# FOUR MOMENTS THEOREMS ON MARKOV CHAOS

S. BOURGUIN, S. CAMPESE, N. LEONENKO, AND M.S. TAQQU

Abstract. We obtain quantitative Four Moments Theorems establishing convergence of the laws of elements of a Markov chaos to a Pearson distribution, where the only assumption we make on the Pearson distribution is that it admits four moments. These results are obtained by first proving a general carré du champ bound on the distance between laws of random variables in the domain of a Markov diffusion generator and invariant measures of diffusions, which is of independent interest, and making use of the new concept of chaos grade. For the heavy-tailed Pearson distributions, this seems to be the first time that sufficient conditions in terms of (finitely many) moments are given in order to converge to a distribution that is not characterized by its moments.

## 1. Introduction

Four Moments Theorems are results which imply or characterize convergence in law of some approximating sequence $\{F_k : k \geq 0\}$ of random variables towards some target measure $\nu$. A typical example of such an approximating sequence (with the target measure $\nu$ being Gaussian) are homogeneous sums of the form

$$F_k = \sum_{j_1, \ldots, j_p = 1}^{k} a_{j_1 \cdots j_p} W_{j_1} \cdots W_{j_p}, \tag{1}$$

normalized to have unit variance. Here, $\left\{W_j : j \geq 1\right\}$ is an i.i.d. sequence of standard Gaussian random variables and the constants $a_{j_1 \cdots j_p}$ are symmetric in the indices and vanish on diagonals. The classical fourth moment theorem of Nualart and Peccati (see [NP05]) states that $F_k$ converges in law to a standard Gaussian distribution if and only if the fourth moment of $F_k$ converges to the fourth moment of the standard Gaussian distribution, namely 3. In fact, the aforementioned two authors have proven their result in infinite dimensions, where the sequence of $F_k$ are sequences of multiple Wiener-Itō integrals of fixed order $p$. The original proof in [NP05] uses stochastic analysis and shortly after its publication another proof via Malliavin calculus was given by Nualart and Ortiz-Latorre in [NOL08]. Later, in [NP09a], Nourdin and Peccati used this approach to obtain a similar result for approximation of the Gamma distribution. They showed, again for a sequence of normalized Wiener-Itō integrals, that convergence of the third and fourth moments is enough to converge to a Gamma distribution.

2010 *Mathematics Subject Classification.* 60F05, 60J35, 60J99.

*Key words and phrases.* Markov operator, Diffusion generator, Gamma calculus, Pearson distributions, Stein's method, Malliavin-Stein, Limit theorems.

S. Campese was supported by InterMobility grant LILAC of the Luxembourg National Research Fund. M.S. Taqqu and N. Leonenko were supported by Cardiff Incoming Visiting Fellowship Scheme. N. Leonenko was supported partly by Australian Research Council's Discovery Projects funding scheme (project DP160101366) and by project MTM2015-71839-P of MINECO, Spain (cofunded with FEDER funds).

In [Led12], Ledoux provided a new perspective. He gave new proofs of the above results from the abstract point of view of Markov diffusion generators. In this context, given a Markov diffusion generator satisfying a certain spectral condition, convergence of a sequence of eigenfunctions of such a generator to a Gaussian or Gamma distribution is still controlled by convergence of just the first four moments. Multiple Wiener-Itō integrals fit the framework as they are eigenfunctions of the infinite-dimensional Ornstein-Uhlenbeck operator.

Building on [Led12], Azmoodeh et al. showed in [ACP14] that the spectral condition can be replaced with a Markov chaos property of the eigenfunctions which is less restrictive than the earlier notion of Markov chaos introduced in [Led12]. In addition to Four Moments Theorems for convergence towards the Gaussian and Gamma distributions, a Four Moments Theorem for the approximation of the Beta distribution was proven.

In this paper, we derive bounds on probabilistic distances $d(G_k, Z)$, where $\{G_k : k \geq 1\}$ is a sequence of random variables related to Markov diffusion generators and a target random variable $Z$ whose distribution is an invariant measure of a diffusion.

When the target distribution is viewed from the point of view of diffusion theory, it is interesting to note that the Gaussian, Gamma and Beta distributions share a common feature: they are the only invariant measures of a diffusion on the real line admitting an orthonormal basis of polynomial eigenfunctions (see [Maz97]). They also are the only members of the Pearson family of distributions (introduced in [Pea95], see for example [FS08] for a modern treatment) which have moments of all orders. This naturally leads to the question whether Four Moments Theorems can also be proven when the target measure $\nu$ is one of the three remaining heavy-tailed classes of Pearson distributions, commonly known as skew-t, F- and inverse Gamma-distributions (see Subsection 2.3 for details). Here, heavy-tailed is understood in the sense that only a finite number of moments exist.

Target distributions $\nu$ belonging to the Pearson family (or more generally absolutely continuous invariant measures of diffusions) have already been considered in [EVq15, KT12] as possible limit laws $\nu$ for sequences of multiple Wiener-Itō integrals. These integrals have the infinite dimensional Ornstein-Uhlenbeck as underlying Markov generator. For such multiple Wiener-Itō integrals, however, as was also observed in [KT12], the only possible limit distributions belonging to the Pearson family are the Gaussian and Gamma laws.

In this paper, we present a systematic approach to the problem and prove quantitative Four Moments Theorems for convergence to all six classes of the Pearson distribution. The only assumption we make, which seems to be unavoidable in this context, is that the parameters of the distribution are chosen in such a way that the first four moments exist. Compared to [ACP14], we are not only able to cover the full Pearson class as a target distribution $\nu$, but also extend the admissible chaos structures, so that, for example, the laws of the converging sequence of chaotic random variables can be heavy-tailed as well. In particular, no assumption of hypercontractivity or diagonalizability of the underlying generator is made. Our main result (Theorem 3.9 along with Proposition 3.12, to which we refer for full details and any unexplained notation) is a quantitative Four Moments Theorem of the form

$$d(G_k, Z) \leq c\sqrt{\int_E P(G_k)\,\mathrm{d}\,\mu_k + \xi_k \int_E Q(G_k)\,\mathrm{d}\,\mu_k}, \tag{2}$$

where $d$ is a suitable probabilistic distance metrizing convergence in law, $Z$ is a random variable whose law $\nu$ belongs to the Pearson family, $G_k$ is a chaotic random variable defined below with some underlying Markov diffusion generator $L_k$ with invariant measure $\mu_k$ for each $k \geq 0$. Moreover, in (2), $c$ is a positive constant, $P$ and $Q$ are polynomials of degree four, whose coefficients are explicitly given in terms of the parameters of the law of $Z$. In comparison to earlier Four Moments Theorems, the linear combination of moments, given by the integral involving the polynomial $Q$ appearing in the bound on the right hand side of (2) is new (and only appears in certain cases). The deterministic non-negative real sequence $\{\xi_k : k \geq 0\}$ in (2) is defined in terms of a new notion of *chaos grade*, which, heuristically speaking, measures how similar the chaotic sequence $\{G_k : k \geq 0\}$ is to the target random variable $Z$, when the latter is viewed as an element of the Markov chaos of a Pearson generator.

To prove (2), we first obtain a generic bound of the form

$$d(G, Z) \leq c \int_E \left|\Gamma(G, -L^{-1}G) - \tau(G)\right| \mathrm{d}\,\mu \tag{3}$$

(see Theorem 3.2) for probabilistic distances $d(G, Z)$ between a target random variable $Z$ whose law can belong to a large class of absolutely continuous distributions, and a random variable $G$ coming from a Markov structure which involves a generator $L$ with invariant measure $\mu$, the *carré du champ* operator $\Gamma$, the pseudo inverse $L^{-1}$ of the underlying Markov generator $L$ and a function $\tau$ related to the target $Z$. Again, we stress that both, the laws of $Z$ and $G$ do not need to have moments of all orders. The bound (3) is of independent interest and obtained using a combination of Stein's method and the so-called Gamma calculus. It can be seen as an abstract version of the *Malliavin-Stein method* on Wiener chaos, first introduced in [NP09b].

Then, in order to further bound (3) by the right hand side of (2) when $G$ is a chaotic element of the Markov structure and the law of $Z$ belongs to the Pearson family, we again make use of the Gamma calculus and spectral arguments that, in a similar spirit as in [ACP14], allow us to obtain (2), and hence linear combinations of the first four moments as a bound for the right hand side of (3).

Note that, in general, one cannot a priori use moments to prove convergence towards a heavy-tailed distribution. However, our results provide a context in which this is not only possible, but where convergence of only the first four moments suffices.

As particular examples of Markov structures fitting our framework, we study (tensorized) Pearson generators, which have multivariate Pearson distributions as invariant measures. In this context, the chaos grade provides a heuristic for the question which Pearson laws are compatible with each other, in the sense that chaotic random variables (for example homogeneous sums of the type (1) with the Gaussian laws replaced by arbitrary other Pearson laws with finite first four moments) with respect to one Pearson generator can converge in distribution to the invariant measure of another (possibly different) Pearson generator.

The paper is organized as follows. In Section 2, we introduce the Markov framework we will be working in and give a quick summary of Stein's method as well as an overview of the Pearson distributions. Our main results, in particular the bounds (3) and (2) as well as the definition of Markov chaos are presented in Section 3. As an application, we study in Section 4 the case of Pearson chaos, whose chaos structure fits our framework.

As a last remark, let us mention that speaking of Four Moments Theorems as opposed to a Fourth Moment or generally a Third and Fourth Moment is merely a question of style, depending on whether one normalizes the approximating sequences to have the correct mean and variance or not. We chose not to impose any normalization.

## 2. Preliminaries

2.1. **Markov diffusion generators.** Our main results will be proven in the setting of Markov diffusion generators, that is, we have some underlying diffusive Markov process $\{X_t : t \geq 0\}$ with invariant measure $\mu$, associated semigroup $\{P_t : t \geq 0\}$, infinitesimal generator $L$ and carré du champ $\Gamma$, where all of these objects are inherently connected. The operators $L$ and $\Gamma$ play an important role here. From an abstract point of view, a standard and elegant way to introduce this setting is through so called Markov triples, where one starts from the invariant measure $\mu$, the carré du champ $\Gamma$ and a suitable algebra of functions (random variables), from which the generator $L$, the semigroup $\{P_t : t \geq 0\}$ (including their $L^2$-domains) and thus also the Markov process $\{X_t : t \geq 0\}$ are constructed. The assumptions we will make here are those of a so-called *Full Markov Triple* $(E, \mu, \Gamma)$ in the sense of [BGL14, Part I, Chapter 3]. Before introducing this setting rigorously, let us give an informal description. Random variables are viewed as elements of an algebra $\mathcal{A}$ of functions $F\colon E \to \mathbb{R}$, where $(E, \mathcal{F}, \mu)$ is some probability space. On this algebra $\mathcal{A}$, the generator $L$ and the bilinear and symmetric carré du champ operator $\Gamma$ are defined and related via the identity

$$\Gamma(F, G) = \frac{1}{2}\left(L(FG) - FLG - GLF\right).$$

They satisfy a diffusion property, which in its simplest form reads

$$L\varphi(F) = \varphi'(F)LF + \varphi''(F)\Gamma(F, F)$$

or, expressed using the carré du champ,

$$\Gamma(\varphi(F), G) = \varphi'(F)\Gamma(F, G).$$

The subset of random variables with finite mean and variance is then $L^2(E, \mu) \subseteq \mathcal{A}$. On this smaller space, $L$ and $\Gamma$ are typically only densely defined on their domains $\mathcal{D}(L)$ and $\mathcal{D}(\mathcal{E}) \times \mathcal{D}(\mathcal{E})$. The symbol $\mathcal{E}$, defined below, stands for a Dirichlet form (the so-called energy functional), which is used to construct the domains. On these domains, an important relation between $L$ and $\Gamma$ holds, namely the integration by parts formula

$$\int_E \Gamma(F, G)\, \mathrm{d}\mu = -\int_E FLG\, \mathrm{d}\mu.$$

We are now going to introduce this setting in a rigorous way, following closely [BGL14, Part I, Chapter 3]. The needed definitions and assumptions are as follows.

