(e^{-n(\lambda_p - \mu)t + \lambda_p |I|t} V_t^{|I^c|}(\Pi_{I^c} f_J) \right)^2 \right)^{1/2}. \end{aligned}$$

Let us note that by $\lambda_p > 2\mu$ and (51) we get $n(\lambda_p - \mu) - \lambda_p |I| \geq n(\lambda_p - \mu) - 2(\lambda_p - \mu)|I| \geq |I^c|(\lambda_p - \mu)$. Thus, the summands on the right-hand side above for $I \neq \{1, \dots, k\}$ can be bounded using Corollary 23 and Proposition 29 by

$$C \exp(c\|x\|) \|\Pi_{I^c} f_J\|_{\mathcal{P}} \leq C^2 \exp(c\|x\|) \|f(2n\cdot)\|_{\mathcal{P}}$$

(the last inequality is analogous as in the proof of Corollary 26).

The contribution from $I = \{1, \dots, k\}$ (in the case $l = 0$) also can be bounded like in Corollary 26. Namely, $I^c = \emptyset$, so $|V_t^{|I^c|}(\Pi_{I^c} f_J)| = |\Pi_{I^c} f_J| = |\langle \varphi^{\otimes k}, f_J \rangle| \leq C \|f_J\|_{\mathcal{P}} \leq C^2 \|f(2n\cdot)\|_{\mathcal{P}}$ and $\exp(-n(\lambda_p - \mu)t + \lambda_p |I|t) \leq \exp(-(n - 2|I|)(\lambda_p - \mu)) \leq 1$, which easily gives the desired estimate. $\square$

7 Remarks on the Non-degenerate Case

As in the case of U -statistics of i.i.d. random variables, by combining the results for completely degenerate U -statistics with the Hoeffding decomposition, we can obtain limit theorems for general U -statistics, with normalization, which depends on the order of degeneracy of the kernel. For instance, in the slow branching case Theorem 3, the Hoeffding decomposition and the fact that $\Pi_k : \mathcal{P}(\mathbb{R}^{nd}) \rightarrow \mathcal{P}(\mathbb{R}^{kd})$ is continuous, give the following

Corollary 30 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Assume that $\lambda_p < 2\mu$ and let $f \in \mathcal{P}(\mathbb{R}^{nd})$ be symmetric and degenerate of order $k - 1$. Then conditionally on Ext^c , $|X_t|^{-(n-k/2)} U_t^n(f - \langle f, \varphi^{\otimes n} \rangle)$ converges in distribution to $\binom{n}{k} L_1(\Pi_k f)$.*

Similar results can be derived in the remaining two cases. Using the fact that on the set of non-extinction $|X_t|$ grows exponentially in t , we obtain

Corollary 31 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Assume that $\lambda_p = 2\mu$ and let $f \in \mathcal{P}(\mathbb{R}^{nd})$ be symmetric and degenerate of order $k - 1$. Then conditionally on Ext^c , $t^{-k/2} |X_t|^{-(n-k/2)} U_t^n(f - \langle f, \varphi^{\otimes n} \rangle)$ converges in distribution to $\binom{n}{k} L_2(\Pi_k f)$.*

Similarly, using (8) and the definition of W we obtain

Corollary 32 *Let $\{X_t\}_{t \geq 0}$ be the OU branching system starting from $x \in \mathbb{R}^d$. Assume that $\lambda_p > 2\mu$ and let $f \in \mathcal{P}(\mathbb{R}^{nd})$ be symmetric and degenerate of order $k - 1$. Then conditionally on Ext^c , $\exp(-(\lambda_p n - \mu k)t) U_t^n(f - \langle f, \varphi^{\otimes n} \rangle)$ converges in probability to $\binom{n}{k} W^{n-k} L_3(\Pi_k f)$.*

Since in all the corollaries above the normalization is strictly smaller than $|X_t|^n$, they in particular imply that $|X_t|^{-n} U_t^n(f - \langle f, \varphi^{\otimes n} \rangle) \rightarrow 0$ in probability, which proves the second part of Theorem 9 (as announced in Sect. 5.3).

