The method of cumulants for the normal approximation[*]

Hanna Döring

Osnabrück University, Institute of Mathematics,
Albrechtstr. 28a, 49076 Osnabrück, Germany
e-mail: hanna.doering@uni-osnabrueck.de

Sabine Jansen

LMU Munich, Department of Mathematics,
Theresienstr. 39, 80333 Munich, Germany

Munich Center for Quantum Science and Technology (MCQST),
Schellingstr. 4, 80799 Munich, Germany
e-mail: jansen@math.lmu.de

and

Kristina Schubert

TU Dortmund University, Faculty of Mathematics,
Vogelpothsweg 87, 44227 Dortmund, Germany
e-mail: kristina.schubert@tu-dortmund.de

Abstract: The survey is dedicated to a celebrated series of quantitave results, developed by the Lithuanian school of probability, on the normal approximation for a real-valued random variable. The key ingredient is a bound on cumulants of the type $|\kappa_j(X)| \leq j!^{1+\gamma}/\Delta^{j-2}$, which is weaker than Cramér's condition of finite exponential moments. We give a self-contained proof of some of the "main lemmas" in a book by Saulis and Statulevičius (1989), and an accessible introduction to the Cramér-Petrov series. In addition, we explain relations with heavy-tailed Weibull variables, moderate deviations, and mod-phi convergence. We discuss some methods for bounding cumulants such as summability of mixed cumulants and dependency graphs, and briefly review a few recent applications of the method of cumulants for the normal approximation.

Contents

1. Introduction

1.1. Aims and scope of the article

The method of cumulants is a central tool in comparing the distribution of a random variable with the normal law. It enters the proof of central limit theorems, moderate and large deviation principles, Berry-Esseen bounds, and concentration inequalities in various fields of probability theory: stochastic geometry, random matrices, random graphs, random combinatorial structures, functionals of stochastic processes, mathematical biology—the list is not exhaustive.

The present survey shines a spotlight on a celebrated series of bounds developed by the Lithuanian school, notably Rudzkis, Saulis, and Statulevičius [99] and Bentkus and Rudzkis [9]. The bounds work under a condition on cumulants that allows for heavy-tailed behavior and is considerably weaker than the Cramér condition of finite exponential moments frequently invoked in the theory of large deviations. The conditions on cumulants can be verified in many situations of interest, beyond sums of independent random variables. In their monograph [101], Saulis and Statulevičius study applications, for example, to random processes with mixing, multiple stochastic integrals, and U-statistics.

Our primary aim is to give a self-contained and accessible presentation, including proofs, of the "main lemmas" from Chapter 2 in the book [101] by Saulis and Statulevičius; along the way, we correct a few minor errors in the proofs. We have not aimed at a reconstruction of all numerical constants. The presentation should be accessible to a reader with little exposure to complex-analytic or Fourier methods in probability or slightly arcane concepts such as the Cramér-Petrov series. The reader familiar with the classical books by Ibragimov and Linnik [64], Petrov [92], Gnedenko and Kolmogorov [44], or even Feller [35] will easily recognize a classical set of ideas, however an in-depth study of characteristic functions, inversion formulas, and asymptotic expansions are nowadays frequently eschewed in graduate probability courses and these methods are no longer part of every probabilist's toolbox. Moreover the presentation in [101] is extremely technical, making it very hard to extract the proof philosophy from the long series of technical estimates. To remedy this situation, we have strived to make explicit the logical structure and key ideas, notably the truncation procedures needed to deal with Taylor expansions with zero radius of convergence.

In addition, we mention a few exemplary applications and discuss the relation of the aforementioned results with other techniques and fields, in particular, large deviations for sums of heavy-tailed variables [33, 83], analytic combinatorics [43], and mod-phi convergence [39].

In the remaining part of the introduction, we define the cumulants, summarize the main bounds studied in the present survey, address methods for bounding cumulants, and list a few recent applications.

1.2. Cumulants

The *cumulants* of a real-valued random variable X are the numbers $\kappa_j \equiv \kappa_j(X)$, $j \in \mathbb{N}$, given by

$$\kappa_j(X) := (-\mathrm{i})^j \frac{\mathrm{d}^j}{\mathrm{d}t^j} \log \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big]\Big|_{t=0},$$

provided the derivative exists. Equivalently, assuming r-fold differentiability of the characteristic function at the origin, the cumulants of order $j = 1, \ldots, r$ are related to the Taylor expansion by

$$\log \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big] = \sum_{j=1}^{r} \frac{\kappa_j}{j!} (\mathrm{i}t)^j + o(t^r) \quad (t \to 0).$$

The cumulant of order 1 is the expected value $\kappa_1 = \mathbb{E}[X]$, the cumulant of order 2 is the variance $\kappa_2 = \mathbb{V}(X)$. More generally, there exists a recurrence formula between the centered moments $\mathbb{E}[(X - \mathbb{E}[X])^j]$ and the cumulants, which is how Thiele [110] originally defined them. Cumulants are often called *semi-invariants* because of the homogeneity $\kappa_j(\lambda X) = \lambda^j \kappa_j(X)$ and shift-invariance $\kappa_j(X + c) = \kappa_j(X)$ for $j \geq 2$. The name *cumulants* was proposed by Fisher and Wishart [42]; see Hald [52] for a historical account and a translation of Thiele's article from Danish to English.

There are many moment-to-cumulants formulas. We list a few to give a first impression of cumulants but emphasize that none of them, except perhaps the first, is used in the sequel. The most common relation, obtained from Faà di Bruno's formula, is

$$\kappa_j(X) = \sum_{m=1}^{j} \frac{(-1)^{m-1}}{m} \sum_{\substack{k_1 + \cdots + k_m = j \\ k_1, \ldots, k_m \geq 1}} \frac{j!}{k_1! \cdots k_m!} \prod_{\ell=1}^{m} \mathbb{E}\big[X^{k_\ell}\big].$$

Equivalently, the j-th cumulant is a sum over set partitions $\{B_1, \ldots, B_m\}$ of $\{1, \ldots, j\}$,

$$\kappa_j(X) = \sum_{m=1}^{j} \sum_{\{B_1, \ldots, B_m\}} (-1)^{m-1} (m-1)! \prod_{\ell=1}^{m} \mathbb{E}\big[X^{\#B_\ell}\big].$$

It can be obtained by a Möbius inversion on the lattice of set partitions, the function $(-1)^{m-1}(m-1)!$ is the Möbius function [108]. Another set of relations is obtained by differentiating the logarithm of the characteristic function: The relation

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big] = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big] \frac{\mathrm{d}}{\mathrm{d}t} \log \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big]$$

implies the recurrence relations

$$\kappa_j(X) = \mathbb{E}\big[X^j\big] - \sum_{r=1}^{j-1} \binom{j-1}{r-1} \mathbb{E}\big[X^{j-r}\big] \kappa_r(X).$$

Cramer's rule for solving linear equations yields an expression of the cumulant as the determinant of a Toeplitz matrix [97, Cor. 3.1],

$$\kappa_j(X) = (-1)^{j-1}(j-1)! \begin{vmatrix} \mathbb{E}X & \frac{\mathbb{E}X^2}{2!} & \cdots & \frac{\mathbb{E}X^{j-1}}{(j-1)!} & \frac{\mathbb{E}X^k}{k!} \\ 1 & \mathbb{E}X & \cdots & \frac{\mathbb{E}X^{j-2}}{(j-2)!} & \frac{\mathbb{E}X^{j-1}}{(j-1)!} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbb{E}X & \frac{\mathbb{E}X^2}{2!} \\ 0 & 0 & \cdots & 1 & \mathbb{E}X \end{vmatrix}.$$

Cumulants offer some advantages over moments. Crucially, the j-th order cumulant of a sum of independent random variables is simply the sum of the j-th order cumulants. For a standard normal random variable all cumulants of order $j \geq 3$ vanish. Cumulants of a Poisson random variable with parameter λ are all given by $\kappa_j \equiv \lambda$, while formulas for moments are more involved.

Cumulants help prove central limit theorems: A sequence $(X_n)_{n \in \mathbb{N}}$ of random variables converges in distribution to a standard normal variable if and only if the expectation and the variance go to zero and one, respectively, and in addition all higher-order cumulants go to zero. In fact for the higher-order cumulants it is enough to check that the cumulants with j larger than any fixed order $s \geq 3$ go to zero, see Janson [67]. Bounds on cumulants translate into quantitative bounds for the normal approximation.

1.3. Short description of the "main lemmas"

The main theorems discussed in the present survey apply to real-valued random variables X for which all moments, hence also all cumulants, exist and for which the cumulants can be bounded as

$$|\kappa_j(X)| \leq \frac{(j!)^{1+\gamma}}{\Delta^{j-2}} \qquad (j \geq 3)$$

with $\gamma \geq 0$, $\Delta > 0$. The variable X is assumed to be centered and normalized, $\mathbb{E}[X] = 0$ and $\mathbb{V}(X) = 1$. Following [1] we shall refer to this condition as the *Statulevičius condition*. The main results, roughly, are the following:

1. *Normal approximation with Cramér corrections.* Let $Z \sim \mathcal{N}(0, 1)$ be a standard normal variable. Then for $x \in (0, c\Delta^{1/(1+2\gamma)})$ with suitable constant $c > 0$,

$$\mathbb{P}(X \geq x) = e^{\tilde{L}(x)} \mathbb{P}(Z \geq x) \Big(1 + O\Big(\frac{x+1}{\Delta^{1/(1+2\gamma)}} \Big) \Big)$$

where $\tilde{L}(x)$ is related to the so-called *Cramér-Petrov series* (reviewed in Appendix A) and satisfies $|\tilde{L}(x)| = O(x^3/\Delta^{1/(1+2\gamma)})$. See Rudzkis, Saulis, and Statulevičius [99], Lemma 2.3 in [101], and Theorem 2.3 below.

2. *Bound on the Kolmogorov distance.* The following bound of Berry-Esseen type holds true:

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(X \leq x) - \mathbb{P}(Z \leq x) \right| \leq \frac{C}{\Delta^{1/(1+2\gamma)}}$$

for some constant $C > 0$. See Rudzkis, Saulis, and Statulevičius [99], Corollary 2.1 in [101], and Theorem 2.4 below.

3. *Concentration inequality.* Assuming $|\kappa_j| \leq \frac{1}{2} j!^{1+\gamma} H/\Delta^{j-2}$ for some parameters $H, \Delta > 0$ and all j (a minor variant of the Statulevičius condition), one has

$$\mathbb{P}(X \geq x) \leq \exp\left(-\frac{1}{2} \frac{x^2}{H + x^{2-\alpha}/\Delta^{\alpha}}\right), \quad \alpha := \frac{1}{1+\gamma}$$

$$\leq \exp\left(-\frac{1}{4} \min\left(\frac{x^2}{H}, (x\Delta)^{\alpha}\right)\right)$$

for all $x \geq 0$. See Bentkus and Rudzkis [9], the corollary to Lemma 2.4 in [101], and Theorem 2.5 below.

Let us briefly put these results in perspective and discuss the nature of the Statulevičius condition. When $\gamma = 0$, the condition implies that the cumulant generating function $\varphi(t) = \log \mathbb{E}[\exp(tX)] = \sum_j \kappa_j t^j/j!$ is analytic in a neighborhood of the origin and finite for $|t| < \Delta$. This is precisely Cramér's condition of finite exponential moments. The normal approximation with Cramér corrections is proven with standard techniques also employed for sums of independent identically distributed random variables [64, Chapter 8]. The concentration inequality is similar to the Bernstein inequality [111].

The bounds are more intriguing for $\gamma > 0$. In that case the Taylor expansions $\sum_j \kappa_j t^j/j!$ of the cumulant generating function can have zero radius of convergence and the random variable can be heavy-tailed, meaning that it has infinite exponential moments $\mathbb{E}[\exp(tX)]$ for arbitrarily small $t \neq 0$. The concentration inequality shows that the tails of X have at least stretched exponential decay $O(\exp(-\text{const } x^{\alpha}))$ with $\alpha = 1/(1+\gamma)$, i.e., X has (in the worst case) Weibull-like tails. In fact there is equivalence: It is known that the Statulevičius condition holds true if and only if Linnik's condition $\mathbb{E}[\exp(\delta|X|^{\alpha})] < \infty$, for some $\delta > 0$, is satisfied, see Section 2.4. Results on large deviations under conditions more general than Linnik's condition are available as well, see Section 2.4. Hence, the concentration inequality morally says that *if a random variable has cumulants similar to those of a Weibull-like variable, then it also has Weibull-like tails.*

This result is rather amazing. True, it is well-known that statements on the tails of a random variable can be inferred from information on the characteristic function $\chi(t) = \mathbb{E}[\exp(itX)]$ near $t = 0$— or, if the random variable is non-negative or with values in $\mathbb{N}_0$, from the behavior Laplace transform $\mathbb{E}[\exp(-\lambda X)]$ as $\lambda \downarrow 0$ or the behavior of the probability generating function $G(z) = \mathbb{E}[z^X]$ as $z \to 1$. Such relations are at the heart of analytic proofs of limit theorems in probability and combinatorics with Fourier analysis, Tauberian theorems, or

complex analysis [64, 43]. However, it is not clear at all that the coefficients of a divergent Taylor expansion carry enough information to allow for rigorous statements.

Proofs for $\gamma > 0$ require ingenious truncation procedures. Some of them are similar to procedures employed for large deviations of sums of heavy-tailed random variables, see Section 3.1. The critical scale $\Delta^{1/(1+2\gamma)}$ is comparable to the monomial zones of attraction to the normal law and "Cramér's system of limiting tails" for sums of i.i.d. random variables discussed by Ibragimov and Linnik [64, Chapters 9 and 10]. Let us stress that the critical scale is not just some technical limitation. For sums of i.i.d. Weibull-like variables, it corresponds exactly to the scale at which the normal approximation ceases to be good and heavy-tailed effects kick in [85], see also Section 2.3.2 for an elementary example.

Note that the same cumulant bound also implies Rosenthal type bounds i.e. estimates for the difference of the k-th moment to the corresponding moment of the normal distribution, see [31]. Connections with mod-phi convergence and moderate deviation principles are discussed in Section 3.1 and 3.2.

1.4. How to bound cumulants

There is no one-size-fits-all way to bound cumulants. Nevertheless, there are some repeating features, depending on the type of bound one seeks to establish and the type of random variable under investigation. For the bound, it matters whether one is after an analytic bound $\gamma = 0$ or after the weaker case $\gamma > 0$. Models fall broadly in two classes: first, random variables built out of fields or processes with underlying independence or good control on dependencies and decay of correlations, for example, empirical mean for a stationary ergodic process or functionals of Poisson point processes; second, models directly defined on more complex structures, including random matrices, models from analytic and probabilistic number theory, or random combinatorial structures.

The Statulevičius bound for $\gamma = 0$ implies that the cumulant generating function is analytic. Conversely, when the cumulant generating function is analytic in a neighborhood of the origin, then the cumulants can be bounded by applying Cauchy's formula. Accordingly one may shift perspective away from the cumulants and focus directly on generating functions. This is especially helpful for random combinatorial structures, where often the recursive structure of combinatorial objects translates into functional equations for generating functions and information on the analytic behavior [43]. Working directly with generating functions is also of advantage for random matrices and probabilistic number theory [39].

Analyticity of the cumulant generating function is equivalent to zero-freeness of the moment generating function. The role of analyticity and zero-freeness was already emphasized in Bryc's central limit theorem [16]. In statistical mechanics, controlling zeros is related to Lee-Yang theory and relations with central limit theorems were explored, for example, by Iagolnitzer and Souillard [63] or Ruelle, Pittel, Lebowitz and Speer [75]; see also [39, Chapter 8.1]. For new results and an account of modern developments, the reader is referred to Michelen

and Sahasrabudhe [81, 82] (Section 7 in [82] has the telling title "Taming the cumulant sequence").

When a direct control of generating functions is not possible, cumulants are often treated with combinatorial bounds and correlation estimates. Two prototypes are sums of dependent random variables and U-statistics of sequences of independent random variables [101, Chapters 4 and 5.1] or m-dependent vectors [53, 55]. Consider for example a sequence of independent random variables $(X_n)_{n\in\mathbb{N}}$ and the random variable $Y = \sum_{i=1}^{n-1} X_i X_{i+1}$. The cumulant of Y is not the sum of the cumulants $\kappa(X_i X_j)$ because $X_i X_{i+1}$ and $X_k X_{k+1}$ are not independent for $\{i, i+1\} \cap \{k, k+1\} \neq \varnothing$. But clearly for a given nearest-neighbor pair $\alpha = \{i, i+1\}$ the number of pairs $\beta = \{j, j+1\}$ with $\alpha \cap \beta \neq \varnothing$ is bounded by 2. This can be exploited with *dependency graphs*, see Section 3.4.