(i) $(E, \mathcal{F}, \mu)$ is a probability space and $L^2(E, \mathcal{F}, \mu)$ is separable.

(ii) $\mathcal{A}$ is a vector space of real-valued, measurable functions (random variables) on $(E, \mathcal{F}, \mu)$, stable under products (i.e. $\mathcal{A}$ is an algebra) and under the action of $C^\infty$-functions $\Psi\colon \mathbb{R}^k \to \mathbb{R}$.

(iii) $\mathcal{A}_0 \subseteq \mathcal{A}$ is a subalgebra of $\mathcal{A}$ consisting of bounded functions which are dense in $L^p(E,\mu)$ for every $p \in [1,\infty)$. We assume that $\mathcal{A}_0$ is also stable under the action of smooth functions $\Psi$ as above and also that $\mathcal{A}_0$ is an ideal in $\mathcal{A}$ (if $F \in \mathcal{A}_0$ and $G \in \mathcal{A}$, then $FG \in \mathcal{A}_0$)

(iv) The carré du champ operator $\Gamma\colon \mathcal{A}_0 \times \mathcal{A}_0 \to \mathcal{A}_0$ is a bilinear symmetric map such that $\Gamma(F,F) \geq 0$ for all $F \in \mathcal{A}_0$. For every $F \in \mathcal{A}_0$ there exists a finite constant $c_F$ such that for every $G \in \mathcal{A}_0$

$$\left| \int_E \Gamma(F,G) \,\mathrm{d}\mu \right| \leq c_F \left\| G \right\|_2 ,$$

where $\left\| G \right\|_2^2 = \int_E G^2 \,\mathrm{d}\mu$. The Dirichlet form $\mathcal{E}$ is defined on $\mathcal{A}_0 \times \mathcal{A}_0$ by

$$\mathcal{E}(F,G) = \int_E \Gamma(F,G) \,\mathrm{d}\mu.$$

(v) The domain $\mathcal{D}(\mathcal{E}) \subseteq L^2(E,\mu)$ is obtained by completing $\mathcal{A}_0$ with respect to the norm $\left\| F \right\|_{\mathcal{E}} = (\left\| F \right\|_2 + \mathcal{E}(f,f))^{1/2}$. The Dirichlet form $\mathcal{E}$ and the carré du champ operator $\Gamma$ are extended to $\mathcal{D}(\mathcal{E}) \times \mathcal{D}(\mathcal{E})$ by continuity and polarization. We thus have that $\Gamma\colon \mathcal{D}(\mathcal{E}) \times \mathcal{D}(\mathcal{E}) \to L^1(E,\mu)$.

(vi) $L$ is a linear operator, defined on $\mathcal{A}_0$ via the *integration by parts formula*

$$\int_E FLG \,\mathrm{d}\mu = -\int_E \Gamma(F,G) \,\mathrm{d}\mu \tag{4}$$

for all $F, G \in \mathcal{A}_0$. We assume that $L(\mathcal{A}_0) \subseteq \mathcal{A}_0$.

(vii) The domain $\mathcal{D}(L) \subseteq \mathcal{D}(\mathcal{E})$ consists of all $F \in \mathcal{D}(\mathcal{E})$ such that

$$\left| \mathcal{E}(F,G) \right| \leq c_F \left\| G \right\|_2$$

for all $G \in \mathcal{D}(\mathcal{E})$, where $c_F$ is a finite constant. The operator $L$ is extended from $\mathcal{A}_0$ to $\mathcal{D}(L)$ by the integration by parts formula (4). On $\mathcal{D}(L)$, $L$ is by construction self-adjoint (as $\Gamma$ is symmetric). We assume that $L1 = 0$ and that $L$ is ergodic: $LF = 0$ implies that $F$ is constant for all $F \in \mathcal{D}(L)$.

(viii) The operator $L\colon \mathcal{A} \to \mathcal{A}$ is an extension of $L\colon \mathcal{A}_0 \to \mathcal{A}_0$. On $\mathcal{A} \times \mathcal{A}$, the carré du champ $\Gamma$ is defined by

$$\Gamma(F,G) = \frac{1}{2} \left( L(FG) - FLG - GLF \right). \tag{5}$$

(ix) For all $F \in \mathcal{A}$, we assume $\Gamma(F,F) \geq 0$ with equality if, and only if, $F$ is constant. By the integration by parts formula (4), this implies in particular that $-L$ is a positive symmetric operator, and therefore the spectrum of $L$ is contained in $(-\infty, 0]$, with 0 always being an eigenvalue given that $L1 = 0$.

(x) The *diffusion property* holds. For all $C^\infty$-functions $\Psi\colon \mathbb{R}^p \to \mathbb{R}$ and $F_1, \dots, F_p, G \in \mathcal{A}$ one has

$$\Gamma\left( \Psi\left( F_1, \dots, F_p \right), G \right) = \sum_{j=1}^{p} \partial_j \Psi(F_1, \dots, F_p) \, \Gamma(F_j, G) \tag{6}$$

and

$$L\Psi(F_1,\ldots,F_p)=\sum_{i=1}^{p}\partial_i\Psi(F_1,\ldots,F_p)LF_i+\sum_{i,j=1}^{p}\partial_{ij}\Psi(F_1,\ldots,F_p)\,\Gamma(F_i,F_j). \tag{7}$$

(xi) The integration by parts formula (4) continues to hold if $F\in\mathcal{A}$ and $G\in\mathcal{A}_0$ (or vice versa).

Of course, one can also introduce a symmetric Markov semigroup and associated Markov process with infinitesimal generator $L$ defined on its domain $\mathcal{D}(L)$ but as we will not make direct use of both of these objects in this paper, we again refer to [BGL14, Part I, Chapter 3] for details.

To summarize, we have an algebra $\mathcal{A}$ of random variables on some probability space $(E,\mathcal{F},\mu)$ on which the carré du champ operator $\Gamma$ and the generator $L$ act. The measure $\mu$ is called the invariant measure of $L$. Note that there is no integrability assumption on the elements of $\mathcal{A}$. The $L^2(E,\mu)$-domains of $L$ and $\Gamma$ are denoted by $\mathcal{D}(L)$ and $\mathcal{D}(\mathcal{E})\times\mathcal{D}(\mathcal{E})$, respectively, and both, $\mathcal{D}(L)$ and $\mathcal{D}(\mathcal{E})$ are dense in $L^2(E,\mu)$. By construction, one has $\mathcal{A}_0\subseteq\mathcal{D}(L)\subseteq\mathcal{D}(\mathcal{E})\subseteq L^2(E,\mu)\subseteq\mathcal{A}$.

A model example of the setting described above is the Markov triple $(\mathbb{R}^d,\gamma_d,\Gamma)$, where $\gamma_d$ is the $d$-dimensional Gaussian measure and $\Gamma=\langle\nabla f,\nabla f\rangle_{\mathbb{R}^d}$ the carré du champ of the $d$-dimensional Ornstein-Uhlenbeck generator $L$ given by $Lf=x\cdot\nabla f+\Delta f$. A suitable algebra $\mathcal{A}$ is given by polynomials in $d$ variables. In infinite dimension, one obtains the infinite-dimensional Ornstein-Uhlenbeck generator on Wiener space with Wiener measure as invariant distribution. In this case, we have that $L=-\delta D$, where $\delta$ is the Malliavin divergence operator (also called Skorohod integral) and $D$ the Malliavin derivation operator with carré du champ operator $\Gamma$ given by $\Gamma(F,G)=\langle DF,DG\rangle_{\mathfrak{H}}$, where $\mathfrak{H}$ denotes the underlying Hilbert space. For further details on this example, see [BH91, Nua06], or [NP12].

The Ornstein-Uhlenbeck generator is a particular example of Pearson generators which will be discussed in detail in Section 4. General references with more examples fitting our framework are [BÉ85, Bak14, BGL14, FOT11].

In what follows, we will also make use of the pseudo-inverse $L^{-1}$ of $L$, satisfying for any $F\in\mathcal{D}(L)$,

$$LL^{-1}F=L^{-1}LF=F-\pi_0(F), \tag{8}$$

where $\pi_0(F)=\int_E F\,\mathrm{d}\mu$ denotes the orthogonal projection of $F$ onto $\ker(L)$ (recall that the kernel of $L$ by assumption only consists of constants). For completeness, we recall how this pseudo-inverse is constructed. By self-adjointness of $L$, considered as an operator on $\mathcal{D}(L)$, we have that $\mathcal{D}(L)=\ker(L)\oplus(\operatorname{ran}(L)\cap\mathcal{D}(L))$. Therefore, we can define $L^{-1}$ on $\operatorname{ran}(L)\cap\mathcal{D}(L)$ (as $L$ is injective there) and then extend it to $\mathcal{D}(L)$ by setting $L^{-1}F=0$ if $F\in\ker(L)$.

We end this subsection by a useful lemma which combines the integration by parts formula (4) and the diffusion property (6).

**Lemma 2.1.** *In the setting introduced above, let $F\in\mathcal{D}(\mathcal{E})$, $G\in\mathcal{D}(L)$ such that $\int_E G\,\mathrm{d}\mu=0$ and $\varphi\colon\mathbb{R}\to\mathbb{R}$ be a differentiable function such that $\varphi(F)\in\mathcal{D}(\mathcal{E})$. Furthermore, assume that $\varphi(F)G\in L^1(E,\mu)$. Then, one has*

$$\int_E\varphi(F)G\,\mathrm{d}\mu=\int_E\varphi'(F)\Gamma(F,-L^{-1}G)\,\mathrm{d}\mu.$$

*Proof.* By (8) and the assumption that $G$ is centered, we have that $G = LL^{-1}G$. Therefore, by the integration by parts formula (4) and the diffusion property (6), one has

$$\begin{aligned}\int_E \varphi(F)G \,\mathrm{d}\mu &= \int_E \varphi(F)LL^{-1}G \,\mathrm{d}\mu \\ &= \int_E \Gamma\left(\varphi(F), -L^{-1}G\right) \mathrm{d}\mu \\ &= \int_E \varphi'(F)\Gamma(F, -L^{-1}G) \,\mathrm{d}\mu.\end{aligned}$$

□

2.2. **Stein's method for invariant measures of diffusions.** In this section, we present Stein's method for invariant measures of diffusions. Note that if $\mu$ is a measure which is absolutely continuous with respect to the Lebesgue measure and admits a density $p$ as well as a second moment, then under very minimal assumptions there exists a Markov diffusion generator $L$ having $\mu$ as its invariant measure.

To be more precise, let $\mu$ be a probability measure admitting a density $p$ with support $(l, u) \subseteq \mathbb{R}$, $-\infty \le l < u \le +\infty$. Furthermore, let $\theta > 0$, denote $m = \int_{\mathbb{R}} xp(x)dx$ and

$$\sigma^2(x) = \frac{-2\theta \int_{-\infty}^{x} (y - m)p(y) \,\mathrm{d}y}{p(x)}, \; x \in (l, u), \tag{9}$$

which is a non-negative quantity. Then, the stochastic differential equation

$$\mathrm{d}X_t = -\theta(X_t - m) \,\mathrm{d}t + \sigma(X_t) \,\mathrm{d}B_t, \; X_t \in (l, u), \tag{10}$$

where $\{B_t : t \ge 0\}$ is a Brownian motion, has a unique weak Markovian solution with invariant measure $\mu$ (see [BSSr05], Theorem 2.3). The support of the density $p$ could very well be taken to be a union of open intervals, but we treat here the case of one open interval in order not to make the notation heavier than it needs to be.