Acknowledgments Research of R.A. was partially supported by the MNiSW grant N N201 397437 and by the Foundation for Polish Science. Research of P.M. was partially supported by the MNiSW grant N N201 397537. The authors would like to thank the referees for their constructive remarks.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Adamczak, R., Miłoś, P.: CLT for Ornstein–Uhlenbeck branching particle system. arXiv:1111.4559 (2011)
- Athreya, K.B., Ney, P.E.: Branching Processes. Springer, New York (1972). Die Grundlehren der mathematischen Wissenschaften, Band 196
- Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1999)
- Birkner, M., Zähle, I.: A functional CLT for the occupation time of state-dependent branching random walk. Ann. Probab. **35**(6), 2063–2090 (2007)
- Borisov, I.S., Volodko, N.V.: Orthogonal series and limit theorems for canonical U - and V -statistics of stationarily connected observations. Mat. Tr. **11**(1), 25–48 (2008)
- Borovkova, S., Burton, R., Dehling, H.: Limit theorems for functionals of mixing processes with applications to U -statistics and dimension estimation. Trans. Am. Math. Soc. **353**(11), 4261–4318 (2001). (electronic)
- de la Peña, V.H., Giné, E.: Decoupling. Probability and Its Applications. From Dependence to Independence. Randomly Stopped Processes. U -statistics and Processes. Martingales and Beyond. Springer, New York (1999)
- Dehling, H., Taqqu, M.S.: The limit behavior of empirical processes and symmetric statistics for stationary sequences. In: Proceedings of the 46th Session of the International Statistical Institute, vol. 4 (Tokyo, 1987), vol. 52, pp. 217–234 (1987)
- Del Moral, P., Doucet, A., Peters, G.W.: Sharp propagation of chaos estimates for Feynman–Kac particle models. Teor. Veroyatn. Primen. **51**(3), 552–582 (2006)
- Del Moral, P., Patras, F., Rubenthaler, S.: Tree based functional expansions for Feynman–Kac particle models. Ann. Appl. Probab. **19**(2), 778–825 (2009)
- Del Moral, P., Patras, F., Rubenthaler, S.: Convergence of U -statistics for interacting particle systems. J. Theor. Probab. **24**(4), 1002–1027 (2011)
- Dudley, R.M.: Real Analysis and Probability, volume 74 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge. Revised reprint of the 1989 original (2002)
- Dynkin, E.B.: Representation for functionals of superprocesses by multiple stochastic integrals, with applications to self-intersection local times. Astérisque (157–158), 147–171 (1988). Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987)

14. Dynkin, E.B.: An Introduction to Branching Measure-Valued Processes, volume 6 of CRM Monograph Series. American Mathematical Society, Providence, RI (1994)
15. Dynkin, E.B., Mandelbaum, A.: Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Stat.* **11**(3), 739–745 (1983)
16. Engländer, J., Harris, S.C., Kyprianou, A.E.: Strong law of large numbers for branching diffusions. *Ann. Inst. H. Poincaré* **46**, 279–298 (2010)
17. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. II, 2nd edn. Wiley, New York (1971)
18. Filippova, A.A.: The theorem of von Mises on limiting behaviour of functionals of empirical distribution functions and its statistical applications. *Teor. Veroyatnost. i Primenen.* **7**, 26–60 (1962)
19. Halmos, P.R.: The theory of unbiased estimation. *Ann. Math. Stat.* **17**, 34–43 (1946)
20. Hoeffding, W.: A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19**, 293–325 (1948)
21. Janson, S.: Gaussian Hilbert Spaces, volume 129 of Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge (1997)
22. Lee, A.J.: *U*-statistics: Theory and Practice, volume 110 of Statistics: Textbooks and Monographs. Marcel Dekker Inc., New York (1990)
23. Rubenthaler, S.: Expansion of the propagation of chaos for bird and Nanbu systems. Preprint. Available at <http://arxiv.org/abs/0901.3476>
24. Rubin, H., Vitale, R.A.: Asymptotic distribution of symmetric statistics. *Ann. Stat.* **8**(1), 165–170 (1980)
25. Rudin, W.: Functional Analysis. McGraw-Hill Series in Higher Mathematics. McGraw-Hill Book Co, New York (1973)