Another example is when the input variables X_n, $n \in \mathbb{N}$, are not independent. The cumulants of the partial sum are given by the sum of mixed cumulants $\kappa(X_{\alpha_1}, \ldots, X_{\alpha_j})$ [76] (see Eq. (3.4) below) as

$$\kappa_j(X_1 + \cdots + X_n) = \sum_{1 \leq \alpha_1, \ldots, \alpha_j \leq n} \kappa(X_{\alpha_1}, \ldots X_{\alpha_j}),$$

hence

$$\left|\kappa_j(X_1 + \cdots + X_n)\right| \leq n \sup_{\alpha_1 \in \mathbb{N}} \sum_{\alpha_2, \ldots, \alpha_j \in \mathbb{N}} \left|\kappa(X_{\alpha_1}, \ldots, X_{\alpha_j})\right|.$$

Thus bounds on the cumulants of the sum are intimately tied to summability properties of mixed cumulants. Analogous relations apply in the context of point processes; here summability of mixed cumulants is replaced with bounds on the total variation of reduced (factorial) cumulant measures, which leads to the notions of weak and strong *Brillinger mixing* [15, 65]. In the language of statistical mechanics, Brillinger mixing corresponds to integrability of truncated correlation functions, a condition typically satisfied by Gibbs point processes at low density [100].

Brillinger mixing can be rather difficult to establish directly. An alternative approach, still focused on the decay of correlations, was devised by Baryshnikov and Yukich [8] and then further developed, see Błaszczyszyn, Yogeshwaran and Yukich [11] and the references therein. Eichelsbacher, Raič and Schreiber [32] pushed the method all the way up to the Statulevičius condition. The method works with approximate factorization properties of moment measures; cumulants are represented as combinations of *semi-cluster measures*. Combinatorics enter when bounding the number of summands in the latter representation [32, Lemma 3.2].

For functionals of Markov chains or stochastic processes with mixing, it is convenient to work with another set of quantities, called *centered moments* [101, Chapter 4], *higher-order covariances* [57] or *Boolean cumulants* [37, Section 10]. Mixing properties of the underlying stochastic process lead to good bounds on the centered moments, and then bounds on cumulants are deduced from a Boolean-to-classical cumulants formula.

For Poisson or Gaussian input data, another class of methods builds on diagrammatic formulas for cumulants and chaos decompositions (related to Feynman diagrams and Fock spaces in mathematical physics). This applies to multiple stochastic integrals of Brownian motion or of Poisson point processes [101, Chapter 5.3]. A modern account of diagrammatic formulas is given by Peccati and Taqqu [89]. Useful formulas for cumulants can also be derived using Malliavin calculus and the infinite-dimensional Ornstein-Uhlenbeck operator, see Nourdin and Peccati [87].

1.5. Some recent applications

In recent years the method of cumulants has attracted interest in various areas of application. We list a few; the examples below involve bounds on cumulants though not necessarily of the form $|\kappa_j| \leq j!^{1+\gamma}/\Delta^{j-2}$.

In stochastic geometry, the method of cumulants is used in numerous ways. In [48] and [49] it is used for the volume and the number of faces of a Gaussian polytope to show concentration inequalities and a Marcinkiewicz-Zygmund-type strong law of large numbers as well as a central limit theorem including error bounds and moderate deviations. Furthermore for the volumes of random simplices a central limit theorem, mod-phi convergence as well as moderate and large deviations are proved in [47], where the dimension and the number of points grow to infinity. Poisson cylinder processes are studied in [59], [60] and volumes of simplices in high-dimensional Poisson-Delaunay tessellations in [51]. The method is also applicable to the covered volume in the Boolean model, [46, 56, 57]. General functionals of random m-dependent fields are studied in [46, 53, 55, 58]. A survey covering m-dependent variables, Markov chains, and mixing random variables is given by Heinrich [54]. Moderate deviation results for classical stabilizing functionals in stochastic geometry can be found in [32].

The method of cumulants also plays an important role in the theory of random matrices and determinantal point processes. For the latter cumulants for the sine kernel were studied by Costin and Lebowitz [18]. In [106] and [107] Soshnikov studied the Gaussian limit for linear statistics of Gaussian fields and determinantal random point fields. The methods of [18] were also applied to spacing distribution in the circular unitary ensemble [105]. The method of [106] was further extended e.g. to study general one-cut unitary-invariant matrix models [72] and to prove mesoscopic fluctuations for unitary invariant ensembles [73]. Further the methods were applied to (generalized) Ginibre ensembles, where linear eigenvalue statistics and characteristic polynomials were studied [95] and moderate deviations for counting statistics were shown [36]. The method of cumulants is also applied for eigenvalue counting statistics and determinants of Wigner matrices [26, 27]. This was generalized to the linear spectrum statistics of orthogonal polynomial ensembles in [88]. Studying the cumulants also works for characteristic polynomials of random matrices from the circular unitary ensemble [17] as well as the determinants of random block Hankel matrices, [24]. Cumulants also enter the analysis of matrix models in which the eigenvalues do not form a determinantal point process, see Borot and Guionnet [14].

The method of cumulants, combined with the concept of dependency graphs, yields useful results in random graphs, in particular subgraph count statistics in Erdős-Rényi random graphs [28, 39], or more generally for graphons in [41], for the profile of a branching random walk [50] as well as for the winding number of Brownian motion in the complex plane [21]. Féray gives criteria for normality convergence of dependency graphs by using the method of cumulants [37]. This is applied to vincular permutation patterns in [61] (yielding the same order of convergence in the Kolmogorov distance than achieved by applying Stein's method).

The generalization to weighted dependency graphs allows for applications to the Ising model as [29] shows. Already in [80] models of statistical mechanics such as the Curie-Weiss and the Ising model are studied.

For an application for multiple Wiener-Itô integrals in stochastic analysis see [87], the survey of a series of articles in the book [89] as well as [102].

Another area of applications are random logarithmic structures in combinatorics, random permutations and random integer partitions as well as so-called character values, see [4, 39, 41]. Random arithmetic functions, the Riemannian ζ function and L functions over finite fields are considered in the series of publications [21, 39, 66, 70, 71]. The latest publication [38] shows a central limit theorem for weighted dependency graphs and generalizes results for the number of occurrences of any fixed pattern in multiset permutations and in set partitions.

In mathematical biology Möhle and Pitters derived the absorption time and tree length of the Kingman coalescent by bounding the cumulants, see [84]. Restricted to the infinitely many sites model of Kimura, in [93] Pitters derives a formula for the cumulants of the number of segregating sites of a Kingman coalescent implying a central limit theorem and a law of large numbers. Studying the multivariate cumulants, the number of cycles in a random permutation and the number of segregating sites jointly converge to the Brownian sheet [94].

To conclude this non-exhaustive list, we stress that cumulants are highly relevant in areas somewhat outside the scope of this survey. Perhaps closest to traditional probability are asymptotic techniques in statistics, see Barndorff-Nielsen and Cox [7]. A tensorial view of cumulants, with applications in statistics, is given by McCullagh [79]. Cumulants are also useful in algebraic statistics, see the book by Zwiernik [114]. For example, Sturmfels and Zwiernik [109] discuss cumulants for algebraic varieties and binary random variables on hidden subset models. A completely different line of inquiry is free probability and non-commutative probability, in which different notions of independence come with different notions of cumulants. The relation between free, monotone and Boolean cumulants is studied by Arizmendi, Hasebe, Lehner and Vargas [3]. Finally, cumulants also feature prominently in kinetic theory and the analysis of time-dependent models in mathematical physics [77, 13].

2. The main lemmas

Here we state four theorems, roughly the "main lemmas" in Saulis and Stat-ulevičius [101, Chapter 2]. Two theorems are about the normal approximation, with Cramér corrections, to a random variable under conditions on finitely many cumulants (Theorem 2.2) or under the growth condition $(\mathcal{S}_\gamma)$, which allows for heavy-tailed random variables (Theorem 2.3). An inequality of the Berry-Esseen type and a concentration inequality are given in Theorems 2.4 and 2.5, again under the condition $(\mathcal{S}_\gamma)$.

The main theorems are illustrated with two elementary examples—Gamma and Weibull distributions—in Section 2.3. The meaning of the Statulevičius condition $(\mathcal{S}_\gamma)$ is further clarified in Section 2.4.

2.1. Normal approximation with Cramér corrections

In the following X is a real-valued random variable, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with cumulants $\kappa_j(X) = \kappa_j$. We assume throughout the text that the variable X is normalized, i.e., $\mathbb{E}[X] = 0$ and $\mathbb{V}(X) = 1$. Two important quantities are the cumulant generating function

$$\varphi(t) := \log \mathbb{E}\big[\exp(tX)\big] \in \mathbb{R} \cup \{\infty\} \qquad (t \in \mathbb{R}) \tag{2.1}$$

and its Legendre transform

$$I(x) := \sup_{t \in \mathbb{R}}\big(tx - \varphi(t)\big) \qquad (x \in \mathbb{R}). \tag{2.2}$$

The first theorem works under the condition that there exists some $\Delta > 0$ such that

$$\forall j \geq 3: \quad |\kappa_j(X)| \leq \frac{(j-2)!}{\Delta^{j-2}}. \tag{$\mathcal{S}$}$$

Under this condition the cumulant generating function is finite on $(-\Delta, \Delta)$ with absolutely convergent Taylor expansion

$$\varphi(t) = \frac{t^2}{2} + \sum_{j=3}^{\infty} \frac{\kappa_j}{j!} t^j \quad (|t| < \Delta), \tag{2.3}$$

where we have used $\kappa_1 = \mathbb{E}[X] = 0$ and $\kappa_2 = \mathbb{V}(X) = 1$. The Cramér rate function admits a Taylor expansion as well (Proposition A.1), with radius of convergence at least $0.3\,\Delta$ (Proposition A.2). The expansion is of the form

$$I(x) = \frac{x^2}{2} - \sum_{j=3}^{\infty} \lambda_j x^j \qquad (|x| < \frac{3}{10}\,\Delta). \tag{2.4}$$

The series $\sum_{j=3}^{\infty} \lambda_j\, x^{j-3}$ is called *Cramér series* or *Cramér-Petrov series* af-ter [19, 91]. Cramér's original article [19] was recently made accessible in elec-tronic form, together with an English translation, by Touchette [20]. Appendix A

collects some relevant background on the series. It is convenient to set

$$L(x) := \sum_{j=3}^{\infty} \lambda_j x^j \tag{2.5}$$

so that $I(x) = \frac{1}{2}x^2 - L(x)$. From now on Z is always a standard normal variable, $Z \sim \mathcal{N}(0, 1)$.

Theorem 2.1. *Under condition* $(\mathcal{S})$ *there exist universal constants* $c, C, C' > 0$ *such that for all* $x \in [0, c\Delta]$ *and some* $\theta = \theta(x) \in [-1, 1]$,

$$\mathbb{P}(X \geq x) = \mathrm{e}^{L(x)} \mathbb{P}(Z \geq x)\Big(1 + C\theta \, \frac{x+1}{\Delta}\Big)$$

and $|L(x)| \leq C'x^3/\Delta$.

The theorem is proven in Section 6. It is a special case of Theorem 2.3 below; we have chosen to provide a separate statement as it is easier to grasp, and its proof is a helpful warm-up for the proof of Theorem 2.3.

The next theorem asks what subsists when the cumulants satisfy the bound in $(\mathcal{S})$ only up to some order, i.e.,

$$\forall j \in \{3, \ldots, s+2\}: \quad |\kappa_j(X)| \leq \frac{(j-2)!}{\Delta^{j-2}} \tag{$\mathcal{S}^*$}$$

for some $s \in \mathbb{N}$. We say that X satisfies condition $(\mathcal{S}^*)$ if all moments $\mathbb{E}[X^j]$, $j \leq s+2$ exist—hence also all cumulants κ_j with $j \leq s+2$—and the cumulants satisfy the required inequality. Under condition $(\mathcal{S}^*)$ the random variable X need not have exponential moments and the cumulant generating function may be infinite, therefore the definitions of $\varphi(t)$ and $I(x)$ are modified as follows. We set

$$\tilde{\varphi}(t) = \frac{t^2}{2} + \sum_{j=3}^{s} \frac{\kappa_j}{j!} t^j. \tag{2.6}$$

For small x and t the equation $\tilde{\varphi}'(t) = x$ reads $t + O(t^2) = x$ and it has a solution $t(x) = x + \sum_{j=2}^{\infty} \tilde{b}_j x^j$ with suitably defined coefficients $\tilde{b}_j$. We define

$$\tilde{I}(x) := t(x)x - \tilde{\varphi}(t(x)), \quad \tilde{L}(x) := \frac{x^2}{2} - \tilde{I}(x) \tag{2.7}$$

and note that $\tilde{L}(x)$ has a Taylor expansion $\tilde{L}(x) = \sum_{j=3}^{\infty} \tilde{\lambda}_j x^j$ with radius of convergence at least $0.3\,\Delta$ (Propositions A.4, A.5). In addition, $\tilde{\lambda}_j = \lambda_j$ for $j \leq s$ (Eq. (A.8)).

Theorem 2.2. *Let* X *be a real-valued random variable with* $\mathbb{E}[X] = 0$ *and* $\mathbb{V}(X) = 1$. *Assume that* X *satisfies condition* $(\mathcal{S}^*)$ *for some even* $s \geq 2$ *and* $\Delta > 0$ *with* $s \leq 2\Delta^2$. *Then for all* $x \in [0, \sqrt{s}/(3\sqrt{\mathrm{e}}))$ *and some* $\theta = \theta(x) \in [-1, 1]$,

$$\mathbb{P}(X \geq x) = \mathrm{e}^{\tilde{L}(x)} \mathbb{P}(Z \geq x)\Big(1 + \theta f(\delta, s) \, \frac{x+1}{\sqrt{s}}\Big)$$

with

$$\delta = \frac{x}{\sqrt{s}/(3\sqrt{\mathrm{e}})} \in [0,1), \quad f(\delta,s) = \frac{1}{1-\delta}\Big(127 + 113\,s\,\mathrm{e}^{-(1-\delta)s^{1/4}/2}\Big)$$

and $|\tilde{L}(x)| \le 1.2\,x^3/\Delta$.

The theorem is proven in Section 7. It corresponds to Lemma 2.2 in [101] and is due to Rudzkis, Saulis, and Statulevičius [99]. The constants are slightly worse than the constants given in [101] but of a similar order of magnitude. We are not aware of any application of the concrete formula for $f(\delta,s)$. Instead, what matters is that $f(\delta,s)$ is bounded on $[0,\delta_0] \times \mathbb{N}$, for all $\delta_0 < 1$.

The next theorem works under a condition that allows for heavy-tailed variables. For $\gamma > 0$ and $\Delta > 0$, consider the *Statulevičius condition*

$$\forall j \ge 3: \quad |\kappa_j(X)| \le \frac{j!^{1+\gamma}}{\Delta^{j-2}}. \tag{$\mathcal{S}_\gamma$}$$

The relation of this condition with Weibull tails and *Linnik's condition*, which reads $\mathbb{E}[\exp(\delta|X|^{1/(1+\gamma)})] < \infty$, is clarified by Lemma 2.8 below, see [1]. Define

$$\Delta_\gamma := \frac{1}{6}\Big(\frac{\Delta}{\sqrt{18}}\Big)^{1/(1+2\gamma)}, \qquad s_\gamma := \quad 2\Big\lfloor \frac{1}{2}\Big(\frac{\Delta^2}{18}\Big)^{1/(1+2\gamma)}\Big\rfloor - 2, \tag{2.8}$$

$$m_\gamma := \min\Big(\Big\lceil \frac{1}{\gamma}\Big\rceil + 1, s_\gamma\Big). \tag{2.9}$$

Theorem 2.3. *Let X be a real-valued random variable with $\mathbb{E}[X] = 0$ and $\mathbb{V}(X) = 1$. Suppose that X satisfies condition ($\mathcal{S}_\gamma$). Then there exist some universal constant $C > 0$ such that for all $x \in [0, \Delta_\gamma)$ and some $\theta = \theta(x) \in [-1, 1]$,*

$$\mathbb{P}(X \ge x) = \mathrm{e}^{\tilde{L}_\gamma(x)}\mathbb{P}(Z \ge x)\Big(1 + \theta g(\delta, \Delta_\gamma)\frac{x+1}{\Delta_\gamma}\Big)$$

with $|\tilde{L}_\gamma(x)| \le x^3/(1.54\,\Delta_\gamma)$ *and*

$$\tilde{L}_\gamma(x) = \begin{cases} \theta(\frac{x}{\Delta_\gamma})^3, & \gamma \ge 1, \\ \sum_{j=3}^{m_\gamma} \lambda_j x^j + \theta C_\gamma(\frac{x}{\Delta_\gamma})^{m_\gamma+1}, & \gamma < 1. \end{cases}$$

Here we set

$$g(\delta, \Delta_\gamma) := \frac{1}{1-\delta}\Big(24 + 749\Delta_\gamma^2 \exp\big(-(1-\delta)\sqrt{\Delta_\gamma}\big)\Big),$$

and $\delta = x/\Delta_\gamma$.