Stein's density approach (see [Ste86] for a detailed treatment) allows us to characterize the invariant measure $\mu$ of the diffusion (10) through the following theorem, called Stein's lemma for invariant measures of diffusions (see [NP09b, Proposition 6.4] or [EVq15, Lemma 6]).

**Theorem 2.2.** *Let $\mu$ be a probability measure admitting a density $p$ with support $(l, u) \subseteq \mathbb{R}$, $-\infty \le l < u \le +\infty$, such that $\int_{\mathbb{R}} |x|\,p(x)dx < \infty$ and $\int_{\mathbb{R}} xp(x)dx = m$. Define the function*

$$\tau(x) = \frac{1}{2}\sigma^2(x)\mathbb{1}_{(l,u)}(x), \quad x \in \mathbb{R},$$

*where $\sigma^2$ is defined in terms of $p$ by (9) and let $Z$ be a random variable having distribution $\mu$. Suppose*

*(i) For every differentiable $\varphi$ such that $\tau(Z)\varphi'(Z) \in L^1(\Omega)$, one has that $(Z - m)\varphi(Z) \in L^1(\Omega)$ and*

$$\mathrm{E}\left(\tau(Z)\varphi'(Z) - \theta(Z - m)\varphi(Z)\right) = 0.$$

*(ii) Let $X$ be a real-valued random variable with an absolutely continuous distribution. If for every differentiable $\varphi$ such that $\tau(X)\varphi'(X) \in L^1(\Omega)$ and $(X - m)\varphi(X) \in L^1(\Omega)$, one has that*

$$\mathrm{E}\left(\tau(X)\varphi'(X) - \theta(X-m)\varphi(X)\right) = 0, \tag{11}$$

*then $X$ has distribution $\mu$.*

Based on the above Stein lemma, one can use the now well established Stein methodology to quantitatively measure the distance between the law of a random variable $X$ and the law of a random variable $Z$ corresponding to an invariant measure of a diffusion. The generalization of the original Stein method to invariant measures of diffusions has been recently studied in [KT12] and further developed in [EVq15]. In order to present this method, we need to introduce separating classes of functions and probabilistic distances.

**Definition 2.3.** Let $\mathscr{H}$ be a collection of Borel-measurable functions on $\mathbb{R}$. We say that the class $\mathscr{H}$ is *separating* if the following property holds: any two real-valued random variables $X, Y$ verifying $h(X), h(Y) \in L^1(\Omega)$ and $\mathrm{E}(h(X)) = \mathrm{E}(h(Y))$ for every $h \in \mathscr{H}$, are necessarily such that $X$ and $Y$ have the same distribution.

Separating classes of functions can be used to introduce distances between probability measures in the following way.

**Definition 2.4.** Let $\mathscr{H}$ be a separating class in the sense of Definition 2.3 and let $X, Y$ be real-valued random variables such that $h(X), h(Y) \in L^1(\Omega)$ for every $h \in \mathscr{H}$. Then the *distance* $d_{\mathscr{H}}(X, Y)$ between the distributions $X$ and $Y$ is given by

$$d_{\mathscr{H}}(X,Y) = \sup_{h\in\mathscr{H}} \left|\mathrm{E}(h(X)) - \mathrm{E}(h(Y))\right|. \tag{12}$$

One can show that $d_{\mathscr{H}}$ is a metric on some subset of the class of all probability measures on $\mathbb{R}$ (see [Dud02, Chapter 11]). With some abuse of language, one often speaks of the "distance between random variables" when really the distance between the laws of these random variables is meant. We will call a given distance $d_{\mathscr{H}}$ *admissible* for a set $\mathcal{M}$ of random variables if $d_{\mathscr{H}}(X, Y)$ is well defined for all $X, Y \in \mathcal{M}$, i.e. if it holds that $\mathrm{E}[h(X)] < \infty$ for all $X \in \mathcal{M}$ and $h \in \mathscr{H}$. As an example of a distance as introduced above, one can take $\mathscr{H}$ to be the class of Lipschitz continuous and bounded functions. This yields the well-known Fortet-Mourier (or bounded Wasserstein) distance denoted by $d_{FM}$, which metrizes convergence in distribution and is defined for *all* real-valued random variables. It is therefore admissible for any set $\mathcal{M}$ of random variables. Other distances (with smaller domains) are the total variation, Kolmogorov or Wasserstein distance (see [NP12, Appendix C]). A Stein equation is an ordinary differential equation linking the notion of distance (through the left-hand side of (12)) to the characterizing expression of a distribution appearing in Stein's lemma (the right-hand side of (11) for instance). More precisely, a Stein equation associated to the Stein characterization (11) is given by

$$\tau(x)f'(x) - \theta(x-m)f(x) = h(x) - \mathrm{E}(h(Z)), \tag{13}$$

where $Z$ is a random variable with distribution $\mu$ given by the invariant measure of the diffusion in (10). It is straightforward to check that this equation has a continuous solution on $\mathbb{R}$ for each $h \in \mathscr{H}$, denoted by $f_h$, and given by

$$\begin{aligned} f_h(x) &= \frac{1}{\tau(x)p(x)} \int_l^x (h(y) - \mathrm{E}(h(Z)))p(y)dy \\ &= -\frac{1}{\tau(x)p(x)} \int_x^u (h(y) - \mathrm{E}(h(Z)))p(y)dy \end{aligned}$$

for $x \in (l, u)$, and by

$$f_h(x) = -\frac{h(x) - \mathrm{E}(h(Z))}{\theta(x-m)}$$

when $x \notin (l, u)$ (as in that case, $\tau(x) = 0$). Now, let $X$ be a real-valued random variable with an absolutely continuous distribution. By letting $x = X$ in (13), taking expectations and the supremum over the separating class of test functions $\mathscr{H}$ on both sides, we can express the distance $d_{\mathscr{H}}$ in (12) as

$$d_{\mathscr{H}}(X, Z) = \sup_{h \in \mathscr{H}} \left| \mathrm{E}\left(\tau(X) f_h'(X)\right) - \mathrm{E}\left(\theta(X-m) f_h(X)\right) \right|. \tag{14}$$

The following result, a proof of which can be found in [EVq15, Lemma 7], combines results from [KT12] and [EVq15] and provides sufficient conditions under which useful estimates for $f_h'$ can be obtained.

**Lemma 2.5.** *Let the function $\sigma^2$, associated to a density $p$ with support $(l, u) \subseteq \mathbb{R}$, $-\infty \le l < u \le +\infty$, be given by* (9). *If $u = \infty$, then assume that $\underline{\lim}_{x\to u} \sigma^2(x) > 0$, and if $l = -\infty$, assume that $\underline{\lim}_{x\to l} \sigma^2(x) > 0$. Furthermore, suppose that there exists a positive function $g \in C^1((l, u), \mathbb{R}_+)$ such that*

*(i)* $0 < \underline{\lim}_{x\to u} \sigma^2(x)/g(x) \le \overline{\lim}_{x\to u} \sigma^2(x)/g(x) < \infty$;
*(ii)* $\lim_{x\to u} g'(x) \in [-\infty, +\infty]$;
*(iii)* $0 < \underline{\lim}_{x\to l} \sigma^2(x)/g(x) \le \overline{\lim}_{x\to l} \sigma^2(x)/g(x) < \infty$;
*(iv)* $\lim_{x\to l} g'(x) \in [-\infty, +\infty]$.

*Then the solution $f_h$ to the Stein equation* (13)*, for a given test function $h \in \mathscr{H}$ such that $\|h'\|_\infty < \infty$, satisfies*

$$\left\| f_h' \right\|_\infty \le k \left\| h' \right\|_\infty,$$

*where the constant $k$ does not depend on $h$.*

2.3. **Pearson diffusions.** Pearson distributions were first classified in [Pea95] by Pearson who noticed that some of the most important distributions in statistics, namely the Gaussian, exponential, gamma, uniform, beta, Student-t, F, and inverse gamma distributions, share the common feature that their logarithmic derivative can be represented as the ratio of a linear and a quadratic polynomial (see (16)). The corresponding class of diffusions having these distributions as invariant measures play an equally central role and include, for example, the Ornstein-Uhlenbeck and Cox-Ingersoll-Ross processes, which are ubiquitous in physics and finance.

Mathematically, Pearson diffusions are Itō diffusions with mean reverting linear drift whose squared diffusion coefficient is a quadratic polynomial, i.e., a stationary solution of the stochastic differential equation

$$\mathrm{d}X_t = a(X_t)\,\mathrm{d}t + \sqrt{2\theta b(X_t)}\,\mathrm{d}B_t, \tag{15}$$

where $a(x) = -\theta(x-m)$ and

$$b(x) = b_2 x^2 + b_1 x + b_0.$$

Here, $m, b_2, b_1, b_0$ are real constants, $\theta > 0$ determines the speed of mean reversion and $m$ is the stationary mean. Recall that the scale and speed densities $s$ and $p$, respectively, are defined as

$$s(x) = \exp\left(-2\int_{x_0}^{x} \frac{a(u)}{\sigma^2(u)}\, \mathrm{d}u\right) \qquad \text{and} \qquad p(x) = \frac{1}{s(x)\sigma^2(x)}.$$

In our case, we have the above-mentioned relation

$$p'(x) = -\frac{(2b_2+1)x - m + b_1}{b_2x^2 + b_1x + b_0}\, p(x), \tag{16}$$

which was originally used by Pearson (see [Pea95, page 360]) to introduce these distributions. From (16) one also sees that the class of Pearson diffusions is closed under linear transformations. Explicitly, if $X_t$ satisfies the stochastic differential equation (15), then $\widetilde{X}_t = \gamma X_t + \delta$ satisfies

$$\mathrm{d}\widetilde{X}_t = \widetilde{a}(\widetilde{X}_t)\,\mathrm{d}t + \widetilde{\sigma}(\widetilde{X}_t)\,\mathrm{d}B_t,$$

where $\widetilde{a}(x) = -\theta\,(x - \gamma m - \delta)$ and

$$\widetilde{\sigma}^2(x) = 2\theta\left(b_2x^2 + (b_1\gamma - 2b_2\delta)\,x + b_0\gamma^2 - b_1\gamma\delta + b_2\delta^2\right).$$

Up to such linear transformations, Pearson diffusions can be categorized into the six classes listed below together with their invariant distributions, densities, means and diffusion coefficients. A detailed analysis and classification of Pearson diffusions can for example be found in [JKB94, JKB95, FS08].

1. Gaussian distribution with parameters $m \in \mathbb{R}$ and $\sigma > 0$. It has state space $\mathbb{R}$, mean $m$, as well as density function and diffusion coefficients given by

$$p(x) \propto e^{-\frac{(x-m)^2}{2\sigma^2}}, \qquad b(x) = \sigma^2.$$

   The Gaussian distribution has moments of all orders.
2. Gamma distribution with parameters $\alpha, \beta > 0$. It has state space $(0, \infty)$, mean $\frac{\alpha}{\beta}$, as well as density function and diffusion coefficients given by

$$p(x) \propto x^{\alpha-1}e^{-\beta x}, \qquad b(x) = \frac{x}{\beta}.$$

   The Gamma distribution has moments of all orders.
3. Beta distribution with parameters $\alpha, \beta > 0$. It has state space $(0, 1)$, mean $\frac{\alpha}{\alpha+\beta}$, as well as density function and diffusion coefficients given by

$$p(x) \propto x^{\alpha-1}(1-x)^{\beta-1}, \qquad b(x) = -\frac{x^2}{\alpha+\beta} + \frac{x}{\alpha+\beta}.$$

   The Beta distribution has moments of all orders.
4. Skew $t$-distribution with parameters $m, \nu, \lambda \in \mathbb{R}$, $\alpha > 0$. It has state space $\mathbb{R}$, mean $\frac{(2m-1)\lambda + \alpha\nu}{2(m-1)}$, as well as density function and diffusion coefficients given by

$$p(x) \propto \left(1 + \left(\frac{x-\lambda}{\alpha}\right)^2\right)^{-m} e^{-\nu \arctan\left(\frac{x-\lambda}{\alpha}\right)},$$

$$b(x) = \frac{x^2}{2(m-1)} - \frac{\lambda x}{2(m-1)} + \frac{\lambda^2+\alpha^2}{2(m-1)}.$$

   The skew $t$-distribution has moments of order $p$ for $p < 2m-1$.