The theorem is proven in Section 8. It corresponds to Lemma 2.3 in [101] and is due to Rudzkis, Saulis, and Statulevičius [99]. The constants given in [101] are 60 and 600 instead of 24 and 749. Our second constant 749 is worse but our first constant 24 is better.

We shall see that under the conditions of the theorem, s_γ is larger or equal to 4 so that $m_\gamma = 2$ for $\gamma \geq 1$ and $m_\gamma \geq 3$ for $\gamma < 1$. Let us briefly comment on the two bounds for $\tilde{L}_\gamma(x)$ in Theorem 2.3. The global bound $|\tilde{L}_\gamma(x)| \leq x^3/(1.54\,\Delta_\gamma)$ is similar to the bounds for $L(x)$ and $\tilde{L}(x)$ in Theorems 2.1 and 2.2. It gives the leading order of $\tilde{L}_\gamma(x)$. Note, however, that it can be quite large, since x^3/Δ_γ can be of order up to Δ_γ^2. The case distinction in Theorem 2.3 provides a representation of $\tilde{L}_\gamma(x)$ that is precise in the sense that the remainder is small when x is small compared to Δ_γ.

Remark. Correction terms from the Cramér series need only be taken into account when $\gamma < 1$. This corresponds to variables with tails decaying like $\exp(-c\,x^\alpha)$ with $\alpha = 1/(1+\gamma) > 1/2$, see Section 2.3.2 and Lemma 2.8 below. It should be pointed out that the value $\alpha = 1/2$ plays a role as well for large deviations of i.i.d. heavy-tailed variables with Weibull tails, see Mikosch and Nagaev [83, Proposition 3.1].

The error term in $O((x+1)/\Delta_\gamma)$ after the exponential is known to be not optimal for sums of i.i.d. random variables. If $X_n = S_n/\sqrt{n}$ is a normalized sum of i.i.d. random variables that satisfy the Statulevičius condition for some fixed Δ, then X_n satisfies the Statulevičius condition with a n-dependent $\Delta(n)$ proportional to $\sqrt{n}$ (see Section 3.1), and $1/\Delta_\gamma$ is of the order of $(x+1)/\sqrt{n}^\beta$ for some $\beta < 1$, which is larger than the error term $O(x/\sqrt{n})$ proven e.g. in Ibragimov and Linnik [64, Eq. (13.4.4)].

The principal idea in the proof of Theorem 2.3 is to apply Theorem 2.2 for suitably chosen s and Δ_s such that condition $(\mathcal{S}^*)$ is satisfied if condition $(\mathcal{S}_\gamma)$ holds true. Thus, we seek s and Δ_s such that

$$\frac{j!^{1+\gamma}}{\Delta^{j-2}} \leq \frac{(j-2)!}{\Delta_s^{j-2}} \qquad (j = 3, \ldots, s+2).$$

The inequality is equivalent to

$$\big(j(j-1)\big)^{1+\gamma}(j-2)!^\gamma \leq \Big(\frac{\Delta}{\Delta_s}\Big)^{j-2} \qquad (j = 3, \ldots, s+2)$$

and

$$\max_{k=1,\ldots,s} \Big(\frac{\gamma}{k}\log k! + \frac{1+\gamma}{k}\big(\log(k+2) + \log(k+1)\big)\Big) \leq \log\frac{\Delta}{\Delta_s}. \qquad (2.10)$$

By Stirling's formula, the term to be maximized behaves like

$$\gamma \log\frac{k}{e} + O\Big(\frac{1}{k}\log k\Big) = \gamma\big(1 + o(1)\big)\log k \qquad (k \to \infty).$$

For a heuristic evaluation of (2.10), let us keep the leading order term only, then (2.10) becomes $\gamma \log s \leq \log(\Delta/\Delta_s)$ hence $\Delta_s = \Delta/s^\gamma$. Then $s \leq 2\Delta_s^2$ if and only if $s \leq 2(\Delta^2)^{1/(1+2\gamma)}$. This suggests to pick $\sqrt{s} = \text{const}\,\Delta^{1/(1+2\gamma)}$, which is precisely the power of Δ appearing in Theorem 2.3, via Δ_γ.

2.2. Berry-Esseen bound and concentration inequality

Theorem 2.3 is complemented by a Berry-Esseen bound and a concentration inequality that provides statements for all of $x \geq 0$—no need to restrict to $x \in (0, \Delta_\gamma)$.

Theorem 2.4. *Under the Statulevičius condition* $(\mathcal{S}_\gamma)$, *we have*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(X \geq x) - \mathbb{P}(Z \geq x) \right| \leq \frac{C_\gamma}{\Delta^{1/(1+2\gamma)}}$$

for some constant C_γ that does not depend on the random variable X or on Δ.

Theorem 2.4 is proven in Section 9, it corresponds to Corollary 2.1 in [101]. The precise bound given by [101] is $18/\Delta_\gamma$ with Δ_γ defined in (2.8), we have not checked the numerical constants.

Theorem 2.5. *Suppose $\mathbb{E}[X] = 0$ and*

$$|\kappa_j| \leq \left(\frac{j!}{2}\right)^{1+\gamma} \frac{H}{\overline{\Delta}^{j-2}} \tag{2.11}$$

for some $\gamma \geq 0$ and $H, \overline{\Delta} > 0$. Set $\alpha := 1/(1+\gamma)$. Then there exists $C > 0$ such that for all $x \geq 0$,

$$\mathbb{P}(X \geq x) \leq C \exp\left(-\frac{1}{2} \frac{x^2}{H + x^{2-\alpha}/\bar{\Delta}^\alpha}\right). \tag{2.12}$$

The constant does not depend on X, H, $\overline{\Delta}$, or γ.

Theorem 2.5 is proven in Section 5, it corresponds to Lemma 2.4 in [101] and is due to Bentkus and Rudzkis [9]. As noted by Kallabis and Neumann [68], the statement of Lemma 2.4 in [101] contains a typo. Bentkus and Rudzkis give the constant $C = 1$, we give a shorter proof but provide no concrete numerical bound on C.

The exponent in (2.12) can be expressed in terms of the harmonic mean $\text{Harm}(a, b) = 2(a^{-1} + b^{-1})^{-1}$ as

$$\frac{x^2}{H + x^{2-\alpha}/\bar{\Delta}^\alpha} = \left(\frac{H}{x^2} + \frac{1}{(x\bar{\Delta})^\alpha}\right)^{-1} = \frac{1}{2}\text{Harm}\left(\frac{x^2}{H}, (x\bar{\Delta})^\alpha\right).$$

Thus Theorem 2.5 provides an upper bound that smoothly interpolates between Gaussian tails $\exp(-x^2/(2H))$ and stretched exponential tails $\exp(-(x\bar{\Delta})^\alpha/2)$.

2.3. Two examples

Before we turn to the proofs, we provide two examples that illustrate the theorems and explain the role of Δ and $\Delta^{1/(1+2\gamma)}$.

2.3.1. A Gamma-distributed random variable

Pick $\Delta > 0$ and let Y be a Gamma random variable with parameters $\beta = \Delta$ and $\alpha = \Delta^2$. Thus the random variable Y has probability density function $\Gamma(\alpha)^{-1}\mathbb{1}_{(0,\infty)}(x)\beta^\alpha x^{\alpha-1}e^{-\beta x}$, moment generating function $(1-t/\beta)^{-\alpha}$, variance $\alpha/\beta^2 = 1$ and expected value $\alpha/\beta = 1$. Set $X := Y - \Delta$. Then X has pdf

$$\rho_\Delta(x) = \frac{\Delta^{\Delta^2}}{\Gamma(\Delta^2)}(x+\Delta)^{\Delta^2-1}e^{-\Delta(x+\Delta)}\mathbb{1}_{[-\Delta,\infty)}(x) \qquad (2.13)$$

and cumulant generating function, for $|t| < \Delta$, given by

$$\varphi(t) = \log \mathbb{E}\big[e^{tX}\big] = -\Delta^2 \log\Big(1 - \tfrac{t}{\Delta}\Big) - t\Delta = \sum_{j=2}^\infty \frac{1}{j}\frac{t^j}{\Delta^{j-2}},$$

from which we read off the cumulants

$$\kappa_j = \frac{(j-1)!}{\Delta^{j-2}} \qquad (j \geq 2).$$

The explicit formula for the probability density function allows us to check that the normal approximation for X is good when Δ is large and x is small compared to Δ.

Proposition 2.6. *As $\Delta \to \infty$ and $x/\Delta \to 0$, the probability density function* (2.13) *satisfies*

$$\rho_\Delta(x) = \frac{\exp(-\tfrac{1}{2}x^2[1 + O(\tfrac{x}{\Delta})])}{\sqrt{2\pi}}\Big(1 + O\big(\tfrac{x+1}{\Delta}\big)\Big).$$

Proof. We rewrite

$$\rho_\Delta(x) = \frac{\Delta^{2\Delta^2-1}}{\Gamma(\Delta^2)}(1 + x/\Delta)^{\Delta^2-1}e^{-\Delta^2(1+x/\Delta)}\mathbb{1}_{[-1,\infty)}(x/\Delta)$$

$$= \frac{\Delta^{2\Delta^2-1}e^{-\Delta^2}}{\Gamma(\Delta^2)}\frac{1}{1+x/\Delta}e^{\Delta^2[\log(1+x/\Delta)-x/\Delta]}\mathbb{1}_{[-1,\infty)}(x/\Delta). \qquad (2.14)$$

Using $\Gamma(x+1) = x\Gamma(x)$ and Stirling's approximation, we have

$$\Gamma(\Delta^2) = \Delta^{-2}\Gamma(\Delta^2+1) = \Big(1 + O\big(\tfrac{1}{\Delta}\big)\Big)\Delta^{-2}\sqrt{2\pi\Delta^2}\Big(\tfrac{\Delta^2}{e}\Big)^{\Delta^2}$$

$$= \Big(1 + O\big(\tfrac{1}{\Delta}\big)\Big)\sqrt{2\pi}\Delta^{2\Delta^2-1}e^{-\Delta^2}. \qquad (2.15)$$

Combining this with the Taylor expansion $\log(1 + u) = u - u^2/2 + O(u^3)$ of the logarithm, we deduce

$$\rho_\Delta(x) = \Big(1 + O\big(\tfrac{1}{\Delta}\big)\Big)\Big(1 + O\big(\tfrac{x}{\Delta}\big)\Big)\frac{1}{\sqrt{2\pi}}\exp\Big(-\tfrac{1}{2}x^2\big(1 + O\big(\tfrac{x}{\Delta}\big)\big)\Big). \qquad \square$$

The tilted normal approximation (compare Bahadur-Rao [5], [22, Theorem 3.7.4] or the proof of the lower bound in Cramér large deviation principle [22, Chapter 2.2]) consists in the following. Let $I(x) := \sup_{t \in \mathbb{R}}(tx - \varphi(t))$ be the rate function in the Cramér large deviation principle. An explicit computation yields

$$I(x) = \Delta x - \Delta^2 \log(1 + x/\Delta) = \frac{1}{2}x^2 + \sum_{j=3}^{\infty} \frac{(-1)^j}{j\Delta^{j-2}} x^j,$$

from which we read off the Cramér-Petrov series

$$L(x) = -\sum_{j=3}^{\infty} \frac{(-1)^j}{j\Delta^{j-2}} x^j \qquad (|x| < \Delta).$$

For later purposes we extend the definition of $L(x)$ to all of $\mathbb{R}$ by putting $L(x) = I(x) - x^2/2$. Notice that as $x/\Delta \to 0$,

$$L(x) = x^2\, O(x/\Delta).$$

Given $x \geq 0$, let $h \geq 0$ be the solution of $\varphi'(h) = x$ and let $\widehat{X}_h$ be a random variable with distribution $\mathbb{P}(\widehat{X}_h \leq y) = e^{-\varphi(h)}\mathbb{E}[e^{hX}\mathbb{1}_{\{X \leq y\}}]$. The variable $\widehat{X}_h$ has expected value $\varphi'(h) = x$, variance $\varphi''(h)$, and probability density function $e^{-\varphi(h)}e^{hy}\rho_\Delta(y)$. The approximation $\mathscr{L}(\widehat{X}_h) \approx \mathcal{N}(x, \varphi''(h))$ suggests

$$e^{-\varphi(h)}e^{hy}\rho_\Delta(y) = \rho_{\widehat{X}_h}(y) \approx \frac{1}{\sqrt{2\pi\varphi''(h)}} e^{-(y-x)^2/[2\varphi''(h)]}.$$

Remember $I(x) = hx - \varphi(h)$ and $I''(x) = 1/\varphi''(h)$, so

$$\rho_\Delta(y) \approx \sqrt{I''(x)}e^{-I(x)} \times \frac{1}{\sqrt{2\pi}}e^{-(y-x)^2/[2\varphi''(h)]-h(y-x)}.$$

Proposition 2.7. *As $\Delta \to \infty$, we have for all $x \geq -\Delta$*

$$\rho_\Delta(x) = \left(1 + O\left(\frac{1}{\Delta}\right)\right)\sqrt{\frac{I''(x)}{2\pi}}\, e^{-I(x)} = \left(1 + O\left(\frac{1}{\Delta}\right)\right)\frac{\exp(L(x))}{1 + x/\Delta}\frac{\exp(-x^2/2)}{\sqrt{2\pi}}$$

with an error term $O(1/\Delta)$ uniform in x.

Notice that this approximation is much more precise than the direct normal approximation.

Proof. An explicit computation yields

$$I'(x) = \frac{x}{1 + x/\Delta}, \quad I''(x) = \frac{1}{(1 + x/\Delta)^2}.$$

Hence for all $x \geq -\Delta$

$$\frac{1}{\sqrt{2\pi}}\sqrt{I''(x)}e^{-I(x)} = \frac{1}{\sqrt{2\pi}}\frac{1}{(1 + x/\Delta)}e^{-\Delta x}(1 + x/\Delta)^{\Delta^2}$$

$$= \frac{1}{\sqrt{2\pi}}(1 + x/\Delta)^{\Delta^2-1} e^{-\Delta x}.$$

A quick look at (2.14) reveals that this expression differs from the density $\rho_\Delta(x)$ only through the prefactor, $1/\sqrt{2\pi}$ vs. $\Delta^{2\Delta^2-1} e^{-\Delta^2}/\Gamma(\Delta^2)$. The ratio between these two prefactors is independent of x and behaves like $1 + O(1/\Delta)$ by Stirling's approximation (2.15). □

2.3.2. Weibull tails

Fix $\gamma > 0$. Set $\alpha := 1/(1+\gamma)$ and consider a non-negative random variable with survival function

$$\mathbb{P}(Y \geq y) = \exp(-y^\alpha) \quad (y \geq 0).$$

The moments of Y are given by

$$\mathbb{E}[Y^m] = \Gamma\left(1 + \frac{m}{\alpha}\right).$$

Notice that as $m \to \infty$,

$$\mathbb{E}[Y^m] = (1 + o(1))\sqrt{2\pi m/\alpha}\left(\frac{m}{\alpha e}\right)^{m/\alpha} = \frac{m!^{1+\gamma}}{\alpha^{m/\alpha}} e^{o(m)}, \qquad (2.16)$$

where we have used Stirling's approximation and $1/\alpha = 1+\gamma$. Let $Z \sim \mathcal{N}(0,1)$, independent of Y. For small $\varepsilon > 0$, set

$$X_\varepsilon := Z + \varepsilon Y.$$

Then the expected value $\mu_\varepsilon = \mathbb{E}[X_\varepsilon]$ and the variance $\sigma_\varepsilon^2 = \mathbb{V}(X_\varepsilon)$ satisfy $\mu_\varepsilon = O(\varepsilon)$ and $\mathbb{V}[X_\varepsilon] = 1 + O(\varepsilon^2)$. The centered variable $\widehat{X}_\varepsilon = (X_\varepsilon - \mu_\varepsilon)/\sigma_\varepsilon$ has cumulants

$$\kappa_j(\widehat{X}_\varepsilon) = \left(\frac{\varepsilon}{\sigma_\varepsilon}\right)^j \kappa_j(Y) \quad (j \geq 3).$$

In view of (2.16) and Lemma 2.8, it seems plausible that the centered variable $\widehat{X}_\varepsilon$ satisfies condition $(\mathcal{S}_\gamma)$ with ε-dependent $\Delta = \Delta(\varepsilon)$ of the order of

$$\Delta(\varepsilon) \approx \frac{\sigma_\varepsilon}{\varepsilon} = \frac{1}{\varepsilon}(1 + O(\varepsilon^2)).$$

We would like to understand the behavior of the tails $\mathbb{P}(X_\varepsilon \geq x_\varepsilon) = \mathbb{P}(\widehat{X}_\varepsilon \geq (x_\varepsilon - \mu_\varepsilon)/\sigma_\varepsilon)$ as $x_\varepsilon \to \infty$ and $\varepsilon \to 0$. Theorem 2.3 suggests that the normal approximation $\mathbb{P}(X_\varepsilon \geq x_\varepsilon) \approx \mathbb{P}(Z \geq x_\varepsilon)$ should be good as long as x_ε is small compared to

$$\Delta(\varepsilon)^{1/(1+2\gamma)} \approx \varepsilon^{-1/(1+2\gamma)}. \qquad (2.17)$$

We are not going to provide a precise statement on the normal approximation. Instead we would like to emphasize two key facts. First, the critical scale (2.17) is explained with a very simple heuristics. Second, the normal approximation

cannot apply beyond that scale. As a consequence, the scale $\Delta^{1/(1+2\gamma)}$ in Theorem 2.3 is not due to technical restrictions but is in fact sharp.