5. Inverse gamma distribution with parameters $\alpha, \beta > 0$. It has state space $(0, \infty)$, mean $\frac{\beta}{\alpha-1}$, as well as density function and diffusion coefficients given by
$$p(x) \propto x^{-(\alpha-1)} e^{-\frac{\beta}{x}}, \qquad b(x) = \frac{x^2}{\alpha - 1}.$$
The inverse gamma distribution has moments of order $p$ for $p < \alpha$.
6. $F$-distribution with parameters $d_1, d_2 > 0$. It has state space $(0, \infty)$, mean $\frac{d_2}{d_2-2}$, as well as density function and diffusion coefficients given by
$$p(x) \propto x^{\frac{d_1}{2}-1} \left(1 + \frac{d_1}{d_2} x\right)^{-\frac{d_1+d_2}{2}}, \qquad b(x) = \frac{2x^2}{d_2 - 2} + \frac{2d_2 x}{d_1(d_2 - 2)}.$$
The $F$-distribution has moments of order $p$ for $p < \frac{d_2}{2}$.

Pearson diffusions are particular (one-dimensional) examples fitting the Markov triple structure introduced in Subsection 2.1. The generator $L$ acts on $L^2(E, \mu)$ via
$$Lf(x) = -(x - m)f'(x) + b(x)f''(x), \tag{17}$$
where $b$ is the quadratic polynomial appearing in (15). Its invariant measure $\mu$ is a Pearson distribution and it is furthermore symmetric, ergodic and diffusive (in the sense of (7)). The set $\Lambda$ of eigenvalues of $L$ is given by (see for example [FS08])
$$\Lambda = \left\{ -n\,(1 - (n-1)b_2)\,\theta :\ n \in \mathbb{N}_0,\ b_2 < \frac{1}{2n-1} \right\} \tag{18}$$
and the corresponding eigenfunctions are the well-known orthogonal polynomials associated with the respective laws (Hermite, Laguerre and Jacobi polynomials for the Gaussian, Gamma and Beta distributions, respectively, and Romanovski-Routh, Romanovski-Bessel and Romanovski-Jacobi polynomials for the skew $t$-, inverse gamma and $F$-distributions. From formula (18), we see that polynomials up to degree $n$, where $n$ is the largest integer strictly less than $\frac{1+b_2}{2b_2}$, are (square integrable) eigenfunctions, so that $\mu$ has moments up to order $2n$. Note that the cardinality of $\Lambda$ is infinite if $b_2 \leq 0$ and finite if $b_2 > 0$. Consistent with the general theory of Markov generators presented in Subsection 2.1 (see (ix)), zero is always contained in $\Lambda$ and all other eigenvalues are negative.

The structure of the spectrum $S$ of such a Pearson generator can thus be described as follows.

(i) If $\mu$ is a Gaussian, Gamma or Beta distribution, then $S$ is purely discrete and consists of infinitely many eigenvalues, each of multiplicity one. In the Gaussian and Gamma case, where $b_2 = 0$, these eigenvalues are the negative integers (up to the common scaling factor $\theta$) including zero. Eigenfunctions are the associated orthogonal polynomials (Hermite, Laguerre or Jacobi).

(ii) If $\mu$ is a skew $t$-, inverse Gamma or scaled $F$-distribution, then $S$ contains a discrete and a continuous part. The discrete part consists of only finitely many eigenvalues.

For later reference, we note that for a Pearson distribution $\mu$, the Stein characterization (11) in Theorem 2.2 becomes
$$\mathrm{E}\Big[b(X)\mathbb{1}_{(l,u)}(X)\varphi'(X) - (X - m)\varphi(X)\Big] = 0, \tag{19}$$
where again $b(x) = b_2x^2 + b_1x + b_0$ is the associated quadratic polynomial.

Identity (19) gives a recursion formula for computing the moments of a given Pearson distribution. Indeed, if the law of $X$ is a Pearson distribution with moments up to order $p+2$, then (19) with $\varphi(x) = x^{p+1}$ reads

$$(p+1)\,\mathrm{E}\big[b(X)X^p\big] - \mathrm{E}\Big[(X-m)X^{p+1}\Big] = 0.$$

This yields

$$(b_2(p+1)-1)\,\mathrm{E}\Big[X^{p+2}\Big] + (b_1(p+1)+m)\,\mathrm{E}\Big[X^{p+1}\Big] + (p+1)b_0\,\mathrm{E}\big[X^p\big] = 0$$

with $\mathrm{E}[X] = m$. Recall from the previous discussion that the condition for the existence of moments of order $p$ is $p < 1 + b_2^{-1}$, so that four moments exist if and only if $b_2 < \frac{1}{3}$. In this case, we start with with $\mathrm{E}[X] = m$ and get

$$\begin{aligned}
\mathrm{E}\Big[X^2\Big] &= \frac{(b_1+m)m+b_0}{1-b_2},\\
\mathrm{E}\Big[X^3\Big] &= \frac{(2b_1+m)\left((b_1+m)\,m+b_0\right)}{(1-b_2)(1-2b_2)} + \frac{2b_0m}{1-2b_2},\\
\mathrm{E}\Big[X^4\Big] &= \frac{(3b_1+m)(2b_1+m)\left((b_1+m)\,m+b_0\right)}{(1-b_2)(1-2b_2)(1-3b_2)} + \frac{(3b_1+m)2b_0m}{(1-2b_2)(1-3b_2)}\\
&\quad + \frac{3b_0\left((b_1+m)\,m+b_0\right)}{1-3b_2}.
\end{aligned}$$

For further analysis of the spectrum of such Pearson generators and general motivation on studying Pearson diffusions, see [ALŠ13].

## 3. Main results

Throughout this section, we always work in the Markov setting introduced in Subsection 2.1. We thus have a probability space $(E, \mathcal{F}, \mu)$ and the two operators $L$ and $\Gamma$ with their $L^2$-domains $\mathcal{D}(L)$ and $\mathcal{D}(\mathcal{E}) \times \mathcal{D}(\mathcal{E})$ respectively, where $\mathcal{D}(L) \subseteq \mathcal{D}(\mathcal{E}) \subseteq L^2(E,\mu)$. As is customary in this context, we continue to use the integral notation for mathematical expectation, so that for example the expectation of a random variable $G \in L^1(E,\mu)$ is denoted by $\int_E G\,\mathrm{d}\mu$.

### 3.1. Carré du champ characterization.

As a first result, we show how the Stein characterization (11) can be used in order to naturally characterize, in terms of the carré du champ operator $\Gamma$, when a random variable $G$ has a given probability distribution $\nu$.

**Theorem 3.1.** *Let $\nu$ be a probability measure admitting a density $p$ with support $(l,u) \subseteq \mathbb{R}$, $-\infty \leq l < u \leq +\infty$, such that $\int_{\mathbb{R}} |x|\,p(x)dx < \infty$ and $\int_{\mathbb{R}} xp(x)dx = m$. Define the function*

$$\tau(x) = \frac{1}{2}\sigma^2(x)\mathbb{1}_{(l,u)}(x), \quad x \in \mathbb{R},$$

*where $\sigma^2$ is defined in terms of $p$ by (9). Let $G \in \mathcal{D}(L)$ with an absolutely continuous distribution and mean $m$. Then $G$ has distribution $\nu$ if, and only if,*

$$\Gamma(G, -L^{-1}G) = \theta^{-1}\tau(G)$$

*almost surely.*

*Proof.* Let $\varphi \in C^\infty(\mathbb{R}, \mathbb{R})$ be such that $\tau(G)\varphi'(G) \in L^1(E)$ and $(G-m)\varphi(G) \in L^1(E)$. By Lemma 2.1, one has

$$\int_E (G-m)\varphi(G)\,\mathrm{d}\mu = \int_E \varphi'(G)\Gamma(G, -L^{-1}G)\,\mathrm{d}\mu. \tag{20}$$

This implies

$$\int_E \tau(G)\varphi'(G) - \theta(G-m)\varphi(G)\,\mathrm{d}\mu = \theta \int_E \varphi'(G)\left(\theta^{-1}\tau(G) - \Gamma(G, -L^{-1}G)\right)\mathrm{d}\mu, \tag{21}$$

so that the assertion follows from Theorem 2.2. □

Using Stein's method, we obtain the following quantitative version of Theorem 3.1.

**Theorem 3.2.** *Let $\nu$ be a measure with density $p$ and let $\sigma^2$ be given by* (9)*. Assume that $\sigma^2$ satisfies the assumptions of Lemma 2.5, and let $\tau(x) = \frac{1}{2}\sigma^2(x)\mathbb{1}_{(l,u)}(x)$, $x \in \mathbb{R}$. Furthermore, let $G \in \mathcal{D}(L)$ such that $\int_E \tau(G)\,\mathrm{d}\mu < \infty$ and $\int_E G\,\mathrm{d}\mu = m$. Finally, let $Z$ be a random variable with distribution $\nu$. Then one has*

$$d_{\mathscr{H}}(G, Z) \leq c_{\mathscr{H}} \int_E \left|\Gamma(G, -L^{-1}G) - \theta^{-1}\tau(G)\right| d\mu, \tag{22}$$

*where $d_{\mathscr{H}}$ is an admissible distance for $G$ and $Z$, defined via* (12) *using a separating class $\mathscr{H}$ of absolutely continuous test functions such that $\sup_{h\in\mathscr{H}} \|h'\|_\infty < \infty$ and $c_{\mathscr{H}}$ is a positive constant depending solely on the class $\mathscr{H}$.*

*Remark* 3.3. Note that the Fortet-Mourier metric always satisfies the assumptions of Theorem 3.2, as, by definition, $\|h'\|_\infty \leq 1$ for all $h$ in the Fortet-Mourier class of test functions (see for example [NP12, Appendix C]). In concrete situations, when both the law $\nu$ and the generator $L$ are explicit, one can often take stronger distances such as Kolmogorov or total variation.

*Proof of Theorem 3.2.* On the one hand, by using Stein's method for invariant measures of diffusions (see Subsection 2.2), we can write, using (14),

$$d_{\mathscr{H}}(G, Z) = \sup_{h\in\mathscr{H}} \left| \int_E \tau(G) f_h'(G) d\mu - \int_E \theta(G-m) f_h(G) d\mu \right|, \tag{23}$$

where $f_h$ denotes the solution to the Stein equation (13). On the other hand, by Lemma 2.1, one has

$$\int_E (G-m) f_h(G)\,\mathrm{d}\mu = \int_E f_h'(G)\Gamma(G, -L^{-1}G)\,\mathrm{d}\mu.$$

Plugged into (23) and applying the Hölder inequality, we obtain

$$\begin{aligned} d_{\mathscr{H}}(G, Z) &= \sup_{h\in\mathscr{H}} \left| \int_E \left(\tau(G) f_h'(G) - \theta f_h'(G)\Gamma(G, -L^{-1}G)\right)\mathrm{d}\mu \right| \\ &\leq \sup_{h\in\mathscr{H}} \left\|f_h'\right\|_\infty \theta \int_E \left|\Gamma(G, -L^{-1}G) - \theta^{-1}\tau(G)\right| \mathrm{d}\mu, \end{aligned}$$

so that the assertion follows by Lemma 2.5 with $c_{\mathscr{H}} = k\theta \sup_{h\in\mathscr{H}} \|h'\|_\infty < \infty$. □

*Remark* 3.4. Let us point out some key features of the results of this subsection.

Firstly, to avoid any confusion, note that the target measure $\nu$ appearing in Theorems 3.1 and 3.2 is not related to the invariant measure $\mu$ of the generator

$L$, in the domain of which $G$ lies. As pointed out in Subsection 2.2, almost any distribution admitting a density can be regarded as an invariant measure of a diffusion, and is therefore admissible as a target distribution $\nu$. Secondly, observe that our assumptions on the random variable $G$ are quite mild, hence providing a lot of flexibility for applications.

In the specific case where the underlying generator is the infinite-dimensional Ornstein-Uhlenbeck generator (Wiener space) and the target law $\nu$ is Gaussian (constant diffusion coefficient), a bound of this type had been obtained in [NP09b], which has been applied in several contexts, for example to obtain Berry-Esséen theorems for parameter estimation of stochastic partial differential equations (see [KP17]) or in the context of fractional Ornstein-Uhlenbeck processes (see [HNZ17]).

Theorems 3.1 and 3.2 in this section extend the class of possible target distributions from Gaussian to general invariant measures of diffusions, and also allows to consider functionals of non-Gaussian random fields. In particular, both the target law and the underlying random field can have heavy tails.

3.2. **Markov chaos and Four Moments Theorems.** This subsection introduces the concept of chaotic eigenfunctions, for which the general bound obtained in Theorem 3.2 can further be bounded by a finite linear combination of moments. Chaotic eigenfunctions have first been introduced in [Led12] and a more general definition has been given in [ACP14]. In order to also be able to deal with heavy-tailed invariant measures, we have to extend this definition once again by introducing the new notion of chaos grade.

We continue to assume as given a Markov structure as introduced in Subsection 2.1 and denote the spectrum of the generator $L$ (defined on $\mathcal{D}(L)$) by $S$. As $-L$ is non-negative and symmetric, one has $S \subseteq (-\infty, 0]$. Let $\Lambda \subseteq S$ denote the set of eigenvalues of $L$. We always have that $0 \in \Lambda$ as by assumption $L1 = 0$. Chaotic random variables are then defined as follows.

**Definition 3.5.** An eigenfunction $F$ with respect to an eigenvalue $-\lambda$ of $L$ is called *chaotic*, if there exists $\eta > 1$ such that $-\eta\lambda$ is an eigenvalue of $L$ and

$$F^2 \in \bigoplus_{\substack{-\kappa \in \Lambda \\ \kappa \leq \eta\lambda}} \ker\left(L + \kappa \operatorname{Id}\right). \tag{24}$$

In this case, the smallest $\eta$ satisfying (18) is called the *chaos grade* of $F$.

In other words, an eigenfunction is called chaotic if its square can be expressed as a sum of eigenfunctions.

*Remark* 3.6. (i) As we assume that $L^2(E, \mathcal{F}, \mu)$ is separable, the set $\Lambda$ and therefore the direct orthogonal sum (24) of eigenspaces is at most countable.
(ii) The chaos grade is invariant under scaling of the generator, in the sense that if $F$ is a chaotic random variable of $L$ with chaos grade $\eta$, then the chaos grade of $F$ remains unchanged when viewed as a chaotic random variable of $\alpha L$ for any $\alpha \in \mathbb{R}$.

Let us give some examples to illustrate the concept.

*Example* 3.7. 1. An example is the generator $L$ of a Pearson distribution and we will study this example in detail in Section 4. At this point, let us briefly illustrate the chaos grade concept by treating the concrete case of the Gaussian distribution $\mu$. Here, the generator is the one-dimensional Ornstein-Uhlenbeck generator, acting on $L^2(\mathbb{R}, \mu)$. As is well known, the spectrum of $L$ consists of the negative integers and zero, which all are eigenvalues with the respective Hermite polynomials as eigenfunctions (the Hermite polynomial $H_p$ of order $p$ being an eigenfunction with respect to the eigenvalue $-p$). The square of such a Hermite polynomial $H_p$ can of course be expressed as a linear combination of Hermite polynomials up to order $2p$, so that the chaos grade of $H_p$ is $\eta = 2$. This expansion is given explicitly by the well-known product formula

$$H_p^2(x) = \sum_{j=0}^{p} c_{p,j} H_{2(p-j)}(x),$$

where $c_{p,j} = j!\binom{p}{j}^2$.

2. The preceding example can also be looked at in infinite dimensions. Here, the one-dimensional Gaussian distribution is replaced with Wiener measure and $L$ is the infinite dimensional Ornstein-Uhlenbeck generator. The spectrum of $L$ still consists of the negative integers and zero, with the eigenfunctions now being multiple Wiener-Itō integrals of the form $F = I_p(f)$ (so that $LI_p(f) = -pI_p(f)$). The product formula for such integrals says that

$$F^2 = I_p(f)^2 = \sum_{j=0}^{p} c_{p,j} I_{2(p-j)}(f_j),$$

where the constants $c_{p,j}$ are defined as in the previous example and the kernels $f_j$ are given in terms of so-called contractions of the original kernel $f$. This shows that any such multiple Wiener-Itō integral is a chaotic eigenfunction in the sense of Definition 3.5 with chaos grade 2.

3. Another example in dimension one is obtained by taking $L$ to be the Jacobi generator acting on $L^2([0,1], \nu)$, with invariant measure $\nu$ given by $\nu(\mathrm{d}x) = c_{\alpha,\beta} x^{\alpha-1}(1-x)^{\beta-1} 1_{[0,1]}(x)\,\mathrm{d}x$ for some positive parameters $\alpha, \beta$. Then $L$ is such that

$$Lf(x) = x(1-x)f''(x) + (\alpha - (\alpha+\beta)x)f'(x).$$

It is well known that the eigenvalues of $L$ are given by the Jacobi polynomials. The chaos grade of an eigenfunction associated to the eigenvalue $\lambda_n = -n\left(1 + \frac{n-1}{\alpha+\beta}\right)$ is given by $2\left(1 + \frac{n}{n-1+\alpha+\beta}\right)$ (see Section 4 for a full treatment of chaos grade characterizations). Note that the chaos grade in this case is no longer 2 and depends on the eigenvalue the eigenfunction is associated to. As in the Wiener case, a tensorization procedure (see Section 4) allows to generalize this example to higher dimensions.

*Remark* 3.8. For a systematic study of the chaos grades of eigenfunctions of Pearson generators, see Section 4.

We are now ready to prove Four Moments Theorems for Pearson distributions. In all that follows, $F$ will denote an eigenfunction of $L$, which is necessarily centered, and $G = F + m$ a translated version of $F$ which has then expectation $m \in \mathbb{R}$ as in the previous section. Furthermore, as the six classes of Pearson diffusions given by (15) are invariant under linear transformations (see Section 2.3), we assume from here on without loss of generality that $\theta = \frac{1}{2}$.

**Theorem 3.9.** *Let $\nu$ be a Pearson distribution associated to the diffusion given by* (15) *with mean m and diffusion coefficient*

$$\sigma^2(x) = b(x) = b_2x^2 + b_1x + b_0,$$

*where $b_0, b_1, b_2 \in \mathbb{R}$. Let F be a chaotic eigenfunction of L with respect to the eigenvalue $-\lambda$, chaos grade $\eta$ and moments up to order* 4. *Set $G = F + m$. Then, if $\eta \leq 2(1 - b_2)$, one has*

$$\int_E \left(\Gamma(G, -L^{-1}G) - b(G)\right)^2 \mathrm{d}\,\mu \leq 2\left(1 - b_2 - \frac{\eta}{4}\right) \int_E U(G)\,\mathrm{d}\,\mu, \tag{25}$$

*whereas if $\eta > 2(1 - b_2)$, one has*

$$\begin{aligned}\int_E \left(\Gamma(G, -L^{-1}G) - b(G)\right)^2 \mathrm{d}\,\mu \\ \leq 2\left(1 - b_2 - \frac{\eta}{4}\right) \int_E U(G)\,\mathrm{d}\,\mu + \frac{\xi(1 - b_2)}{2} \int_E Q^2(G)\,\mathrm{d}\,\mu,\end{aligned} \tag{26}$$

*where*

$$\xi = \eta - 2(1 - b_2) > 0,$$

*and where the polynomials Q and U are given respectively by*

$$Q(x) = x^2 + \frac{2(b_1 + m)}{2b_2 - 1}x + \frac{1}{b_2 - 1}\left(b_0 + \frac{m(b_1 + m)}{2b_2 - 1}\right),$$

*and*

$$U(x) = (1 - b_2)Q^2(x) - \frac{1}{12}(Q'(x))^3(x - m).$$

*Remark* 3.10.

(i) Observe that both $\int_E U(G)\,\mathrm{d}\,\mu$ and $\int_E Q^2(G)\,\mathrm{d}\,\mu$ are linear combinations of the first four moments of $G$, i.e. there exists coefficients $c_j, d_j$, $j = 0, \ldots, 4$ such that

$$\int_E U(G)\,\mathrm{d}\,\mu = \sum_{j=0}^{4} c_j \int_E G^j d\mu \quad \text{and} \quad \int_E Q^2(G)\,\mathrm{d}\,\mu = \sum_{j=0}^{4} d_j \int_E G^j d\mu.$$

The coefficients $c_j, d_j$ only depend on the coefficients of the polynomial $b$ and the mean $m$ of the target distribution, and hence only on $\nu$. For convenience, they are given in Table 1. We provide some examples of such linear moment combinations below.

(ii) In Theorem 3.13, we will use Theorem 3.9 to obtain moment conditions for the convergence in law of a sequence $\{G_k \colon k \geq 1\}$ to a random variable $Z$ with distribution $\nu$. Consider for example (25). If $\int_E U(G_k)\,\mathrm{d}\,\mu \to 0$ as $k \to \infty$, then the left-hand side of (25) converges to zero, and hence the distribution of $G_k$ converges to the Pearson distribution $\mu$ by Theorem 3.2.

(iii) Note that by the identities (31) and (27) in the forthcoming proof of Theorem 3.9 and the Cauchy-Schwarz inequality,

$$\int_E U(G)\,\mathrm{d}\,\mu \leq \sqrt{\int_E Q^2(G)\,\mathrm{d}\,\mu}\sqrt{\int_E (\Gamma(G, -L^{-1}G) - b(G))^2\,\mathrm{d}\,\mu},$$

TABLE 1. Coefficients in the linear combinations of moments in Remark 3.10.i

| j | $c_j$ |
|---|---|
| 0 | $\dfrac{\left(b_0 + \frac{m(b_1+m)}{2b_2-1}\right)^2}{1-b_2} + \dfrac{2m(b_1+m)^3}{3(2b_2-1)^3}$ |
| 1 | $\dfrac{4b_0(b_1+m)}{1-2b_2} + \dfrac{2(b_1+m)^2(b_1+2m(3b_2-1))}{3(1-2b_2)^3}$ |
| 2 | $-2b_0 - \dfrac{2(b_1+m)^2}{2b_2-1}$ |
| 3 | $-2b_1 - \dfrac{4m}{3}$ |
| 4 | $\dfrac{1}{3} - b_2$ |
| j | $d_j$ |
| 0 | $\dfrac{(b_0(2b_2-1)+m(b_1+m))^2}{(1-2b_2)^2(1-b_2)^2}$ |
| 1 | $\dfrac{4(b_1+m)(b_0(2b_2-1)+m(b_1+m))}{(1-2b_2)^2(b_2-1)}$ |
| 2 | $\dfrac{2\left(b_0(1-2b_2)^2+(b_1+m)(2b_1(b_2-1)+(4b_2-3)m)\right)}{(1-b_2)(1-2b_2)^2}$ |
| 3 | $\dfrac{4(b_1+m)}{2b_2-1}$ |
| 4 | $1$ |

showing that the moment combination $\int_E U(G)\,\mathrm{d}\,\mu$ indeed vanishes for a random variable $G$ having the law $\nu$ of the target distribution (as the $\Gamma$ expression is zero if the law of $G$ is $\nu$ by Theorem 3.1).