For the heuristic derivation of the critical scale (2.17), notice

$$\mathbb{P}(X_\varepsilon \geq x_\varepsilon) \geq \mathbb{P}(Z \geq 0,\, \varepsilon Y \geq x_\varepsilon) = \frac{1}{2}\,\exp\Big(-\big(\frac{x_\varepsilon}{\varepsilon}\big)^\alpha\Big) \qquad (2.18)$$

but also, since Y is non-negative,

$$\mathbb{P}(X_\varepsilon \geq x_\varepsilon) \geq \mathbb{P}(Z \geq x_\varepsilon) = (1 + o(1))\frac{\exp(-x_\varepsilon^2/2)}{x_\varepsilon\sqrt{2\pi}},$$

where we have used the well-known asymptotic behavior of Gaussian tails, see Eq. (4.16) below. The two lower bounds correspond to two different ways of realizing the unlikely event that $X_\varepsilon = Z + \varepsilon Y \geq x_\varepsilon$—either Z stays small but εY is very large, or εY stays small but Z is large. Which of the two effects dominates the other? The answer depends on how large x_ε is. In view of the equivalence

$$\exp\Big(-\frac{x_\varepsilon^2}{2}\Big) \geq \exp\Big(-\big(\frac{x_\varepsilon}{\varepsilon}\big)^\alpha\Big) \;\Leftrightarrow\; x_\varepsilon \leq \Big(\frac{2^{1/\alpha}}{\varepsilon}\Big)^{\alpha/(2-\alpha)}$$

we should expect the probability $\mathbb{P}(X_\varepsilon \geq x_\varepsilon)$ to be similar to $\exp(-x_\varepsilon^2/2)$ when $x_\varepsilon \ll \varepsilon^{-\alpha/(2-\alpha)}$ and similar to $\exp(-(x_\varepsilon/\varepsilon)^\alpha)$ when $x_\varepsilon \gg \varepsilon^{-\alpha/(2-\alpha)}$. Because of $\alpha = 1/(1+\gamma)$, we have

$$\varepsilon^{-\alpha/(2-\alpha)} = \varepsilon^{-1/(1+2\gamma)},$$

which is exactly the right-hand side of (2.17). Thus we have recovered, heuristically, the critical scale from Theorem 2.3.

In addition, for $x_\varepsilon \gg \varepsilon^{-\alpha/(2-\alpha)}$, the lower bound (2.18) yields the rigorous asymptotic lower bound

$$\mathbb{P}(X_\varepsilon \geq x_\varepsilon) \geq \frac{1}{2}\exp\Big(-\big(\frac{x_\varepsilon}{\varepsilon}\big)^\alpha\Big) \gg \mathbb{P}(Z \geq x_\varepsilon).$$

Hence, the normal approximation cannot be good beyond the scale (2.17).

2.4. On the Cramér, Linnik, and Statulevičius conditions

For $\gamma > 0$, the Statulevičius condition $(\mathcal{S}_\gamma)$ seems technical and not immediately accessible to probabilistic intuition. For $\gamma = 0$, the situation is simpler: If $|\kappa_j| \leq j!/\Delta^{j-2}$ for some $\Delta > 0$ and all $j \geq 3$, then $\sum_{j\geq 1} \kappa_j t^j/j!$ is absolutely convergent on $(-\Delta, \Delta)$ and $\mathbb{E}[\exp(t|X|)] < \infty$ for all $t \in (-\Delta, \Delta)$. Thus Cramér's condition is satisfied and the distribution of X has exponentially decaying tails.

The question arises if there is a similar intuition for the condition $(\mathcal{S}_\gamma)$ when $\gamma > 0$. The answer is yes, if we replace Cramér's condition by *Linnik's condition*, which reads

$$\mathbb{E}\big[\exp(\delta|X|^\alpha)\big] < \infty$$

for some $\alpha \in (0, 1)$ and $\delta > 0$. The correct choice turns out to be $\alpha = 1/(1+\gamma)$, which should not surprise us after Section 2.3.2. In addition, conditions on cumulants may be replaced by conditions on moments.

Lemma 2.8. *Let X be a random variable with $\mathbb{E}[X] = 0$, $\mathbb{V}(X) = 1$, and $\mathbb{E}[|X|^j] < \infty$ for all $j \geq 3$. Fix $\gamma \geq 0$. Then, the following three statements are equivalent:*

(i) X satisfies condition $(\mathcal{S}_\gamma)$.
(ii) There exists $H \geq 1$ such that the moments of X satisfy

$$\left|\mathbb{E}[X^j]\right| \leq j!^{1+\gamma} H^{j-2} \quad (j \geq 3). \qquad (\mathcal{M}_\gamma)$$

(iii) There exists $\delta > 0$ such that $\mathbb{E}\left[\exp\left(\delta|X|^{1/(1+\gamma)}\right)\right] < \infty$.

Similar relations, with explicit control on constants, are proven in [101, Chapter 3.1]. Condition (ii), for $\gamma = 0$, is a variant of the condition $|\mathbb{E}[X^j]| \leq \frac{1}{2}Cj!H^{j-2}$, sometimes called *Bernstein condition* [101] because it allows for a Bernstein inequality with unbounded random variables [64, Chapter 7.5]. Linnik's condition is discussed in depth in the context of "monomial zones of local normal attraction" in [64, Chapter 9]. The name Linnik condition is not used in the book [64], it is used for example by Saulis and Statulevičius [101] or Amosova [2].

Large deviation theorems for sums of i.i.d. variables under conditions of the type $\mathbb{E}\exp[h(X)] < \infty$ are available as well, see Chapter 11 in Ibragimov and Linnik [64] and Nagaev [86] for sums of i.i.d. variables and Heinrich [54, Section 4] for sums of Markov chains. However to the best of our knowledge there is no analogue of Lemma 2.8 for such more general conditions.

Proof. We start with the equivalence of (ii) and (iii).

"$(ii) \Rightarrow (iii)$" First we note that condition $(\mathcal{M}_\gamma)$ implies a similar condition for the moments of $|X|$. For even powers this is immediate. For odd powers, we use $2j + 1 = j + (j + 1)$, the Cauchy-Schwarz inequality, and condition $(\mathcal{M}_\gamma)$. This gives

$$\mathbb{E}\left[|X|^{2j+1}\right] \leq \left((2j)!^{1+\gamma}H^{2j-2}(2j+2)!^{1+\gamma}H^{2j}\right)^{1/2}$$
$$= \left(\frac{2j+2}{2j+1}\right)^{(1+\gamma)/2}(2j+1)!^{1+\gamma}H^{2j-1}.$$

The ratio $(2j + 2)/(2j + 1)$ is not smaller than 1 but is bounded by $4/3$ (for $j \geq 1$). Therefore, choosing $H' \geq H$ large enough, we get

$$\mathbb{E}\left[|X|^{2j+1}\right] \leq (2j+1)!^{1+\gamma}(H')^{2j+1-2}.$$

We conclude with an argument by Mason and Zhou [78, Appendix B]. Set $\alpha := 1/(1+\gamma)$. The function $x \mapsto x^\alpha$ is concave on $\mathbb{R}_+$, therefore $\mathbb{E}[Y^\alpha] \leq \mathbb{E}[Y]^\alpha$ for every non-negative random variable Y. In particular, for all $j \geq 3$,

$$\mathbb{E}\left[|X|^{j\alpha}\right] \leq \left(\mathbb{E}\left[|X|^j\right]\right)^\alpha \leq j!^{(1+\gamma)\alpha}(H')^{(j-2)\alpha} = j!(H')^{(j-2)\alpha},$$

which gives

$$\mathbb{E}\Big[\exp(\delta|X|^{\alpha})\Big] = 1 + \sum_{j=1}^{\infty} \frac{\delta^j}{j!} \mathbb{E}\big[|X|^{j\alpha}\big] < \infty$$

for $\delta < 1/H'$.

The implication "$(iii) \Rightarrow (ii)$" is proven by Amosova [2, Lemma 3], see also Mason and Zhou [78, Appendix B]. We sketch the argument for the reader's convenience, following [78]. Because of $\exp(x) \geq x^k/k!$ for all $x \geq 0$, we have

$$\mathbb{E}\big[|X|^{k\alpha}\big] \leq \frac{k!}{\delta^k} \mathbb{E}\big[\exp(\delta|X|^{\alpha})\big] =: \frac{k!}{\delta^k} C(\delta).$$

Given $m \in \mathbb{N}$, define $k = \lceil m/\alpha \rceil = \lceil m(1+\gamma) \rceil$. Thus $\beta := m/(k\alpha) \leq 1$ and $y \mapsto y^\beta$ is concave. Therefore

$$\mathbb{E}\big[|X|^m\big] \leq \Big(\mathbb{E}\big[|X|^{k\alpha}\big]\Big)^{m/(k\alpha)}.$$

Because of $y^\beta \leq \max(1, y)$ for all $y \geq 0$ and $\beta \in (0, 1)$, we conclude

$$\mathbb{E}\big[|X|^m\big] \leq \max\Big(1, \frac{k!}{\delta^k} C(\delta)\Big).$$

The proof is completed by comparing $k!$ with $m!^{1+\gamma}$, aided by Stirling's formula, see [78, Appendix B] for details.

Next we address the equivalence of moment and cumulant conditions.

"$(ii) \Rightarrow (i)$" We follow Rudzkis, Saulis, and Statulevičius [99, Lemma 2]. Set $m_j := \mathbb{E}\big[X^j\big]$. The moment-cumulant relation yields

$$\frac{\kappa_j}{j!} = [z^j] \log\Big(1 + \sum_{k=1}^{j} \frac{m_k}{k!} z^k\Big)$$

meaning that $\kappa_j/j!$ is equal to the coefficient of z^j in the power series obtained by expanding the logarithm on the right-hand side. Let $r > 0$ small enough so that

$$\sum_{k=1}^{j} \frac{|m_k|}{k!} r^k < 1.$$

Then $z \mapsto \log\Big(1 + \sum_{k=1}^{j} \frac{1}{k!} m_k z^k\Big)$ is analytic in $|z| < r$ and continuous in $|z| \leq r$. By Cauchy's integral formula, the coefficient can be represented by a contour integral over the contour $|z| = r$ and we find

$$\left|\frac{\kappa_j}{j!}\right| = \left|\frac{1}{2\pi i} \oint \log\Big(1 + \sum_{k=1}^{j} \frac{m_k}{k!} z^k\Big) \frac{\mathrm{d}z}{z^{j+1}}\right| \leq r^{-j} \left|\log\Big(1 - \sum_{k=1}^{j} \frac{|m_k|}{k!} r^k\Big)\right|,$$

where we have used $|\log(1+z)| \leq -\log(1-|z|)$ for $z \in \mathbb{C}$ with $|z| < 1$. For $\delta \in (0, 1)$ small enough we check that the choice

$$r := \frac{\delta}{j!^{\gamma/j} H^{(j-2)/j}}$$

is admissible. Notice $r^{-j} = \delta^{-j} j!^\gamma H^{j-2}$ resp. $r^k = \delta^k H^{-k(j-2)/j}(j!)^{-\gamma k/j}$ for $k \leq j$. We have

$$\sum_{k=1}^{j} \frac{|m_k|}{k!} r^k \leq \frac{1}{2} r^2 + \sum_{k=3}^{j} k!^\gamma H^{k-2} r^k = \frac{1}{2} r^2 + \sum_{k=3}^{j} \frac{k!^\gamma}{j!^{\gamma k/j}} \frac{H^{k-2}}{H^{k(j-2)/j}} \delta^k.$$

For $k \leq j$ we have $k!^j \leq j!^k$ (this can be proven by induction over $j \geq k$ at fixed k). In addition $j(k-2) \leq k(j-2)$, hence $H^{k-2} \leq H^{k(j-2)/j}$ and

$$\sum_{k=1}^{j} \frac{|m_k|}{k!} r^k \leq \frac{1}{2} \delta^2 + \sum_{k=3}^{j} \delta^k \leq \frac{1}{2} \delta^2 + \frac{\delta^3}{1-\delta} =: C_\delta.$$

Clearly $C_\delta < 1$ for small δ. Setting $C'_\delta := -\frac{1}{\delta^2} \log(1 - C_\delta)$ we get

$$|\kappa_j| \leq C'_\delta \, j!^{1+\gamma} \left(\frac{1}{\delta} H\right)^{j-2}.$$

Set $\Delta := \frac{\delta}{H} \min(1, \frac{1}{C'_\delta})$, then $|\kappa_j| \leq j!^{1+\gamma}/\Delta^{j-2}$ for all $j \geq 3$.

"$(i) \Rightarrow (ii)$" Suppose that X satisfies $(\mathcal{S}_\gamma)$. From the moment-cumulant relations, we get

$$\frac{m_j}{j!} = [z^j] \exp\Big(\sum_{\ell=2}^{j} \frac{\kappa_\ell}{\ell!} z^\ell\Big),$$

meaning that m_j is equal to the coefficient of z^j on the right-hand side. The Cauchy inequality yields

$$\frac{|m_j|}{j!} \leq \sup_{r>0} \frac{1}{r^j} \exp\Big(\sum_{\ell=2}^{j} \frac{|\kappa_\ell|}{\ell!} r^\ell\Big).$$

From here on the proof is similar to the proof of the implication $(ii) \Rightarrow (i)$ and therefore omitted.   □

3. Related techniques and applications

3.1. Moderate deviations vs. heavy-tailed behavior

Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of normalized real-valued random variables such that each X_n satisfies condition $(\mathcal{S}_\gamma)$ for some n-dependent $\Delta_n > 0$. Further assume that $\Delta_n \to \infty$ as $n \to \infty$. Let us evaluate $\mathbb{P}(X_n \geq x_n)$ for sequences x_n with $x_n \to \infty$ and $x_n = o(\Delta_n^{1/(1+2\gamma)})$. Theorem 2.3, combined with the Gaussian tail estimate (4.16), yields

$$\mathbb{P}(X_n \geq x_n) = \big(1 + o(1)\big) \frac{1}{\sqrt{2\pi} \, x_n} \exp\left(-\frac{x_n^2}{2} + O\Big(\frac{x_n^3}{\Delta_n}\Big)\right).$$

In general the correction term $O(x_n^3/\Delta_n)$ from the Cramér-Petrov series does not go to zero, however for $x_n = o(\Delta_n)$ it is negligible compared to $x_n^2/2$. Hence, if we are only interested in a rough asymptotics on the exponential scale, we may drop it and write

$$\mathbb{P}(X_n \geq x_n) = \exp\left(-\frac{x_n^2}{2}\big(1 + o(1)\big)\right).$$

This asymptotic statement can be lifted to a full moderate deviation principle [22, Chapter 3.7], where probabilities of more general sets are examined.

Theorem 3.1. *[28, Theorem 1.1] Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables with $\mathbb{E}[X_n] = 0$, $\mathbb{V}(X_n) = 1$ that satisfy condition $(\mathcal{S}_\gamma)$ with n-dependent Δ_n but fixed $\gamma \geq 0$. Suppose that $\Delta_n \to \infty$. Then, for every sequence $a_n \to \infty$ with*

$$a_n = o(\Delta_n^{1/(1+2\gamma)}),$$

the sequence $(X_n/a_n)_{n\in\mathbb{N}}$ satisfies the moderate deviation principle *with speed a_n^2 and rate function $I(x) = x^2/2$.*

Thus for every Borel set $A \subset \mathbb{R}$, the lower and upper bounds

$$\liminf_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\left(\frac{X_n}{a_n} \in A\right) \geq -\inf_{x\in\text{int}(A)} \frac{x^2}{2}$$

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\left(\frac{X_n}{a_n} \in A\right) \leq -\inf_{x\in\text{cl}(A)} \frac{x^2}{2}$$

hold, where the infimum and supremum are taken over the interior respectively closure of A.

The scale $\Delta_n^{1/(1+2\gamma)}$ is not merely technical. For sums of i.i.d. random variables that satisfy Cramér's condition or have Weibull tails, the critical scale $\Delta_n^{1/(1+2\gamma)}$ corresponds to the scale at which the tail behavior switches to Cramér large deviations or heavy-tailed behavior. Precisely, let Y_i, $i \in \mathbb{N}$, be i.i.d. random variables with $\mathbb{E}[Y_i] = 0$, $\mathbb{V}(Y_i) = 1$. Set

$$S_n := Y_1 + \cdots + Y_n, \quad X_n = \frac{1}{\sqrt{n}} S_n.$$

If $\mathbb{E}[\exp(tY_i)] < \infty$ for $|t| < \delta$, then X_n satisfies condition $(\mathcal{S}_\gamma)$ with $\gamma = 0$ and $\Delta_n = c\delta\sqrt{n}$ for some suitable constant $c > 0$. Furthermore by the Cramér LDP, for $x \geq 0$,

$$\liminf_{n\to\infty} \frac{1}{\Delta_n^2} \log \mathbb{P}\left(\frac{X_n}{\Delta_n} \geq x\right) = \liminf_{n\to\infty} \frac{1}{(c\delta)^2} \frac{1}{n} \log \mathbb{P}\left(\frac{S_n}{n} \geq xc\delta\right)$$

$$\geq -\frac{I(c\delta x)}{(c\delta)^2}, \quad I(x) = \sup_{t\in\mathbb{R}}\big(tx - \log \mathbb{E}[e^{tY_1}]\big),$$

similarly for the upper bound. Thus $(X_n/\Delta_n)_{n\in\mathbb{N}}$ satisfies a large deviation principle with speed Δ_n^2, however the rate function $I(x)$ is in general different from $x^2/2$.

If Y_i is integer-valued and $\mathbb{P}(Y_i = k) = (1 + o(1))c\exp(-k^\alpha)$ with $\alpha = 1/(1+\gamma) \in (0,1)$, then on the one hand, by Lemma 2.8 applied to the Y_i's and $\kappa_j(X_n) = n^{-j/2}\sum_{i=1}^n \kappa_j(Y_i)$, the variable X_n satisfies condition $(\mathcal{S}_\gamma)$ with Δ_n of the order of $\sqrt{n}$, and on the other hand, for $s_n \gg n^{1/(2-\alpha)}$

$$\mathbb{P}(S_n = s_n) = nc\exp(-s_n^\alpha + o(s_n^\alpha))$$

as $n \to \infty$, see Nagaev [85]. In the previous asymptotics, the unlikely event that $S_n = s_n$ is best realized by making one out of the n summands very large—this is the typical heavy-tailed behavior [33], different from the collective behavior underpinning large and moderate deviations. The scale $s_n^* = n^{1/(2-\alpha)}$ naturally appears when solving for $s_n^\alpha = s_n^2/n$. As a consequence, for

$$a_n \gg \frac{1}{\sqrt{n}}n^{1/(2-\alpha)} = n^{1/[2(1+2\gamma)]} = \text{const }\Delta_n^{1/(1+2\gamma)},$$

(meaning $a_n/\Delta_n^{1/(1+2\gamma)} \to \infty$) and $x > 0$, we have

$$\frac{1}{a_n^2}\log\mathbb{P}\Big(\frac{X_n}{a_n} \geq x\Big) = \frac{1}{a_n^2}\log\mathbb{P}(S_n \geq \sqrt{n}a_n x) = (1 + o(1))\frac{1}{a_n^2}\big(\sqrt{n}a_n x\big)^\alpha \to 0$$

which is again different from $-x^2/2$.

A more involved example illustrating the role of the critical scale for functionals in a fixed Wiener chaos is given by Schulte and Thäle [102, Proposition 3].

More generally, the domain of validity of the moderate deviation principle should be related to the *small steps sequence* for subexponential random variables studied by Denisov, Dieker and Shneer [23], in turn related to the sequence $\Lambda(n)$ in Ibragimov and Linnik [64, Chapter 11] and in Nagaev [86] (N_n^{**}) in [34, Lemma 2.5]. The small steps sequence in general is smaller than the boundary of the big-jump domain. The latter corresponds to sequences b_n for which $\mathbb{P}(S_n = b_n) \sim n\mathbb{P}(Y_1 = b_n)$, i.e. the dominant effect is having one large summand.

The connection with heavy-tailed variables suggests that different bounds on cumulants—reflecting the behavior of other heavy-tailed laws, e.g. log-normal—might lead to generalizations of Theorem 3.1. This corresponds to a generalization of Linnik's condition (see Section 2.4) of the form $\mathbb{E}[\exp(h(X))] < \infty$ with functions $h(x)$ different from cx^α. Such generalized Linnik conditions are treated by Ibragimov and Linnik in a chapter on "narrow zones of normal attraction" [64, Chapter 11], the class of functions h and the domains of attraction were further improved by Nagaev [86]. However we are not aware of a corresponding generalized Statulevičius condition.

3.2. Mod-phi convergence

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of normalized real-valued random variables such that each X_n satisfies condition $(\mathcal{S}_\gamma)$ for some n-dependent $\Delta_n > 0$ as in the

previous subsection. Then we may define a function R_n on the axis of purely imaginary numbers by

$$\mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX_n}\big] = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tZ}\big]R_n(\mathrm{i}t). \tag{3.1}$$

If $\gamma = 0$, then for $|t| < \Delta_n$ and some $\theta = \theta_n(t) \in [-1,1]$,

$$R_n(\mathrm{i}t) = \exp\Big(\frac{\kappa_3(X_n)\Delta_n}{3!}\frac{(\mathrm{i}t)^3}{\Delta_n} + \theta\Delta_n^2\frac{(t/\Delta_n)^4}{1-|t|/\Delta_n}\Big).$$

If $\kappa_3(X_n)\Delta_n$ converges to some constant $c_3 \in \mathbb{R}$, then it is natural to rescale variables as $t = \Delta_n^{1/3}s$. In view of $\Delta_n^2(\Delta_n^{1/3}/\Delta_n)^4 = \Delta_n^{2-8/3} \to 0$ we obtain

$$\lim_{n\to\infty} R_n\big(\mathrm{i}\Delta_n^{1/3}s\big) = \exp\Big(\frac{c_3}{6}(\mathrm{i}s)^3\Big)$$

uniformly on compact sets. Let us define $\eta(z) := z^2/2$, $\psi(z) := \exp(c_3z^3/6)$, and $Y_n := \Delta_n^{1/3}X_n$, then

$$\lim_{n\to\infty} \exp\big(-\Delta_n^{2/3}\eta(\mathrm{i}s)\big)\,\mathbb{E}\big[\mathrm{e}^{\mathrm{i}sY_n}\big] = \psi(\mathrm{i}s) \tag{3.2}$$

uniformly on compact sets. The convergence (3.2) is a key ingredient to the notion of *mod-phi convergence*, here mod-Gaussian convergence of the sequence $(Y_n)_{n\in\mathbb{N}}$ with speed $\Delta_n^{2/3}$ and limiting function $\psi(z)$. Different full definitions of mod-phi convergence, given functions η, ψ, and a speed sequence, impose slightly different additional conditions [66, 21, 39]. For example, one may or may not impose that the convergence (3.2) extends from the purely imaginary axis to a complex strip (which would require $\gamma = 0$ in the Statulevičius condition), that $\eta(z)$ is the Lévy exponent of some infinitely divisible law, or that $\psi(z)$ is non-zero on a complex strip.

Mod-Gaussian convergence was introduced by Jacod, Kowalski, and Nikeghbali [66]. The original motivation was in random matrix theory and analytic number theory. Concretely, a result by Keating and Snaith [69] as summarized in [66] says that the determinant Z_N of a random matrix in $U(N)$, distributed according to the uniform measure (Haar measure) on $U(N)$, satisfies for all $\lambda \in \mathbb{C}$ with $\mathrm{Re}\,\lambda > -1$,

$$\lim_{N\to\infty} \frac{1}{N^{\lambda^2}}\mathbb{E}\big[|Z_N|^{2\lambda}\big] = \frac{(G(1+\lambda))^2}{G(1+2\lambda)}, \tag{3.3}$$

with G some special function. Eq. (3.3) is clearly in the spirit of (3.2)—set $\lambda = \mathrm{i}s$, $Y_N = \log|Z_N|^2$, and make adequate choices of the speed and limiting function. Subsequent developments include mod-Poisson convergence for random combinatorial structures [6], random vectors, proving mod-Gaussian convergence with dependency graphs, and a systematic study of asymptotic statements on probabilities when mod-phi convergence holds true, see the monograph [39]. The asymptotic bounds and their proofs share some similarities with the bounds on which we focus in this survey, see the discussion [39, Section 5.3].

3.3. Analytic combinatorics. Singularity analysis

Discrete probability and analytic combinatorics [43] share a common complex-analytic toolbox. It is often of advantage to work with generating functions $G(z) = \sum_{n=0}^{\infty} g_n z^n$. In discrete probability, the coefficient g_n represent a probability measure on $\mathbb{N}_0$. In analytic combinatorics, the coefficients are instead related to counting problems, e.g. counting the number of trees on n vertices. When the generating function is well-understood, probabilities or cardinalities can be recovered by complex contour integrals, using Cauchy's formula. These formulas are similar to inversion formulas that express a probability density function or cumulative distribution function in terms of the characteristic function (Fourier transform).

Understanding probabilities then boils down to understanding parameter-dependent contour integrals, for which a plethora of methods are available, for example saddle-point and steepest descent methods, and singularity analysis [43]. *Singularity* refers to the singularities of the function $G(z)$ in the complex plane, among which the dominant singularity $z = R$, the radius of convergence of $G(z)$. *Transfer theorems* go from asymptotic expansions of $G(z)$ near its dominant singularity to asymptotic behavior of the coefficients as $n \to \infty$, a process related to Tauberian theorems for inverse Laplace transforms [43, Chapter VI].

Cumulants fit in very naturally: for an integer-valued, heavy-tailed random variable X, the dominant singularity of the probability generating function $G(z) = \mathbb{E}[z^X]$ is at $z = 1$, and

$$G(\mathrm{e}^t) = \exp\Big(\sum_{j \geq 1} \frac{\kappa_j}{j!} t^j\Big) \quad (z = \mathrm{e}^t \to 1, \, t \to 0).$$

However, even though singularity analysis deals with Taylor expansions with zero radius of convergence, Weibull-like variables do not belong to the class amenable to singularity analysis [43, Chapter VI.6] and therefore the methods described in the book by Flajolet and Sedgewick [43] are not directly applicable (see nevertheless [34] and the references therein).

The methods extend to multivariate generating functions [90] and to sequences of generating functions. The latter enter the stage naturally when working with sequences of random variables, and allow for a derivation of limit laws in random combinatorial structures. The simplest setting is when generating functions can be approximated by powers of simpler generating functions (think of independent random variables!), leading to the framework of *quasi-powers*, see Hwang [62] and [43, Chapter IX.5]. The relation between Hwang's quasi-powers and mod-phi convergence is commented upon in [39, Remark 1.2].

3.4. Dependency graphs

Let Y_α, $\alpha \in I$, be real-valued random variables indexed by some set I of cardinality $N \in \mathbb{N}$. The *mixed cumulants* of the Y_α's are given by an inverse Möbius

transform of mixed moments as

$$\kappa(Y_{\alpha_1}, \ldots, Y_{\alpha_r}) = \sum_{\{B_1, \ldots, B_m\}} (-1)^{m-1}(m-1)! \prod_{\ell=1}^{m} \mathbb{E}\Big[\prod_{i \in B_\ell} Y_{\alpha_\ell}\Big] \qquad (3.4)$$

with summation over set partitions $\{B_1, \ldots, B_m\}$ of $\{1, \ldots, r\}$ of variable number of blocks $m \in \{1, \ldots, r\}$. A classical reference for (3.4) is Leonov and Shiryaev [76], an early mention of Möbius inversion is found in Schützenberger [103]; a detailed historical discussion is given by Speicher[1]. Mixed cumulants of independent variables vanish. The cumulants of $X := \sum_{\alpha \in I} Y_\alpha$ are

$$\kappa_r(X) = \sum_{(\alpha_1, \ldots, \alpha_r) \in I^r} \kappa(Y_{\alpha_1}, \ldots, Y_{\alpha_r}).$$

Now assume that the Y_α's have a dependency structure encoded by a *dependency graph* G. The latter is a graph $G = (I, E(G))$ with vertex set I with the following property: if I_1 and I_2 are disjoint subsets of I not linked by an edge $\{\alpha, \beta\}$ in G, then $Y_\alpha, \alpha \in I_1$ and $Y_\beta, \beta \in I_2$ are independent. Féray, Méliot, and Nikeghbali prove a beautiful tree bound for cumulants.

Lemma 3.2. *[39, Section 9.3] Suppose that $G = (I, E(G))$ is a dependency graph for $Y_\alpha, \alpha \in I$. Assume in addition that $|Y_\alpha| \leq A$ almost surely. For $r \in \mathbb{N}$, let $\mathcal{T}_r$ be the set of tree graphs with vertex set $\{1, \ldots, r\}$. Then for all $\alpha_1, \ldots, \alpha_r \in I$,*

$$\big|\kappa(Y_{\alpha_1}, \ldots, Y_{\alpha_r})\big| \leq 2^{r-1} A^r \sum_{T \in \mathcal{T}_r} \prod_{\{i,j\} \in E(T)} \big(\mathbb{1}_{\{\alpha_i = \alpha_j\}} + \mathbb{1}_{\{\alpha_i \neq \alpha_j, \{\alpha_i, \alpha_j\} \in E(G)\}}\big).$$
$$(3.5)$$

The sum over trees is equal to the number of spanning trees of the graph H with vertex set $\{1, \ldots, r\}$ for which $\{i, j\}$ is an edge if and only if either $\alpha_i = \alpha_j$ or $\alpha_i \neq \alpha_j$ and $\{\alpha_i, \alpha_j\}$ is an edge of the dependency graph.

When the dependency graph has maximum degree D, it is easily checked that for each fixed tree $T \in \mathcal{T}_r$ and all $\alpha_1 \in I$,

$$\sum_{\alpha_2, \ldots, \alpha_r \in I} \prod_{\{i,j\} \in E(T)} \big(\mathbb{1}_{\{\alpha_i = \alpha_j\}} + \mathbb{1}_{\{\alpha_i \neq \alpha_j, \{\alpha_i, \alpha_j\} \in E(G)\}}\big) \leq (D+1)^{r-1}.$$

Summing over α_1 gives an additional factor $N = \#I$. Combining with Cayley's formula $\#\mathcal{T}_r = r^{r-2}$, one finds

$$\big|\kappa_r(X)\big| \leq r^{r-2} \, 2NA\big(2A(D+1)\big)^{r-1}, \qquad (3.6)$$

see [39, Theorem 9.8]. Stirling's formula implies that the sum $X = \sum_\alpha Y_\alpha$, suitably normalized, satisfies the Statulevičius condition $(\mathcal{S}_\gamma)$ with $\gamma = 0$.

---

[1]Blog entry from 2 July 2020:
https://rolandspeicher.com/tag/moment-cumulant-formula/. Last consulted on 1 February 2021.

Applications of the bound (3.6) are found in [39] and [28]. We make two additional remarks. The first remark concerns the appearance of trees. Tree bounds as in Lemma 3.2—with functions $u(\alpha, \beta) \geq 0$ rather than indicators that $\{\alpha, \beta\}$ is in some dependency graph—come up naturally in random fields and statistical mechanics, see for instance Duneau, Iagolnitzer, and Souillard [30]. The existence of tree bounds is sometimes called *strong mixing*. Tree bounds feature prominently in the framework of *complete analyticity* for Gibbs measures, see condition IIb in Dobrushin and Shlosman [25]. The bound (3.6) extends to such soft tree bounds because

$$\sum_{\alpha_1,...,\alpha_r \in I} \sum_{T \in \mathcal{T}_r} \prod_{\{i,j\} \in E(\mathcal{T}_r)} u(\alpha_i, \alpha_j) \leq N r^{r-2} \Big(\max_{\alpha \in I} \sum_{\beta \in I} u(\alpha, \beta) \Big)^{r-1}.$$

Similar considerations appear with *weighted dependency graphs* introduced by Féray [37] and in particular *uniform weighted dependency graphs* [40, Definition 43]. Weighted dependency graphs have been applied, for example, to the Ising model [29], and to many other examples (not covered by statistical mechanics) [37].

Our second remark is of a speculative nature: Lemma 3.2, and its proof in [39] is intriguing because of many similarities with the theory of cluster expansions. In fact the proof of Lemma 3.2 brings up, for a connected graph H, the quantity $\sum_{G \subset H} (-1)^{\#E(G)-1}$ (summation over spanning subgraphs) [39, Lemma 9.12]. This object is centerstage in the theory of cluster expansions and bounding it by trees is fairly standard, see Scott and Sokal [104] and the references therein. Connections between cluster expansions, dependency graphs, and combinatorics have been studied intensely in the context of the Lovász local lemma [104]. The combinatorial proofs by Féray, Méliot, and Nikeghbali open up the intriguing perspective of yet another fruitful connection between cluster expansions and (weighted) dependency graphs.

4. Toolbox

Here we collect a few general lemmas that are of independent interest.