(iv) In order to understand the presence of the additional moment combination

$$\frac{\xi(1-b_2)}{2}\int_E Q^2(G)\,\mathrm{d}\,\mu$$

in the bound (26), let $\tilde{L}$ be the Markov diffusion generator of the diffusion (15) with mean $m$ and diffusion coefficient $\sigma^2(x) = b(x)$ as in the statement of Theorem 3.9, so that the Pearson distribution $\nu$ is its invariant measure and its support is $\tilde{E} = (l, u)$. Let $\tilde{F} = x - m$ and $\tilde{G} = \tilde{F} + m = x$. Then, $\tilde{F}$ is an eigenfunction of $\tilde{L}$ (as it is the first orthogonal polynomial with respect to $\nu$) and $\tilde{G}$ has distribution $\nu$. Indeed, for any smooth function $\varphi$, (17) yields

$$0 = \int_{\tilde{E}} \tilde{L}\tilde{G}\,\mathrm{d}\,\nu = \int_{\mathbb{R}} b(x)\mathbb{1}_{(l,u)}(x)\varphi''(x) - (x-m)\varphi'(x)\nu(dx),$$

where the right-hand side is exactly the Stein characterization (19). By Proposition 4.2 for $n = 1$, $\tilde{F}$ has chaos grade $\tilde{\eta} = 2(1-b_2)$. Therefore, $\xi = \eta - \tilde{\eta}$ measures how much the chaos grade $\eta$ of $G$ exceeds the chaos grade $\tilde{\eta}$ of $\tilde{G}$. If $G$ is replaced by a sequence $\{G_k : k \geq 0\}$ with chaos grades $\{\eta_k : k \geq 0\}$, as will be done in Proposition 3.13, then in order to converge, it is necessary that $\eta_k$ converges to $\tilde{\eta}$.

*Proof of Theorem 3.9.* As $LF = -\lambda F$ and $L1 = 0$, we have that $LG = -\lambda(G-m)$. Also, by definition $L^{-1}G = L^{-1}F = -\frac{1}{\lambda}F = -\frac{1}{\lambda}(G-m)$. Therefore, also using the fact that $\Gamma$ vanishes if any of its two arguments is a constant, it follows that

$$\begin{aligned}\Gamma\left(G,-L^{-1}G\right) &= \frac{1}{\lambda}\Gamma\left(G-m,G-m\right)\\ &= \frac{1}{2\lambda}\left(L+2\lambda\,\mathrm{Id}\right)(G-m)^2\\ &= \frac{1}{2\lambda}\left(L+2\lambda\,\mathrm{Id}\right)(G^2-2mG+m^2).\end{aligned}$$

Using this identity, it is straightforward to verify that the polynomial $Q$ satisfies

$$\Gamma\left(G,-L^{-1}G\right) - b(G) = \frac{1}{2\lambda}(L+2(1-b_2)\lambda\,\mathrm{Id})Q(G), \tag{27}$$

so that we can write, using $\xi = \eta - 2(1-b_2)$,

$$\begin{aligned}\int_E \left(\Gamma(G,-L^{-1}G) - b(G)\right)^2 \mathrm{d}\mu &= \int_E \left(\frac{1}{2\lambda}(L+2(1-b_2)\lambda\,\mathrm{Id})Q(G)\right)^2 \mathrm{d}\mu\\ &= \frac{1}{4\lambda^2}\int_E ((L+\eta\lambda\,\mathrm{Id})Q(G) - \xi\lambda Q(G))^2\,\mathrm{d}\mu\\ &= \frac{1}{4\lambda^2}\left(\int_E ((L+\eta\lambda\,\mathrm{Id})Q(G))^2\,\mathrm{d}\mu + R_\eta(G)\right),\end{aligned} \tag{28}$$

where

$$\begin{aligned}R_\eta(G) &= \xi^2\lambda^2\int_E Q^2(G)\,\mathrm{d}\mu - 2\lambda\xi\int_E Q(G)(L+\eta\lambda\,\mathrm{Id})Q(G)\,\mathrm{d}\mu\\ &= -2\lambda\xi\int_E Q(G)(L+2(1-b_2)\lambda\,\mathrm{Id})Q(G)\,\mathrm{d}\mu - \xi^2\lambda^2\int_E Q^2(G)\,\mathrm{d}\mu.\end{aligned} \tag{29}$$

As $L$ is symmetric,

$$\begin{aligned}\int_E ((L+\eta\lambda\,\mathrm{Id})(Q(G)))^2\,\mathrm{d}\mu &= \int_E Q(G)(L+\eta\lambda\,\mathrm{Id})^2 Q(G)\,\mathrm{d}\mu\\ &= \eta\lambda\int_E Q(G)(L+\eta\lambda\,\mathrm{Id})Q(G)\,\mathrm{d}\mu\\ &+ \int_E Q(G)L(L+\eta\lambda\,\mathrm{Id})Q(G)\,\mathrm{d}\mu\\ &\leq \eta\lambda\int_E Q(G)(L+\eta\lambda\,\mathrm{Id})Q(G)\,\mathrm{d}\mu\\ &= \eta\lambda\int_E Q(G)(L+2(1-b_2)\lambda\,\mathrm{Id})Q(G)\,\mathrm{d}\mu\\ &+ \eta\xi\lambda^2\int_E Q^2(G)\,\mathrm{d}\mu,\end{aligned} \tag{30}$$

where the inequality follows from the fact that

$$\int_E Q(G)L(L+\eta\lambda\,\mathrm{Id})Q(G)\,\mathrm{d}\mu \leq 0.$$

Indeed, as by assumption

$$Q(G) = \sum_{-\kappa\in\Lambda:\ \kappa\leq\eta\lambda} \pi_\kappa(Q(G)),$$

where $\pi_\kappa(Q(G))$ denotes the orthogonal projection of $Q(G)$ onto the eigenspace $\ker(L+\kappa\,\mathrm{Id})$, one has

$$\begin{aligned}&\int_E Q(G)L(L+\eta\lambda\,\mathrm{Id})Q(G)\,\mathrm{d}\mu\\ &\qquad=\sum_{-\kappa\in\Lambda:\ \kappa\le\eta\lambda}\int_E \pi_\kappa(Q(G))L(L+\eta\lambda\,\mathrm{Id})\pi_\kappa(Q(G))\,\mathrm{d}\mu\\ &\qquad=-\sum_{-\kappa\in\Lambda:\ \kappa\le\eta\lambda}\kappa(\eta\lambda-\kappa)\int_E \pi_\kappa(Q(G))^2\,\mathrm{d}\mu\le 0.\end{aligned}$$

Plugging (30) and (29) into (28) yields

$$\begin{aligned}\int_E\left(\Gamma(G,-L^{-1}G)-b(G)\right)^2\mathrm{d}\mu\le\frac{\eta-2\xi}{4\lambda}&\int_E Q(G)(L+2(1-b_2)\lambda\,\mathrm{Id})Q(G)\,\mathrm{d}\mu\\ &+\frac{\xi(1-b_2)}{2}\int_E Q^2(G)\,\mathrm{d}\mu.\end{aligned}$$

In order to prove that

$$\int_E Q(G)(L+2(1-b_2)\lambda\,\mathrm{Id})Q(G)\,\mathrm{d}\mu=2\lambda\int_E U(G)\,\mathrm{d}\mu,\tag{31}$$

we use integration by parts and the diffusion property of $\Gamma$, as well as the fact that $(Q'(x)^3)'=6Q'(x)^2$, to write

$$\begin{aligned}\int_E Q(G)LQ(G)\,\mathrm{d}\mu&=-\int_E\Gamma(Q(G),Q(G))\,\mathrm{d}\mu\\ &=-\int_E(Q'(G))^2\Gamma(G,G)\,\mathrm{d}\mu\\ &=-\frac{1}{6}\int_E\Gamma\left((Q'(G))^3,G\right)\mathrm{d}\mu\\ &=\frac{1}{6}\int_E(Q'(G))^3\,LG\,\mathrm{d}\mu\\ &=-\frac{\lambda}{6}\int_E(Q'(G))^3\,(G-m)\,\mathrm{d}\mu.\end{aligned}$$

Hence,

$$\begin{aligned}&\int_E\left(\Gamma(G,-L^{-1}G)-b(G)\right)^2\mathrm{d}\mu\\ &\qquad\le\frac{\eta-2\xi}{2}\int_E U(G)\,\mathrm{d}\mu+\frac{\xi(1-b_2)}{2}\int_E Q^2(G)\,\mathrm{d}\mu\end{aligned}\tag{32}$$

proving (26) since $(\eta-2\xi)/2=2(1-b_2-\eta/4)$. Note finally that if $\eta\le 2(1-b_2)$, then $\xi\le 0$ and hence $\xi(1-b_2)\le 0$, so that the second term in (32) is negative and can be dropped. This proves (25). $\square$

*Remark* 3.11. In view of the general bound (22) obtained in Theorem 3.2, it is natural to ask whether the quantity

$$\int_E\left(\Gamma(G,-L^{-1}G)-\theta^{-1}\tau(G)\right)^2 d\mu,$$

where $G$ is an eigenfunction, can be bounded by the first four moments of $G$, when $\tau$ is the diffusion coefficient of a diffusion with invariant measure outside of the Pearson class. Inspecting the proof of Theorem 3.9, one sees that, after expanding the square, the first four moments appear naturally from the term $\int_E \Gamma(G, -L^{-1}G)^2 \,\mathrm{d}\,\mu$. The remaining terms $\int_E \tau(G)\Gamma(G, -L^{-1}G) \,\mathrm{d}\,\mu$ and $\int_E \tau(G)^2 \,\mathrm{d}\,\mu$ yield moments up to order four if, and only if the diffusion coefficient $\tau$ is a polynomial of degree at most two, for which the corresponding invariant measures are exactly the Pearson distributions. In this sense, the class of Pearson target laws for which we provide four moment theorems is exhaustive.

By combining Theorem 3.9 with Theorem 3.2, we obtain quantitative moment bounds for suitable distances.

**Proposition 3.12.** *In the setting and with the notation of Theorem 3.9, let $Z$ be a random variable with distribution $\nu$. Then, if $\eta \leq 2(1-b_2)$, one has*

$$d_{\mathscr{H}}(G,Z) \leq c_{\mathscr{H}} \sqrt{\left(1 - b_2 - \frac{\eta}{4}\right) \int_E U(G) \,\mathrm{d}\,\mu},$$

*whereas if $\eta > 2(1-b_2)$, one has*

$$d_{\mathscr{H}}(G,Z) \leq c_{\mathscr{H}} \sqrt{\left(1 - b_2 - \frac{\eta}{4}\right) \int_E U(G) \,\mathrm{d}\,\mu + \frac{\xi(1-b_2)}{2} \int_E Q^2(G) \,\mathrm{d}\,\mu}.$$

*Here, $d_{\mathscr{H}}$ denotes an admissible distance for $G$ and $Z$, defined via a separating class $\mathscr{H}$ of absolutely continuous test functions such that $\sup_{h\in\mathscr{H}} \|h'\|_\infty < \infty$. The positive constant $c_{\mathscr{H}}$ depends solely on the class $\mathscr{H}$.*

*Proof.* We have to check that the function $\sigma^2$ satisfies the assumptions of Lemma 2.5. This is immediate by taking $g = \sigma^2$. We then apply Cauchy-Schwarz to (3.2) and use (25) and (26). □

At this point it is straightforward to state the following quantitative Four Moments Theorems for approximation of any Pearson distribution admitting at least four moments by a sequence of chaotic eigenfunctions.

**Theorem 3.13.** *Let $\nu$ be a Pearson distribution associated to the diffusion given by* (15) *with mean $m$ and diffusion coefficient $\sigma^2(x) = b(x) = b_2x^2 + b_1x + b_0$, and let $Z$ be a random variable with law $\nu$. For $k \in \mathbb{N}$, let $F_k$ be a chaotic eigenfunction with chaos grade $\eta_k$ of a Markov diffusion generator $L_k$ and let $G_k = F_k + m$. Furthermore, let $d_{\mathscr{H}}$ be an admissible distance for $\{G_k : k \in \mathbb{N}\} \cup \{Z\}$, defined via a separating class $\mathscr{H}$ of absolutely continuous test functions with uniformly bounded derivative. Then, if $\eta_k \leq 2(1-b_2)$, one has*

$$d_{\mathscr{H}}(G_k,Z) \leq c_{\mathscr{H}} \sqrt{\left(1 - b_2 - \frac{\eta_k}{4}\right) \int_E U(G_k) \,\mathrm{d}\,\mu}$$

*whereas if $\eta_k > 2(1-b_2)$, one has*

$$d_{\mathscr{H}}(G_k,Z) \leq c_{\mathscr{H}} \sqrt{\left(1 - b_2 - \frac{\eta}{4}\right) \int_E U(G_k) \,\mathrm{d}\,\mu + \frac{\xi_k(1-b_2)}{2} \int_E Q^2(G_k) \,\mathrm{d}\,\mu},$$

*where $\xi_k = \eta_k - 2(1-b_2)$. Here, $c_{\mathscr{H}}$ is a positive constant solely depending on the separating class $\mathscr{H}$. In particular, the following two conditions are sufficient for the sequence $\{G_k : n \geq 0\}$ to converge in distribution to $Z$:*

*(i)*

$$\int_E U(G_k)\,\mathrm{d}\mu \to 0. \tag{33}$$

*(ii) For every subsequence $(\eta_{k_r})$ of $(\eta_k)$ such that $\eta_{k_r} > 2(1-b_2)$ for every $r \in \mathbb{N}$ one has*

$$\sup_{r\in\mathbb{N}} \int_E Q^2(G_{k_r})\,\mathrm{d}\mu < \infty$$

*and $\eta_{k_r} \to 2(1-b_2)$.*

*Proof.* This is an immediate consequence of Theorem 3.13. The sufficient condition (ii) ensures that the second term on the right-hand side of (26) converges to zero. □

*Remark* 3.14. To the best of our knowledge, Theorem 3.13 is the first instance where moment conditions are given in order to converge to heavy-tailed distributions (which are not characterized by moments). Furthermore, the main results (all non-heavy-tailed) of [NP05, NP09a, NP09b, Led12, ACP14] are included as particular cases and in a unified way.

*Example* 3.15. Let us give some explicit examples of the moment combinations appearing in Condition (33) for several target distributions. To improve readability, we abbreviate the $p$-th moment $\int_E G_k^p\,\mathrm{d}\mu$ by $m_p(G_k)$.

(i) For convergence towards a centered Gaussian distribution with variance $\sigma^2$, we have $b(x) = \sigma^2$ and $m = 0$, so that by Table 1, we get that $c_0 = \sigma^4$, $c_1 = 0$, $c_2 = -2\sigma^2$, $c_3 = 0$ and $c_4 = \frac{1}{3}$, hence recovering the well-known moment condition

$$\frac{1}{3}m_4(G_k) - 2\sigma^2 m_2(G_k) + \sigma^4 \to 0,$$

which becomes $m_4(G_k) \to 3$ when $m_2(G_k) = \sigma^2 = 1$.

(ii) For a (heavy-tailed) Student $t$-distribution with mean zero and $\tau$ degrees of freedom (which is a particular case of a Skew $t$-distribution with parameters $m = \frac{\tau+1}{2}$, $\lambda = \nu = 0$ and $\alpha = \sqrt{\tau}$), we have $b(x) = \frac{x^2}{\tau-1} + \frac{\tau}{\tau-1}$. Therefore, the moment condition becomes

$$\frac{(\tau-4)}{3(\tau-1)}m_4(G_k) - \frac{2\tau}{(\tau-1)}m_2(G_k) + \frac{\tau^2}{\tau^2-3\tau+2} \to 0.$$

This moment condition is new.

(iii) For the inverse gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, which is non-centered (as opposed to the two previous examples) with mean $\frac{\beta}{\alpha-1}$, we have $b(x) = \frac{x^2}{\alpha-1}$. We hence obtain new moment conditions as well, ensuring convergence to the (heavy-tailed) inverse gamma distribution. For instance, setting the shape parameter $\alpha = 5$, we get that

$$\frac{1}{12}m_4(G_k) - \frac{\beta}{3}m_3(G_k) + \frac{\beta^2}{4}m_2(G_k) - \frac{\beta^3}{24}m_1(G_k) \to 0.$$

## 4. Pearson chaos

As an application of our results, we treat the case where the converging sequence of chaotic eigenfunctions itself comes from a generator associated to a Pearson law. To avoid technicalities, we present here only the finite-dimensional case, analogous results in infinite dimension can be obtained in a similar way. We begin by describing a general and well-known tensorization procedure of Markov generators.

Fix $N \geq 2$ and, for $1 \leq i \leq N$, let $\mathcal{L}_i$ be a generator with invariant probability measure $\mu_i$ and $L^2$-domain $\mathcal{D}(\mathcal{L}_i) \subseteq L^2(E_i, \mathcal{F}_i, \mu_i)$. Let $(E, \mathcal{F}, \mu)$ be the product of the probability spaces $(E_i, \mathcal{F}_i, \mu_i)$. Then we can define a generator $L_N = \otimes_{i=1}^N \mathcal{L}_i$ on $\mathcal{D}(L_N) = \bigotimes_{i=1}^N \mathcal{D}(\mathcal{L}_i)$ by

$$L_N \left( F_1 \times F_2 \times \cdots \times F_N \right) = \sum_{i=1}^{N} F_1 \times \cdots \times F_{i-1} \times (\mathcal{L}_i F_i) \times F_{i+1} \times \cdots \times F_N.$$

From this definition, it follows that if $F_i$ is an eigenfunction of $\mathcal{L}_i$ with eigenvalue $\lambda_i$, then $F = \oplus_{i=1}^N F_i$ is an eigenfunction of $L_N$ with eigenvalue $\lambda = \sum_{i=1}^N \lambda_i$. The following corollary describes how the chaos grade behaves under tensorization.

**Corollary 4.1.** *In the above setting, let each eigenfunction $F_i$ be chaotic with chaos grade $\eta_i$. Then $F$ is chaotic and its chaos grade $\eta$ is bounded as follows:*

$$\min \left\{ \eta_1, \eta_2, \ldots, \eta_N \right\} \leq \eta \leq \max \left\{ \eta_1, \eta_2, \ldots, \eta_N \right\}.$$

*The above inequalities become equalities, if, and only if, all of the chaos grades $\eta_i$ are equal.*

*Proof.* By definition, the squares $F_i^2$ can be expanded as sums of eigenfunctions, with the eigenvalue of largest magnitude in such an expansion being $\lambda_i \eta_i$. Therefore, $F^2$ can also be expanded as a sum of eigenfunctions, with the eigenvalue of largest magnitude, say $\lambda_{\max}$, being given by

$$\lambda_{\max} = \sum_{i=1}^{N} \lambda_i \eta_i.$$

Applying the definition of chaos grade (see Definition 3.5) now yields that

$$\eta = \frac{\lambda_{max}}{\lambda} = \frac{\sum_{i=1}^N \lambda_i \eta_i}{\sum_{i=1}^N \lambda_i},$$

from which the assertion follows as all $\lambda_i$ have the same sign. □

In the following proposition, we calculate the possible range of values of the chaos grade for eigenfunctions related to all six Pearson distributions.

**Proposition 4.2.** *Let $L$ be the generator associated to a Pearson diffusion defined by* (15) *and denote the eigenvalues of $L$ by $-\lambda_n$ where $\lambda_n = n(1-(n-1)b_2)\theta$ for $b_2 < \frac{1}{2n-1}$. Let $F_n$ be an eigenfunction of $L$ with respect to $-\lambda_n$. Then $F_n$ is chaotic, if, and only if, $b_2 < \frac{1}{4n-1}$, and in this case its chaos grade $\eta_n$ is given by*

$$\eta_n = \eta_n(b_2) = \begin{cases} 2 & \text{if } b_2 = 0, \\ 2\left(1 + \frac{n}{n-1-\frac{1}{b_2}}\right) & \text{if } b_2 \neq 0. \end{cases} \tag{34}$$

*Furthermore, the following is true.*

*(i) If $\mu$ is a Student, F- or inverse Gamma distribution, then $\eta_n \in \left(\frac{4}{3}, 2-2b_2\right]$.*
*(ii) If $\mu$ is a Gaussian or Gamma distribution, then $\eta_n = 2$.*
*(iii) If $\mu$ is a Beta distribution then $\eta_n \in (4, 2-2b_2]$, if $b_2 < -1$, $\eta_n = 4$, if $b_2 = -1$ and $\eta_n \in [2-2b_2, 4)$, if $-1 < b_2 < 0$.*

*Proof.* An eigenfunction $F_n$ of a Pearson generator with respect to the eigenvalue $-\lambda_n = n(1-(n-1)b_2)\theta$ is an orthogonal polynomial of degree $n$. Its square is then a polynomial of degree $2n$. In order for $F_n^2$ to be expressable as a sum of square integrable eigenfunctions, we therefore need that the first $2n$ eigenfunctions of $L$ are square integrable, or equivalently that moments up to order $4n$ exist. Hence, by (18), the condition required is

$$b_2 < \frac{1}{4n-1}. \tag{35}$$

Let us assume that the above inequality is satisfied. Then, by its very definition, $\eta_n$ is given by the quotient of the $2n$-th eigenvalue with the $n$-th one. Indeed, as $\eta_n$ is the multiplicative factor that indicates what eigenvalue the highest-order eigenfunction in the decomposition of the square of $F_n$ is associated to. On the other hand, we know that the square of the polynomial eigenfunction of degree $n$ produces a sum of polynomial eigenfunctions up to degree $2n$, corresponding to the eigenvalue $-\lambda_{2n}$. Hence we have

$$\eta_n = \frac{\lambda_{2n}}{\lambda_n} = \frac{2n(1-(2n-1)b_2)\theta}{n(1-(n-1)b_2)\theta}, \tag{36}$$

so that (34) follows. Assertion (ii) is immediate as in this case $b_2 = 0$ and the chaos grade is constant. In order to show assertion (i) in which $b_2 > 0$, note that the function $n \mapsto \eta_n(b_2)$ is decreasing. Therefore, the largest possible chaos grade is obtained by taking $n = 1$ in (36), which gives $2(1-b_2)$. On the other hand, as by (35), $n < \frac{1}{4}\left(\frac{1}{b_2}+1\right)$, the lower bound $\frac{4}{3}$ of the chaos grade is obtained by taking $n = \left\lfloor \frac{1}{4}\left(\frac{1}{b_2}+1\right)\right\rfloor$. Assertion (iii) where $b_2 < 0$ follows in a similar way. □

Proposition 4.2 shows that on a global level, the chaos grade $\eta$ of chaotic eigenfunctions of a Pearson generator lies in the interval $\left(\frac{4}{3}, \infty\right)$. Furthermore, all values in this interval can be attained, in the sense that if $x$ is such a value, then there exists a generator $L$ of a Pearson diffusion (15) which has a chaotic eigenfunction of chaos grade $x$. The six types of Pearson distributions are partitioned into three classes with disjoint intervals for the chaos grade values of the corresponding eigenfunctions. These intervals are all of the form

$$\{2(1-b_2) : b_2 \in I\},$$

where $I$ is the set of allowed values for the corresponding class, i.e., $I = (-\infty, 0)$ for the class of student, $F$- and inverse Gamma distributions, $I = \{0\}$ for Gaussian and Gamma distributions and $I = (0, \infty)$ for the Beta distributions.