4.1. Characteristic functions, Kolmogorov distance, and smoothing inequality

One key ingredient in the proof of Theorem 2.2 is a bound on the Kolmogorov distance of two measures in terms of an integral involving the characteristic functions. Estimates of this type are fairly classical and enter proofs of the Berry-Esseen inequality following Berry's strategy [10], see for example [35, Chapter XVI], [92, Chapter 5.1] and the survey on smoothing inequalities by Bobkov [12]. For asymptotic expansions, e.g. Edgeworth expansions that capture correction terms in normal approximations, it is customary to deal not only with probability measures but also with signed measures whose density is a Gaussian multiplied by a polynomial [35, Chapter XVI.4].

If μ is a finite signed measure on $\mathbb{R}$, write $\mu = \mu_+ - \mu_-$ for the Jordan decomposition of μ. The cumulative distribution function is $F_\mu(x) = \mu((-\infty, x])$ and the characteristic function is $\chi_\mu(t) = \int_{\mathbb{R}} \exp(itx)\mu(dx)$.

Lemma 4.1. *Let μ and ν be two finite signed measures on $\mathbb{R}$ with total mass 1. Let Y be an auxiliary continuous random variable whose probability density function ρ_Y is even, i.e., $\rho_Y(y) = \rho_Y(-y)$ for all $y \in \mathbb{R}$. Assume that μ has a Radon-Nikodym derivative with respect to Lebesgue measure bounded in absolute value by $q > 0$, and that the negative part of ν, if non-zero, satisfies $\nu_-(\mathbb{R}) \leq \eta$. Then*

$$\sup_{x \in \mathbb{R}} |F_\mu(x) - F_\nu(x)|$$
$$\leq \frac{1}{1 - 2\mathbb{P}(|Y| \geq y_0)} \left(\left[\varepsilon q y_0 + \eta\right] \mathbb{P}(|Y| \leq y_0) + \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|\chi_{\varepsilon Y}(t)\right| \left|\chi_\mu(t) - \chi_\nu(t)\right| \frac{dt}{|t|} \right),$$

for all $\varepsilon \geq 0$ and every $y_0 > 0$ with $\mathbb{P}(|Y| \geq y_0) < \frac{1}{2}$.

Remark. If ν is a probability measure, i.e., $\nu_- = 0$ and $\eta = 0$, then Lemma 4.1 is due to Zolotarev [112] as cited in [101, Lemma 2.5]. Our extension to signed measures ν is needed to fix an erroneous application of Zolotarev's lemma to the normal law $\mu \sim \mathcal{N}(0,1)$ and a signed measure ν that is not necessarily absolutely continuous. Another extension to signed measures is found in [113, Theorem 2], under the condition that the atoms of the signed measures form a discrete subset of $\mathbb{R}$.

Lemma 4.1 is applied to a normal law $\mu = \mathcal{N}(0,1)$—thus $q = 1/\sqrt{2\pi}$—and the random variable Y with probability density function and characteristic function given by

$$\rho_Y(y) = \frac{1 - \cos y}{\pi y^2}, \quad \chi_Y(t) = (1 - |t|)\mathbb{1}_{[-1,1]}(t). \tag{4.1}$$

The choice ρ_Y of smoothing density was already made by Berry [10]. Write $\varepsilon = 1/T$, then the integral error term becomes

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \left|\chi_{\varepsilon Y}(t)\right| \left|\chi_\mu(t) - \chi_\nu(t)\right| \frac{dt}{|t|} = \frac{1}{2\pi} \int_{-T}^{T} \left(1 - \frac{|t|}{T}\right) |\chi_X(t) - \chi_Z(t)| \frac{dt}{|t|}. \tag{4.2}$$

In the proof of Theorem 2.2 we choose $T = 1/\varepsilon$ of the order of $\sqrt{s}$, see Section 7.6. We follow [101] and choose $y_0 = 3.55$. A numerical evaluation yields $\mathbb{P}(|Y| \leq y_0) \approx 0.819717$ and

$$C_1 := \frac{y_0 \mathbb{P}(|Y| \leq y_0)}{1 - 2\mathbb{P}(|Y| \geq y_0)} \simeq 4.5509, \quad C_2 := \frac{1}{1 - 2\mathbb{P}(|Y| \geq y_0)} \simeq 1.5639. \tag{4.3}$$

The numerical values are better than the values appearing in the smoothing inequality in [35, Chapter XVI.3, Lemma 1]. Assume that ν is a probability measure and write J_ε for the integral term. Then

$$\sup_{x \in \mathbb{R}} |F_\mu(x) - F_\nu(x)| \leq \frac{24}{\pi} q\varepsilon + 2J_\varepsilon. \tag{4.4}$$

We note $24/\pi \simeq 7.34 > 4.5509$ and $2 > 1.5639$, hence Lemma 4.1 with $y_0 = 3.55$ yields better constants than Lemma 1 in [35, Chapter XVI.3].

The proof of Lemma 4.1 is based on classical inversion formulas. Recall that if the characteristic function χ of a random variable X is integrable, then the variable has a probability density function given by

$$\rho(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \chi(t) dt$$

and the cumulative distribution function F is given by

$$F(x) = \lim_{a \to -\infty} \int_a^x \rho(y) dy = \lim_{a \to -\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} \Big(\int_a^x e^{-ity} dy \Big) \chi(t) dt$$
$$= \lim_{a \to -\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{ita}}{-it} \chi(t) dt.$$

For general random variables, the previous formula for $F(x)$ holds true in every point x of continuity of F.

The proof of Lemma 4.1 is adapted from the proof of Theorem 2 in [113], see also Lemma 1 in [35, Chapter XVI.3]. To help the reader grasp the probabilistic content, we first prove the lemma when μ and ν are probability measures on $\mathbb{R}$.

Proof of Lemma 4.1 when μ and ν are probability measures. Let X and Z be two random variables with respective distributions μ and ν. We may assume without loss of generality that X, Y, Z are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that Y is independent from X and Z. We have, in every point of continuity x of $F_{X+\varepsilon Y}$ and $F_{Z+\varepsilon Y}$,

$$F_{X+\varepsilon Y}(x) - F_{Z+\varepsilon Y}(x) = \lim_{a \to -\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{ita}}{-it} \chi_Y(\varepsilon t) \big(\chi_X(t) - \chi_Z(t) \big) dt.$$

If $\int_{\mathbb{R}} |\frac{1}{t} \chi_Y(\varepsilon t)(\chi_X(t) - \chi_Z(t))| dt = \infty$, the lemma is trivial, so we only need to treat the case where $t \mapsto \frac{1}{t} \chi_Y(\varepsilon t)(\chi_X(t) - \chi_Z(t))$ is integrable. The Riemann-Lebesgue lemma then yields the simplified expression

$$F_{X+\varepsilon Y}(x) - F_{Z+\varepsilon Y}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \chi_Y(\varepsilon t) \big(\chi_X(t) - \chi_Z(t) \big) \frac{dt}{-it},$$

from which we deduce the bound

$$\sup_{x \in \mathbb{R}} \big| F_{X+\varepsilon Y}(x) - F_{Z+\varepsilon Y}(x) \big| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \big| \chi_Y(\varepsilon t) \big| \big| \chi_X(t) - \chi_Z(t) \big| \frac{dt}{|t|} =: J_\varepsilon. \quad (4.5)$$

This proves the lemma in the case $\varepsilon = 0$. For $\varepsilon > 0$, we need to bound the Kolmogorov distance of X and Z by that of $X + \varepsilon Y$ and $Z + \varepsilon Y$. The relevant inequalities are called *smoothing inequalities* [35, 12]. We condition on values of Y and distinguish cases according to $|Y| \geq y_0$ or $|Y| \leq y_0$. We start with

$|Y| \leq y_0$. We wish to exploit $\sup_{\mathbb{R}} |F_X'(x)| \leq q$ and the monotonicity of F_Z. Notice that, for every $y' \in [-y_0, y_0]$,

$$F_Z\big(x_0 + \varepsilon(y_0 - y')\big) - F_X\big(x_0 + \varepsilon(y_0 - y')\big) \geq F_Z(x_0) - F_X(x_0) - q\varepsilon(y_0 - y'),$$
$$F_Z\big(x_0 - \varepsilon(y_0 + y')\big) - F_X\big(x_0 - \varepsilon(y_0 + y')\big) \leq F_Z(x_0) - F_X(x_0) + q\varepsilon(y_0 + y').$$
$$\text{(4.6)}$$

Because of the independence of Y from X and from Z, we can reinterpret the inequalities as almost sure inequalities conditioned on $Y = y'$. The second inequality yields

$$\mathbb{1}_{\{|Y| \leq y_0\}} \mathbb{E}\Big[\big(\mathbb{1}_{\{Z + \varepsilon Y \leq x_0 - \varepsilon y_0\}} - \mathbb{1}_{\{X + \varepsilon Y \leq x_0 - \varepsilon y_0\}}\big) \,\Big|\, Y\Big]$$
$$\leq \mathbb{1}_{\{|Y| \leq y_0\}}\Big(F_Z(x_0) - F_X(x_0) + q\varepsilon y_0 + q\varepsilon Y\Big) \quad \text{a.s.}$$

We take expectations on both sides, use $\mathbb{E}[Y\mathbb{1}_{\{|Y| \leq y\}}] = 0$ from the parity of Y, and deduce

$$\mathbb{P}(Z + \varepsilon Y \leq x_0 - \varepsilon y_0, |Y| \leq y_0) - \mathbb{P}(X + \varepsilon Y \leq x_0 - \varepsilon y_0, |Y| \leq y_0)$$
$$\leq \Big(F_Z(x_0) - F_X(x_0) + q\varepsilon y_0\Big)\mathbb{P}(|Y| \leq y_0).$$

Using the notation
$$D := \sup_{x \in \mathbb{R}} |F_Z(x) - F_X(x)|.$$

this implies

$$F_Z(x_0) - F_X(x_0)$$
$$\geq \frac{\mathbb{P}(Z + \varepsilon Y \leq x_0 - \varepsilon y_0, |Y| \leq y_0) - \mathbb{P}(X + \varepsilon Y \leq x_0 - \varepsilon y_0, |Y| \leq y_0)}{\mathbb{P}(|Y| \leq y_0)} - q\varepsilon y_0$$
$$= \frac{\mathbb{E}\Big[\mathbb{1}_{\{|Y| \leq y_0\}}\mathbb{E}\big[\mathbb{1}_{\{Z \leq x_0 - \varepsilon(Y + y_0)\}} - \mathbb{1}_{\{X \leq x_0 - \varepsilon(Y + y_0)\}} \mid Y\big]\Big]}{\mathbb{P}(|Y| \leq y_0)} - q\varepsilon y_0$$
$$\geq -\frac{\sup_{x \in \mathbb{R}} |F_{X + \varepsilon Y}(x) - F_{Z + \varepsilon Y}(x)|}{\mathbb{P}(|Y| \leq y_0)} - D\frac{\mathbb{P}(|Y| > y_0)}{\mathbb{P}(|Y| \leq y_0)} - q\varepsilon y_0$$
$$\geq -\frac{J_\varepsilon}{\mathbb{P}(|Y| \leq y_0)} - D\frac{\mathbb{P}(|Y| > y_0)}{\mathbb{P}(|Y| \leq y_0)} - q\varepsilon y_0, \tag{4.7}$$

where we used (4.5) for the last estimate and in the third line, we have used the inequality

$$\big|\mathbb{P}(Z + \varepsilon Y \leq x_0, |Y| > y_0) - \mathbb{P}(X + \varepsilon Y \leq x_0, |Y| > y_0)\big|$$
$$= \Big|\mathbb{E}\Big[\mathbb{1}_{\{|Y| > y_0\}}\mathbb{E}\big[\mathbb{1}_{\{Z \leq x_0 - \varepsilon Y\}} - \mathbb{1}_{\{X \leq x_0 - \varepsilon Y\}} \mid Y\big]\Big]\Big|$$
$$= \Big|\mathbb{E}\Big[\mathbb{1}_{\{|Y| > y_0\}}\big(F_Z(x_0 - \varepsilon Y) - F_X(x_0 - \varepsilon Y)\big)\Big]\Big|$$
$$\leq D\mathbb{P}(|Y| > y_0).$$

By the same arguments the first inequality in (4.6) yields

$$\mathbb{P}(Z + \varepsilon Y \leq x_0 + \varepsilon y_0, |Y| \leq y_0) - \mathbb{P}(X + \varepsilon Y \leq x_0 + \varepsilon y_0, |Y| \leq y_0)$$
$$\geq \Big(F_Z(x_0) - F_X(x_0) - q\varepsilon y_0 \Big) \mathbb{P}(|Y| \leq y_0)$$

and

$$F_Z(x_0) - F_X(x_0)$$
$$\leq \frac{\mathbb{P}(Z + \varepsilon Y \leq x_0 + \varepsilon y_0, |Y| \leq y_0) - \mathbb{P}(X + \varepsilon Y \leq x_0 + \varepsilon y_0, |Y| \leq y_0)}{\mathbb{P}(|Y| \leq y_0)} + q\varepsilon y_0$$
$$\leq \frac{J_\varepsilon}{\mathbb{P}(|Y| \leq y_0)} + D\frac{\mathbb{P}(|Y| > y_0)}{\mathbb{P}(|Y| \leq y_0)} + q\varepsilon y_0. \tag{4.8}$$

With (4.7) and (4.8) we get

$$\big|F_Z(x_0) - F_X(x_0)\big| \leq \frac{J_\varepsilon}{\mathbb{P}(|Y| \leq y_0)} + D\frac{\mathbb{P}(|Y| > y_0)}{\mathbb{P}(|Y| \leq y_0)} + q\varepsilon y_0. \tag{4.9}$$

Taking the sup over $x_0 \in \mathbb{R}$ and solving for D gives

$$D \leq \frac{q\varepsilon y_0 \mathbb{P}(|Y| \leq y_0) + J_\varepsilon}{\mathbb{P}(|Y| \leq y_0) - \mathbb{P}(|Y| > y_0)} \leq \frac{q\varepsilon y_0 \mathbb{P}(|Y| \leq y_0) + J_\varepsilon}{1 - 2\mathbb{P}(|Y| > y_0)},$$

where we used $\mathbb{P}(|Y| \leq y_0) - \mathbb{P}(|Y| > y_0) = 1 - 2\mathbb{P}(|Y| > y_0) > 0$ by the assumption on y_0. This concludes the proof of the lemma. □

Proof of Lemma 4.1 for signed measures μ and ν. Let

$$D := \sup_{x \in \mathbb{R}} \big|F_\mu(x) - F_\nu(x)\big| \tag{4.10}$$

and

$$D_\varepsilon := \sup_{x \in \mathbb{R}} \Big| \nu * P_{\varepsilon Y}\big((-\infty, x]\big) - \mu * P_{\varepsilon Y}\big((-\infty, x]\big) \Big|, \tag{4.11}$$

where $P_{\varepsilon Y}$ is the law of εY and $\mu * P_{\varepsilon Y}$ the convolution of μ and $P_{\varepsilon Y}$. Arguments similar to the proof of Eq. (4.5) yield

$$D_\varepsilon \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \big|\chi_Y(\varepsilon t)\big| \big|\chi_\mu(t) - \chi_\nu(t)\big| \frac{\mathrm{d}t}{|t|}. \tag{4.12}$$

It remains to bound D in terms of D_ε. For signed measures the cumulative distribution function is no longer monotone increasing however

$$F_\nu(x_0 + u) - F_\nu(x_0) = \nu((x_0, x_0 + u]) \geq -\nu_-(\mathbb{R}) = -\eta$$

for all $x_0 \in \mathbb{R}$ and $u \geq 0$. Therefore (4.6) becomes

$$F_\nu\big(x_0 + \varepsilon(y_0 - y')\big) - F_\mu\big(x_0 + \varepsilon(y_0 - y')\big) \geq F_\nu(x_0) - F_\mu(x_0) - q\varepsilon(y_0 - y') - \eta,$$
$$F_\nu\big(x_0 - \varepsilon(y_0 + y')\big) - F_\mu\big(x_0 - \varepsilon(y_0 + y')\big) \leq F_\nu(x_0) - F_\mu(x_0) + q\varepsilon(y_0 + y') + \eta \tag{4.13}$$

for all $y' \in [-y_0, y_0]$. We integrate over y' with respect to the law P_Y of Y, use $\int_{-y}^{y} y' P_Y(\mathrm{d}y') = 0$ because of the parity of Y, and obtain

$$
\int_{-y_0}^{y_0} \Big(F_\nu\big(x_0 + \varepsilon(y_0 - y')\big) - F_\mu\big(x_0 + \varepsilon(y_0 - y')\big) \Big) P_Y(\mathrm{d}y')
$$
$$
\geq \Big(F_\nu(x_0) - F_\mu(x_0) - [q\varepsilon y_0 + \eta] \Big) \mathbb{P}(|Y| \leq y_0),
$$
$$
\int_{-y_0}^{y_0} \Big(F_\nu\big(x_0 - \varepsilon(y_0 + y')\big) - F_\mu\big(x_0 - \varepsilon(y_0 + y')\big) \Big) P_Y(\mathrm{d}y')
$$
$$
\leq \Big(F_\nu(x_0) - F_\mu(x_0) + q\varepsilon y_0 + \eta \Big) \mathbb{P}(|Y| \leq y_0),
$$

$$(4.14)$$

which replaces (4.7) and (4.8). The left-hand sides can be rewritten with the help of convolutions. Let $P_{\varepsilon Y}$ be the distribution of εY, then

$$
\int_{-\infty}^{\infty} F_\mu\big(x_0 + \varepsilon(y_0 - y')\big) P_Y(\mathrm{d}y')
$$
$$
= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_{(-\infty, x_0 + \varepsilon y_0]}(x + \varepsilon y') \mu(\mathrm{d}x) \right) P_Y(\mathrm{d}y') = \big(\mu * P_{\varepsilon Y}\big)\big((-\infty, x_0 + \varepsilon y_0]\big).
$$

A similar identity holds true with F_ν and ν instead of F_μ and μ. For the integral over $\mathbb{R} \setminus [-y_0, y_0]$, we note

$$
\left| \int_{\mathbb{R} \setminus [-y_0, y_0]} \Big(F_\mu\big(x_0 + \varepsilon(y_0 - y')\big) - F_\nu\big(x_0 + \varepsilon(y_0 - y')\big) \Big) P_Y(\mathrm{d}y') \right| \leq D\, \mathbb{P}(|Y| > y_0).
$$

We write the integral over $[-y_0, y_0]$ as the difference of the integral over $\mathbb{R}$ and $\mathbb{R} \setminus [-y_0, y_0]$, apply the triangle inequality, and deduce

$$
\int_{-y_0}^{y_0} \Big(F_\nu\big(x_0 + \varepsilon(y_0 - y')\big) - F_\mu\big(x_0 + \varepsilon(y_0 - y')\big) \Big) P_Y(\mathrm{d}y') \leq D_\varepsilon + D\, \mathbb{P}(|Y| > y_0)
$$

with D_ε given by (4.11). Similarly,

$$
\int_{-y_0}^{y_0} \Big(F_\nu\big(x_0 - \varepsilon(y_0 + y')\big) - F_\mu\big(x_0 - \varepsilon(y_0 + y')\big) \Big) P_Y(\mathrm{d}y') \geq -D_\varepsilon - D\, \mathbb{P}(|Y| > y_0).
$$

Combining this with (4.14), we find

$$
D_\varepsilon + D\, \mathbb{P}(|Y| > y_0) \geq \Big(F_\nu(x_0) - F_\mu(x_0) - [q\varepsilon y_0 + \eta] \Big) \mathbb{P}(|Y| \leq y_0),
$$
$$
-D_\varepsilon - D\, \mathbb{P}(|Y| > y_0) \leq \Big(F_\nu(x_0) - F_\mu(x_0) + q\varepsilon y_0 + \eta \Big) \mathbb{P}(|Y| \leq y_0)
$$

$$(4.15)$$

hence

$$
\mathbb{P}(|Y| \leq y_0)\, |F_\nu(x_0) - F_\mu(x_0)| \leq D\, \mathbb{P}(|Y| > y_0) + D_\varepsilon + (q\varepsilon y_0 + \eta)\, \mathbb{P}(|Y| \leq y_0).
$$

We take the supremum over $x_0 \in \mathbb{R}$, obtain an inequality with D on both sides from which we deduce

$$D \leq \frac{1}{\mathbb{P}(|Y| \leq y_0) - \mathbb{P}(|Y| > y_0)} \Big(D_\varepsilon + (q\varepsilon y_0 + \eta) \, \mathbb{P}(|Y| \leq y_0) \Big).$$

To conclude, we note that the denominator on the right-hand side is equal to $1 - 2\mathbb{P}(|Y| > y_0)$ and bound D_ε by (4.12). □

4.2. Tails and exponential moments of the standard Gaussian

The *Mills ratio* of a continuous random variable is the ratio of its survival function and its probability density function. For standard normal laws, the asymptotic behavior of the Mills ratio [45] is

$$\frac{\mathbb{P}(Z \geq x)}{\sqrt{2\pi}^{-1} \exp(-x^2/2)} = (1 + o(1))\frac{1}{x} \qquad (x \to \infty). \tag{4.16}$$

Eq. (4.16) is complemented by the following lemma.

Lemma 4.2. *Let Z be a standard normal variable. Then we have, for all $\beta \geq 0$,*

$$\mathbb{E}\big[\mathrm{e}^{-\beta Z}\mathbb{1}_{\{Z \geq 0\}}\big] = \mathrm{e}^{\beta^2/2}\mathbb{P}(Z \geq \beta) \geq \frac{1}{\sqrt{2\pi}(\beta + 1)}.$$

Moreover for all $\beta > 0$ and $\eta \in (-\beta, \beta)$,

$$\mathbb{E}\big[\mathrm{e}^{-(\beta+\eta)Z}\mathbb{1}_{\{Z \geq 0\}}\big] = \frac{\beta}{\beta + \theta\eta} \, \mathbb{E}\big[\mathrm{e}^{-\beta Z}\mathbb{1}_{\{Z \geq 0\}}\big]$$

for some $\theta \in [-1, 1]$.

Proof. We compute

$$\mathbb{E}\big[\mathrm{e}^{-\beta Z}\mathbb{1}_{\{Z \geq 0\}}\big] = \frac{1}{\sqrt{2\pi}} \int_0^\infty \mathrm{e}^{-\beta u - u^2/2}\mathrm{d}u = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\beta^2/2} \int_0^\infty \mathrm{e}^{-(u+\beta)^2/2}\mathrm{d}u$$
$$= \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\beta^2/2} \int_\beta^\infty \mathrm{e}^{-y^2/2}\mathrm{d}y = \mathrm{e}^{\beta^2/2}\mathbb{P}(Z \geq \beta).$$

Next, we note

$$\mathbb{P}(Z \geq \beta) = \frac{1}{\sqrt{2\pi}} \int_\beta^\infty \mathrm{e}^{-y^2/2}\mathrm{d}y \geq \frac{1}{\sqrt{2\pi}} \int_\beta^\infty \mathrm{e}^{-y^2/2}\Big(1 - \frac{y}{(1+y)^2}\Big)\mathrm{d}y$$
$$= \frac{1}{\sqrt{2\pi}} \Big[-\frac{\exp(-y^2/2)}{1 + y}\Big]_{y=\beta}^{y=\infty} = \frac{1}{\sqrt{2\pi}} \frac{\exp(-\beta^2/2)}{\beta + 1}.$$

We set $\psi(\beta) := \exp(\beta^2/2)\mathbb{P}(Z \geq \beta) = \mathbb{E}[\exp(-\beta Z)\mathbb{1}_{\{Z \geq 0\}}]$ and $q(\beta) := \beta\psi(\beta)$. Clearly ψ is monotone decreasing. We check that $q(\beta)$ is monotone increasing. Indeed,

$$q'(\beta) = (1 + \beta^2)\mathrm{e}^{\beta^2/2}\mathbb{P}(Z \geq \beta) - \frac{\beta}{\sqrt{2\pi}}.$$

We compute

$$\frac{1}{\sqrt{2\pi}}\frac{\beta}{1+\beta^2}\mathrm{e}^{-\beta^2/2} = \frac{1}{\sqrt{2\pi}}\int_\beta^\infty \Big(-\frac{\mathrm{d}}{\mathrm{d}y}\frac{y}{1+y^2}\,\mathrm{e}^{-y^2/2}\Big)\mathrm{d}y$$

$$= \frac{1}{\sqrt{2\pi}}\int_\beta^\infty \mathrm{e}^{-y^2/2}\frac{y^4+2y^2-1}{(y^2+1)^2}\mathrm{d}y$$

$$\leq \frac{1}{\sqrt{2\pi}}\int_\beta^\infty \mathrm{e}^{-y^2/2}\mathrm{d}y = \mathbb{P}(Z \geq \beta)$$

and deduce $q'(\beta) \geq 0$ for all $\beta \geq 0$, so $q(\beta)$ is indeed monotone increasing. By the monotonicity of ψ and q, if $\eta \in (-\beta, 0)$, then

$$\psi(\beta) \leq \psi(\beta+\eta) \leq \frac{1}{\beta+\eta}\beta\psi(\beta).$$

Similarly, if $\eta \in (0, \beta)$, then

$$\frac{\beta}{\beta+\eta}\psi(\beta) \leq \psi(\beta+\eta) \leq \psi(\beta).$$

In both cases $\psi(\beta+\eta)/\psi(\beta) = \beta/(\beta+\theta\eta)$ for some $\theta \in [0,1]$. □

4.3. Integrals of monotone functions and Kolmogorov distance

Lemma 4.3. *Let F and G be two cumulative distribution functions of some probability measures and $f : [0,\infty) \to \mathbb{R}_+$ a monotone decreasing, continuous function with $\lim_{y\to\infty} f(y) = 0$. Then*

$$\Big|\int_{-\infty}^\infty f\mathbb{1}_{[0,\infty)}\mathrm{d}F - \int_{-\infty}^\infty f\mathbb{1}_{[0,\infty)}\mathrm{d}G\Big| \leq 2f(0)\sup_{y\in\mathbb{R}}|F(y)-G(y)|.$$

The lemma extends to cumulative distribution functions of finite signed measures with total mass 1.

If F and G are continuous at 0 (i.e., the associated measures have no atom at 0), we may write $\int_0^\infty f\mathrm{d}F$ and $\int_0^\infty f\mathrm{d}G$ without creating ambiguities.

Proof. The lemma is a straightforward consequence of an integration by parts for Riemann-Stieltjes integrals (roughly, $\int_0^\infty f\mathrm{d}F = -\int_0^\infty F\mathrm{d}f$ + boundary term). For $n \in \mathbb{N}$, set

$$f_n := f(0)\mathbb{1}_{\{0\}} + \sum_{k=0}^\infty f(\tfrac{k+1}{n})\mathbb{1}_{(k/n,(k+1)/n]},$$

further define $a_{k+1}^n := f(\tfrac{k}{n}) - f(\tfrac{k+1}{n}) \geq 0$. Let μ be the measure on $\mathbb{R}$ with $\mu((-\infty, x]) = F(x)$. Summing by parts, we get

$$\int_{-\infty}^\infty f_n\mathbb{1}_{[0,\infty)}\mathrm{d}F = f(0)\mu(\{0\}) + \sum_{k=0}^\infty \big(f(0) - a_1^n - \cdots - a_{k+1}^n\big)\Big(F\big(\tfrac{k+1}{n}\big) - F\big(\tfrac{k}{n}\big)\Big)$$

$$= f(0)\big(\mu(\{0\}) + 1 - F(0)\big) - \sum_{\ell=0}^{\infty} a_{\ell+1}^n \sum_{k=\ell}^{\infty}\Big(F\big(\tfrac{k+1}{n}\big) - F\big(\tfrac{k}{n}\big)\Big)$$

$$= f(0)\lim_{\varepsilon \searrow 0}\big(1 - F(-\varepsilon)\big) - \sum_{\ell=0}^{\infty} a_{\ell+1}^n\big(1 - F\big(\tfrac{\ell}{n}\big)\big).$$

A similar representation holds true for the integral against G.

Since $0 \le \sum_{\ell=0}^{\infty} a_{\ell+1}^n \le f(0)$, we deduce

$$\Big|\int_{-\infty}^{\infty} f_n \mathbb{1}_{[0,\infty)}\mathrm{d}F - \int_{-\infty}^{\infty} f_n \mathbb{1}_{[0,\infty)}\mathrm{d}G\Big| \le 2f(0)\sup_{y}|F(y) - G(y)|.$$

We pass to the limit $n \to \infty$, note $f_n \nearrow f$ because of the continuity and monotonicity of f, and obtain the lemma. $\qquad\square$

4.4. Positivity of truncated exponentials

Let

$$\exp_m(u) := \sum_{k=0}^{m}\frac{u^k}{k!}$$

denote the truncated exponential function.

Lemma 4.4. *We have $\exp_{2n}(u) > 0$ for all $u \in \mathbb{R}$ and $n \in \mathbb{N}_0$.*

The analogous statement for the exponential series truncated after odd integers $2n + 1$ is false, since $\exp_{2n+1}(u)$ is a polynomial that goes to $-\infty$ as $u \to -\infty$.

Proof. We follow [101, pp. 37–38]. For $u \ge 0$ the inequality is obvious, so we need only treat $u < 0$. Set

$$a_k = \frac{u^{2k-1}}{(2k-1)!} + \frac{u^{2k}}{(2k)!} = \frac{u^{2k-1}}{(2k-1)!}\Big(1 + \frac{u}{2k}\Big) \qquad (k \ge 1).$$

If $u \le -2k$, then $a_k \ge 0$, so if $u \le -2n$, then $a_k \ge 0$ for all $k = 1,\dots,n$ and $\exp_{2n}(u) = 1 + a_1 + \dots + a_n > 0$. If $-2n < u < 0$, then for all $k \ge n + 1$, we have $u > -2k$ hence $a_k < 0$. It follows that

$$\exp_{2n}(u) = \exp(u) - \sum_{k=n+1}^{\infty} a_k > \exp(u) > 0. \qquad\square$$

An alternative proof is based on Taylor's theorem with integral remainder: for $u < 0$ and even m,

$$\exp(u) - \exp_m(u) = \frac{1}{m!}\int_0^u (u - t)^{m+1}\mathrm{e}^t\,\mathrm{d}t = -\frac{1}{m!}\int_u^0 s^{m+1}\mathrm{e}^{u-s}\mathrm{d}s > 0$$

hence $\exp_m(u) \ge \exp(u) \ge 0$.

5. Concentration inequality. Proof of Theorem 2.5

For the proof of Theorem 2.5 we follow the proof presented in [101] for the first part (up until (5.5) below), but then take a short-cut in order to avoid cumbersome numerical evaluations; instead we exploit a relation between truncated exponentials and Poisson distributions, which we learnt about from Rudzkis and Bakshaev [98].

The first step of the proof consists in applying Markov's inequality to the truncated exponential $\exp_{2n}(x)$, which is monotone increasing on $\mathbb{R}_+$. This yields, for every $h > 0$ and $n \in \mathbb{N}$, and $x \geq 0$,

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[\exp_{2n}(hX)]}{\exp_{2n}(hx)} = \frac{1}{\exp_{2n}(hx)} \sum_{k=0}^{2n} \frac{h^k}{k!} m_k. \tag{5.1}$$

We check that

$$\sum_{k=0}^{2n} \frac{h^k}{k!} m_k \leq \exp_n\Big(\sum_{j=2}^{2n} \frac{h^j}{j!} |\kappa_j|\Big). \tag{5.2}$$

It is enough to show that for each $k = 2, \ldots, 2n$, the moment m_k is smaller than the coefficient of h^k of the series obtained by expanding the right-hand side of the inequality. By definition of the cumulants, the moment m_k is equal to the coefficient of h^k in the formal power series $\exp(\sum_{j=2}^{\infty} h^j \kappa_j / j!)$, which gives

$$m_k = \sum_{r=1}^{\infty} \frac{1}{r!} \sum_{\substack{j_1,\ldots,j_r \geq 2: \\ j_1 + \cdots + j_r = k}} \frac{\kappa_{j_1}}{j_1!} \cdots \frac{\kappa_{j_r}}{j_r!}.$$

For $k \leq 2n$, the only non-zero contributions are from $r \leq n$ and $2 \leq j_\ell \leq 2n$, therefore

$$|m_k| \leq \sum_{r=1}^{n} \frac{1}{r!} \sum_{\substack{2 \leq j_1,\ldots,j_r \leq 2n: \\ j_1 + \cdots + j_r = k}} \frac{|\kappa_{j_1}|}{j_1!} \cdots \frac{|\kappa_{j_r}|}{j_r!}.$$

The right-hand side is equal to the coefficient of h^k in $\exp_n\big(\sum_{j=2}^{n} |\kappa_j| h^j / j!\big)$. This completes the proof of (5.2).