Applying the tensorization procedure described above to the case where all generators $\mathcal{L}_i$ are equal to some generator $L$ of a Pearson diffusion immediately yields the following result.

**Theorem 4.3.** *Let $\mu$ be a Pearson distribution and $L$ be the associated Markov generator. Denote its eigenvalues by $\{-\lambda_i : 0 \leq i < I\}$, where $I \in \mathbb{N} \cup \{\infty\}$ and such that $\lambda_i < \lambda_{i+1}$. Furthermore, denote by $P_i$ the $i$-th orthogonal polynomial associated to*

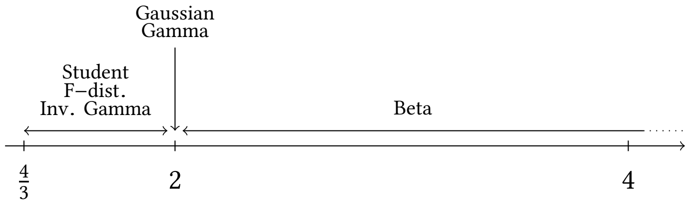


FIGURE 1. Possible chaos grades for the Pearson distributions

*$\mu$. Let $L_N = L^{\otimes N}$ be the generator obtained by the tensorization procedure described above and denote by $\mu_N$ the associated product measure. Then the set of eigenvalues of $L_N$ is given by*

$$S = \left\{ -\sum_{i=1}^{N} \lambda_{k_i} : k_1, \dots, k_N \in I \right\}.$$

*If $-\lambda = -\sum_{i=1}^{N} \lambda_{k_i}$ is such an eigenvalue, then all eigenfunctions $F$ of $L_N$ with respect to $-\lambda$ are of the form*

$$F = \sum_{|\alpha|=p} a_\alpha P_\alpha,$$

*where*

*(i) $p = \sum_{i=1}^{N} k_i$,*
*(ii) the sum is taken over all $N$-dimensional multi-indices $\alpha = (\alpha_1, \dots, \alpha_N)$ of order $p$,*
*(iii) the $a_\alpha$ are real constants,*
*(iv) $P_\alpha(x) = P_\alpha(x_1, x_2, \dots, x_N) = \prod_{i=1}^{N} P_{\alpha_i}(x_i)$.*

Combining Corollary 4.1 with Proposition 4.2 and the discussion thereafter, we see that for the six classes of Pearson distributions the intervals for the chaos grades of the respective chaotic eigenfunctions are invariant under tensorization. In other words, the chaos grades of chaotic eigenfunctions of $L_N$

(i) assume values in the interval $\left(\frac{4}{3}, 2\right)$, if the tensorized distribution is Student, $F$- or inverse Gamma,
(ii) are equal to two in the case of tensorized Gaussian or Gamma distributions,
(iii) lie in the interval $(2, \infty)$ if the distribution is Beta.

Coming back to the Four Moments Theorems proved in Section 3.2, the possible chaos grades also yield a heuristic about "compatible" Pearson distributions, in the sense that one can be obtained as a limit of a chaos of another Pearson distribution. Recall from Section 3.2 (in particular Remark 3.10.iv) that if we want to approximate a random variable $Z$ with a Pearson law and chaos grade $\tilde{\eta}$ to be the limit of a sequence $(G_n)$ of chaotic random variables with corresponding chaos grade sequence $(\eta_n)$, we need that $\eta_n \leq \widetilde{\eta}$ or $\eta_n \to \widetilde{\eta}$, where $\widetilde{\eta}$ is the chaos grade of $Z$ when seen as a chaotic random variable itself. For example, if $Z$ has a Gaussian or Gamma distribution, then $\widetilde{\eta} = 2$. Therefore, chaotic random variables coming from a heavy tailed Pearson chaos are compatible, as in this case we always have $\eta_n \leq 2$. The Gamma and Gaussian chaos is of course compatible as well as here the two chaos grades coincide and for convergence from Beta chaos to a Gaussian or Gamma distribution, our conditions require that $\eta_n \to 2$. This translates to the parameters of the underlying invariant Beta measure growing to infinity.

Taking $Z$ to be a heavy tailed Pearson distribution yields a chaos grade $\widetilde{\eta}$ which is strictly less than two. Here, our conditions suggest that only heavy-tailed chaos are compatible. The aforementioned heuristic could likely be made rigorous by a detailed study of the carré du champ characterization given in Theorem 3.1 and is left for future research.

## References


[ACP14] Ehsan Azmoodeh, Simon Campese, and Guillaume Poly, *Fourth Moment Theorems for Markov diffusion generators*, Journal of Functional Analysis **266** (2014), no. 4, 2341–2359. MR 3150163

[ALŠ13] F. Avram, N. N. Leonenko, and N. Šuvak, *On spectral analysis of heavy-tailed Kolmogorov-Pearson diffusions*, Markov Processes and Related Fields **19** (2013), no. 2, 249–298. MR 3113945

[Bak14] Dominique Bakry, *Symmetric diffusions with polynomial eigenvectors*, Stochastic analysis and applications 2014, Springer Proc. Math. Stat., vol. 100, Springer, Cham, 2014, DOI: 10.1007/978-3-319-11292-3_2, pp. 25–49. MR 3332708

[BÉ85] D. Bakry and Michel Émery, *Diffusions hypercontractives*, Séminaire de probabilités, XIX, 1983/84, Lecture Notes in Math., vol. 1123, Springer, Berlin, 1985, pp. 177–206. MR 889476

[BGL14] Dominique Bakry, Ivan Gentil, and Michel Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 348, Springer, Cham, 2014, DOI: 10.1007/978-3-319-00227-9. MR 3155209

[BH91] Nicolas Bouleau and Francis Hirsch, *Dirichlet forms and analysis on Wiener space*, de Gruyter Studies in Mathematics, vol. 14, Walter de Gruyter & Co., Berlin, 1991. MR 1133391

[BSSr05] Bo Martin Bibby, Ib Michael Skovgaard, and Michael Sø rensen, *Diffusion-type models with given marginal distribution and autocorrelation function*, Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability **11** (2005), no. 2, 191–220. MR 2132002

[Dud02] R. M. Dudley, *Real analysis and probability*, Cambridge Studies in Advanced Mathematics, vol. 74, Cambridge University Press, Cambridge, 2002. MR 1932358

[EVq15] Richard Eden and Juan V\'ı quez, *Nourdin-Peccati analysis on Wiener and Wiener-Poisson space for general distributions*, Stochastic Processes and their Applications **125** (2015), no. 1, 182–216. MR 3274696

[FOT11] Masatoshi Fukushima, Yoichi Oshima, and Masayoshi Takeda, *Dirichlet forms and symmetric Markov processes*, extended ed., De Gruyter Studies in Mathematics, vol. 19, Walter de Gruyter & Co., Berlin, 2011. MR 2778606

[FS08] Julie Lyng Forman and Michael Sørensen, *The Pearson Diffusions: A Class of Statistically Tractable Diffusion Processes*, Scandinavian Journal of Statistics **35** (2008), no. 3, 438–465 (en).

[HNZ17] Yaozhong Hu, David Nualart, and Hongjuan Zhou, *Parameter estimation for fractional ornstein–uhlenbeck processes of general hurst parameter*, Statistical Inference for Stochastic Processes (2017).

[JKB94] Norman L. Johnson, Samuel Kotz, and N. Balakrishnan, *Continuous univariate distributions. Vol. 1*, second ed., Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Inc., New York, 1994. MR 1299979

[JKB95] ———, *Continuous univariate distributions. Vol. 2*, second ed., Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Inc., New York, 1995. MR 1326603

[KP17] Yoon Tae Kim and Hyun Suk Park, *Optimal Berry-Esseen bound for statistical estimations and its application to SPDE*, J. Multivariate Anal. **155** (2017), 284–304. MR 3607896

[KT12] Seiichiro Kusuoka and Ciprian A. Tudor, *Stein's method for invariant measures of diffusions via Malliavin calculus*, Stochastic Processes and their Applications **122** (2012), no. 4, 1627–1651. MR 2914766

[Led12] Michel Ledoux, *Chaos of a Markov operator and the fourth moment condition*, The Annals of Probability **40** (2012), no. 6, 2439–2459 (EN), Zentralblatt MATH identifier: 06114704.

[Maz97] Olivier Mazet, *Classification des semi-groupes de diffusion sur $\bf R$ associés à une famille de polynômes orthogonaux*, Séminaire de Probabilités, XXXI, Lecture Notes in Math., vol. 1655, Springer, Berlin, 1997, pp. 40–53. MR 1478714

[NOL08] David Nualart and Salvatore Ortiz-Latorre, *Central limit theorems for multiple stochastic integrals and Malliavin calculus*, Stochastic Processes and their Applications **118** (2008), no. 4, 614–628. MR 2394845

[NP05] David Nualart and Giovanni Peccati, *Central limit theorems for sequences of multiple stochastic integrals*, The Annals of Probability **33** (2005), no. 1, 177–193. MR 2118863

[NP09a] Ivan Nourdin and Giovanni Peccati, *Noncentral convergence of multiple integrals*, The Annals of Probability **37** (2009), no. 4, 1412–1426 (EN). MR MR2546749

[NP09b] ———, *Stein's method on Wiener chaos*, Probability Theory and Related Fields **145** (2009), no. 1-2, 75–118. MR 2520122

[NP12] ———, *Normal approximations with Malliavin calculus*, Cambridge Tracts in Mathematics, vol. 192, Cambridge University Press, Cambridge, 2012, From Stein's method to universality.

[Nua06] David Nualart, *The Malliavin calculus and related topics*, second ed., Probability and its Applications (New York), Springer-Verlag, Berlin, 2006.

[Pea95] Karl Pearson, *Contributions to the mathematical theory of evolution - II. Skew variation in homogeneous material*, Phil. Trans. R. Soc. Lond. A **186** (1895), 343–414 (en).

[Ste86] Charles Stein, *Approximate computation of expectations*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7, Institute of Mathematical Statistics, Hayward, CA, 1986. MR 882007

Boston University, Department of Mathematics and Statistics, 111 Cummington Mall, Boston, MA 02215, USA
*E-mail address*: bourguin@math.bu.edu

University of Luxembourg, Mathematics Research Unit, 6, rue Richard Coudenhove-Kalergi, 1359 Luxembourg, Luxembourg
*E-mail address*: simon.campese@uni.lu

Cardiff University, School of Mathematics, Senghennydd Road, Cardiff, Wales, UK, CF24 4AG
*E-mail address*: leonenkon@cardiff.ac.uk

Boston University, Department of Mathematics and Statistics, 111 Cummington Mall, Boston, MA 02215, USA
*E-mail address*: murad@bu.edu