Next we show that for a specific x-dependent choice of h and n, we may bound

$$\sum_{j=2}^{2n} \frac{|\kappa_j|}{j!} h^j \leq \frac{hx}{2}. \tag{5.3}$$

Consider first the case $H = 1$. Choose $h := x(x\overline{\Delta})^{1/(1+\gamma)}/(x^2 + (x\overline{\Delta})^{1/(1+\gamma)})$ so that

$$\frac{1}{hx} = \frac{1}{x^2} + \frac{1}{(x\overline{\Delta})^{1/(1+\gamma)}}. \tag{5.4}$$

By the assumptions on the cumulants, $|\kappa_j|/j! \leq j!^\gamma/(2^{1+\gamma}\overline{\Delta}^{j-2})$ and hence

$$\sum_{j=2}^{2n} \frac{|\kappa_j| h^j}{j!} \leq \frac{h^2}{2} \sum_{j=2}^{2n} \Big(\frac{j!}{2}\Big)^\gamma \Big(\frac{h}{\overline{\Delta}}\Big)^{j-2}.$$

A straightforward induction over n shows that for $j \leq 2n$, we have $j!/2 \leq (2n)^{j-2}$, thus if we pick $n \leq hx/2$ then $j!/2 \leq (hx)^{j-2}$ for $j \leq 2n$ and

$$\sum_{j=2}^{2n} \frac{|\kappa_j|}{j!} h^j \leq \frac{h^2}{2} \sum_{j=2}^{2n} \left((hx)^\gamma \frac{h}{\overline{\Delta}} \right)^{j-2} = \frac{h^2}{2} \sum_{j=2}^{2n} q^{j-2}, \quad q := (hx)^\gamma \frac{h}{\overline{\Delta}}.$$

Noticing $q = (hx)^{1+\gamma}/(\overline{\Delta}x) < 1$, we deduce

$$\sum_{j=2}^{2n} \frac{|\kappa_j|}{j!} h^j \leq \frac{h^2}{2} \frac{1}{1-q}.$$

Set $\tilde{q} = q^{1/(1+\gamma)}$, then $q < \tilde{q} < 1$ and by (5.4), $1 = \frac{h}{x} + \tilde{q}$ hence $x = h/(1-\tilde{q})$ and

$$\sum_{j=2}^{2n} \frac{|\kappa_j|}{j!} h^j \leq \frac{h^2}{2} \frac{1}{1-\tilde{q}} = \frac{hx}{2}.$$

This completes the proof of (5.3).

The bounds (5.1), (5.2) and (5.3) yield

$$\mathbb{P}(X \geq x) \leq \frac{\exp_n(hx/2)}{\exp_{2n}(hx)} \tag{5.5}$$

for $2n \leq hx$. If we had exponentials instead of truncated exponentials, then the right-hand side of the previous inequality would be $\exp(-hx/2)$ and we would be done, in the case $H = 1$.

Let us choose $n := \lfloor (hx)/2 \rfloor$, then $2n \leq hx < 2n + 2$. Let N_{2n} be a Poisson random variable with parameter $2n$. Then

$$\mathrm{e}^{-hx} \exp_{2n}(hx) \geq \mathrm{e}^{-(2n+2)} \exp_{2n}(2n) = \mathrm{e}^{-2} \mathbb{P}(N_{2n} \leq 2n).$$

The variable N_{2n} is equal in distribution to the sum of $2n$ i.i.d. Poisson variables with parameter 1. Applying the central limit theorem, one finds that

$$\lim_{n \to \infty} \mathbb{P}(N_{2n} \leq 2n) = \frac{1}{2}.$$

As a consequence, there exist $c, m > 0$ such that for all sufficiently large $hx \geq m$ (hence large n),

$$\exp_{2n}(hx) \geq c\,\mathrm{e}^{hx} \geq c\,\mathrm{e}^{hx/2} \exp_n(hx/2).$$

Together with (5.5) this gives

$$\mathbb{P}(X \geq x) \leq \frac{1}{c}\,\mathrm{e}^{-hx/2}$$

for all $hx \geq m$. Replacing $1/c$ by $C := \max(1/c, \exp(m/2))$ we obtain an inequality that is true for all values of hx. This completes the proof of the theorem when $H = 1$.

For general $H > 0$, define $\tilde{X} := X/\sqrt{H}$. Then $\tilde{X}$ satisfies the assumption of the theorem with $H = 1$ and $\overline{\Delta}$ replaced with $\tilde{\Delta} = \overline{\Delta}\sqrt{H}$ and the proof is easily concluded. □

6. Bounds under Cramér's condition. Proof of Theorem 2.1

Here we prove Theorem 2.1 on random variables that have exponential moments ($\mathcal{S}$), which corresponds to $s = \infty$ in condition ($\mathcal{S}^*$). This helps explain the strategy for finite s, and some of the estimates are reused for finite s. The Cramér-Petrov series $\sum_{j=3}^{\infty} \lambda_j x^j$ is defined in Appendix A. Proposition A.2 shows that under the conditions of Theorem 2.1, we have

$$|\lambda_j| \leq \frac{\Delta^2/2}{(3\Delta/10)^j}.$$

Proof of Theorem 2.1. The proof of Theorem 2.1 comes in several steps:

1. Introduce a tilted variable X_h, as often done in the proof of Cramér's large deviation principle, and a centered normalized version $\widehat{X}_h$ with $\mathbb{E}[\widehat{X}_h] = 0$ and $\mathbb{V}[\widehat{X}_h] = 1$.
2. Estimate the Kolmogorov distance between the law of $\widehat{X}_h$ and the standard normal distribution by comparing characteristic functions.
3. Undo the tilt: express $\mathbb{P}(X \geq x)$ in terms of $\widehat{X}_h$, conclude with the help of step (2).

The exponential moments $\mathbb{E}[\exp(tX)]$ are finite for all $t \in (-\Delta, \Delta)$ (or $t \in \mathbb{C}$ with $|\operatorname{Re} t| \leq \Delta$), and the Taylor series of the cumulant generating function $\varphi(t) = \log \mathbb{E}[\exp(tX)]$ has radius of convergence at least Δ,

$$\varphi(t) = \log \mathbb{E}[e^{tX}] = \frac{t^2}{2} + \sum_{j=3}^{\infty} \frac{\kappa_j}{j!} t^j \qquad (|t| < \Delta).$$

Let $I(x) = \sup_{t \in \mathbb{R}}(tx - \varphi(t))$ be the Cramér rate function. In order to estimate $\mathbb{P}(X \geq x)$ with $x > 0$ we work with a tilted random variable X_h, given by

$$\mathbb{P}(X_h \in B) = e^{-\varphi(h)} \mathbb{E}[\exp(hX) \mathbb{1}_B(X)],$$

assuming the equation

$$x = \varphi'(h)$$

admits a solution $h \in (0, \Delta)$. The use of tilted variables is fairly standard in the proof of Cramér large deviation principle. Now, the key idea is to approximate the distribution of X_h by that of a normal random variable with mean x and variance $\sigma^2(h) := \varphi''(h)$. Equivalently, defining

$$\widehat{X}_h := \frac{X_h - x}{\sigma(h)},$$

we approximate $\mathscr{L}(\widehat{X}_h) \approx \mathcal{N}(0, 1)$. The error in the normal approximation is quantified by looking first at characteristic functions. We have

$$\chi_h(t) := \mathbb{E}[\exp(it\widehat{X}_h)] = \mathbb{E}\left[\exp\left(it \frac{X_h - x}{\sigma(h)}\right)\right]$$

$$= \exp\left(-\varphi(h) - \mathrm{i}t\frac{x}{\sigma(h)}\right)\mathbb{E}\left[\exp\left(hX + \mathrm{i}t\frac{X}{\sigma(h)}\right)\right]$$

$$= \exp\left(\varphi\left(h + \mathrm{i}\tfrac{t}{\sigma(h)}\right) - \varphi(h) - \mathrm{i}t\tfrac{x}{\sigma(h)}\right), \tag{6.1}$$

for all $t \in \mathbb{R}$ that are small enough so that $|h + \mathrm{i}t/\sigma(h)| \leq T < \Delta$. A Taylor approximation for φ at h shows for some $\theta \in [-1, 1]$

$$\varphi\left(h + \mathrm{i}\tfrac{t}{\sigma(h)}\right) - \varphi(h) - \mathrm{i}t\tfrac{x}{\sigma(h)} = \tfrac{1}{2}\varphi''(h)\left(\tfrac{\mathrm{i}t}{\sigma(h)}\right)^2 + \frac{1}{6}\varphi'''(t + \mathrm{i}\theta\tfrac{t}{\bar{\sigma}(h)})t^3 = -\tfrac{t^2}{2} + O(t^3).$$

Our cumulant bounds allow for an easy bound on the third derivative: if $z \in \mathbb{C}$ satisfies $|z| \leq T < \Delta$, then

$$|\varphi'''(z)| \leq \sum_{j=3}^{\infty} |\kappa_j| \frac{|z|^{j-3}}{(j-3)!} \leq \frac{1}{\Delta} \sum_{n=1}^{\infty} n\left(\frac{|z|}{\Delta}\right)^{n-1} \leq \frac{1}{\Delta}\frac{T/\Delta}{(1 - T/\Delta)^2}. \tag{6.2}$$

For T bounded away from Δ, we deduce that

$$\chi_h(t) = \exp\left(-\frac{t^2}{2}\left(1 + O\left(\frac{t}{\Delta}\right)\right)\right)$$

which, for large Δ, is close to the characteristic function $\exp(-\frac{t^2}{2})$ of the standard normal variable. Careful estimates based on the smoothing inequality [35, Lemma 2, Chapter XVI.3]

$$\sup_{y \in \mathbb{R}}\left|\mathbb{P}(\widehat{X}_h \leq y) - \mathbb{P}(Z \leq y)\right| \leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\chi_h(t) - \exp(-t^2/2)}{t}\right|\mathrm{d}t + \frac{24}{\pi T}\frac{1}{\sqrt{2\pi}}$$

show that the Kolmogorov distance between $\mathscr{L}(\widehat{X}_h)$ and the normal distribution is of order $O(\frac{1}{\Delta})$: for all $\delta \in (0, 1)$, there exists a constant $C_\delta > 0$ such that whenever the tilt parameter $h = h(x)$ exists and is in $[0, \delta\Delta]$, then

$$D(\widehat{X}_h, Z) := \sup_{y \in \mathbb{R}}\left|\mathbb{P}(\widehat{X}_h \leq y) - \mathbb{P}(Z \leq y)\right| \leq \frac{C_\delta}{\Delta}.$$

The next step consists in undoing the tilt: we express the probability we are after in terms of the tilted recentered variable $\widehat{X}_h$ by

$$\mathbb{P}(X \geq x) = \mathrm{e}^{\varphi(h)}\mathbb{E}\left[\mathrm{e}^{-hX_h}\mathbb{1}_{\{X_h \geq x\}}\right] = \mathrm{e}^{\varphi(h)-hx}\mathbb{E}\left[\mathrm{e}^{-h\sigma(h)\widehat{X}_h}\mathbb{1}_{\{\widehat{X}_h \geq 0\}}\right].$$

An easy lemma on Kolmogorov distances (Lemma 4.3) shows

$$\left|\mathbb{E}\left[\mathrm{e}^{-h\sigma(h)\widehat{X}_h}\mathbb{1}_{\{\widehat{X}_h \geq 0\}}\right] - \mathbb{E}\left[\mathrm{e}^{-h\sigma(h)Z}\mathbb{1}_{\{Z \geq 0\}}\right]\right| \leq 2D(\widehat{X}_h, Z) \leq \frac{2C_\delta}{\Delta}.$$

Further noting $\varphi(h) - hx = -I(x)$, we get

$$\mathbb{P}(X \geq x) = \mathrm{e}^{-I(x)}\left(\mathbb{E}\left[\mathrm{e}^{-h\sigma(h)Z}\mathbb{1}_{\{Z \geq 0\}}\right] + \theta\frac{2C_\delta}{\Delta}\right) \tag{6.3}$$

for some $\theta \in [-1, 1]$. The tilted variance $\sigma^2(h)$ is close to 1 because

$$|\sigma^2(h) - 1| = |\varphi''(h) - \varphi''(0)| \le \frac{h}{\Delta}$$

by the bound (6.2) on the third derivative. The tilt parameter h is close to x because

$$|x - h| = |\varphi'(h) - h| \le \sum_{j=3}^{\infty} \frac{|h|^{j-1}}{(j-1)!} |\kappa_j| \le \Delta \frac{(h/\Delta)^2}{1 - h/\Delta}, \tag{6.4}$$

which yields $x - h = O(h^2/\Delta)$. Altogether we get

$$h\sigma(h) = x\left(1 + O\left(\tfrac{h}{\Delta}\right)\right).$$

Bounds on the standard normal distribution and a completion of squares given in Lemma 4.2, together with $\frac{1}{1 + O(h/\Delta)} = 1 + O(h/\Delta)$, yield

$$\mathbb{E}\left[e^{-h\sigma(h)Z} \mathbb{1}_{\{Z \ge 0\}}\right] = \frac{1}{1 + O(h/\Delta)} \mathbb{E}\left[e^{-xZ} \mathbb{1}_{\{Z \ge 0\}}\right]$$

$$= \left(1 + O(h/\Delta)\right) e^{x^2/2} \mathbb{P}(Z \ge x).$$

Inserting this expression into (6.3) we get

$$\mathbb{P}(X \ge x) = e^{-I(x) + x^2/2}\left(\left(1 + O(\tfrac{h}{\Delta})\right)\mathbb{P}(Z \ge x) + \theta\frac{C_\delta}{\Delta} e^{-x^2/2}\right).$$

We factor out $\mathbb{P}(Z \ge x)$, use the lower bound for $\mathbb{P}(Z \ge x)$ from Lemma 4.2, remember (6.4), and obtain

$$\mathbb{P}(X \ge x) = e^{-I(x) + x^2/2} \mathbb{P}(Z \ge x)\left(1 + O\left(\tfrac{x+1}{\Delta}\right)\right).$$

By Appendix A, given $\delta \in (0, 1)$, there exists a constant $c_\delta > 0$ such that for all $x \in [0, c_\delta \Delta]$, the Cramér-Petrov series converges, and the equation $\varphi'(h) = x$ has a unique solution $h = h(x)$ and this solution is in $[0, \delta\Delta]$. Going through the previous estimates carefully, we see that there is a constant $C_\delta > 0$ such that for all $x \in [0, c_\delta \Delta]$,

$$\mathbb{P}(X \ge x) = \mathbb{P}(Z \ge x)\left(1 + \theta C_\delta \tfrac{x+1}{\Delta}\right) \exp\left(\sum_{j=3}^{\infty} \lambda_j x^j\right). \qquad \square$$

7. Bounds with finitely many moments. Proof of Theorem 2.2

Now we turn to condition $(\mathcal{S}^*)$ that $|\kappa_j| \le (j-2)!/\Delta^{j-2}$ for $j = 3, \ldots, s+2$ with $3 \le s \le 2\Delta^2$. Remember that the random variables are centered and normalized as $\mathbb{E}[X] = 0$, $\mathbb{V}(X) = 1$.

7.1. Introducing a modified tilted measure

The Taylor series of $\varphi(t)$ may have radius of convergence zero, so we work instead with the truncated functions

$$\tilde{\varphi}(t) = \sum_{j=2}^{s} \frac{\kappa_j}{j!} t^j, \qquad e^{\tilde{\varphi}(t)} =: \sum_{j=0}^{\infty} \frac{\tilde{m}_j}{j!} t^j. \tag{7.1}$$

Notice $\tilde{m}_j = \mathbb{E}[X^j] =: m_j$ for $j \leq s+2$. The random variable X may have infinite exponential moments, so the exponential tilt is no longer possible; we replace the exponential $\exp(tx)$ in the tilt by

$$g_t(x) := \exp_s(tx) + x^2 \tilde{r}(t) \tag{7.2}$$

with

$$\exp_s(tx) := \sum_{j=0}^{s} \frac{(tx)^j}{j!}, \quad \tilde{r}(t) := \sum_{j=s+1}^{\infty} \frac{1}{j!} \tilde{m}_j t^j. \tag{7.3}$$

The truncated exponential is fairly natural, the additional term $x^2 \tilde{r}(t)$ in $g_t(x)$ ensures that the tilted measure μ_h defined below has total mass 1. For small $h = h(x)$ such that

$$x = \tilde{m}(h) = \tilde{\varphi}'(h) \quad \text{and} \quad \tilde{\sigma}_h^2 := \tilde{\varphi}''(h),$$

we introduce a signed measure μ_h on $\mathbb{R}$ by

$$\mu_h(B) = e^{-\tilde{\varphi}(h)} \mathbb{E}\Big[g_h(X) \mathbb{1}_{\{(X-x)/\tilde{\sigma}_h \in B\}}\Big]. \tag{7.4}$$

Because of

$$\mathbb{E}[g_h(X)] = \sum_{j=0}^{s} \frac{m_j}{j!} h^j + \mathbb{E}[X^2] \sum_{j=s+1}^{\infty} \frac{\tilde{m}_j}{j!} h^j = e^{\tilde{\varphi}(h)}, \tag{7.5}$$

the signed measure is normalized, i.e., $\mu_h(\mathbb{R}) = 1$, however the function g_h may take negative values so μ_h is not necessarily a probability measure. Nevertheless, we will see that μ_h is close to a normal distribution. For later purposes we note the inverse relation to (7.4): for all Borel sets $B \subset \mathbb{R}$,

$$\mathbb{P}(X \in B) = e^{\tilde{\varphi}(h)} \int_{\mathbb{R}} \frac{1}{g_h(\tilde{\sigma}_h y + x)} \mathbb{1}_B(\tilde{\sigma}_h y + x) \mathrm{d}\mu_h(y). \tag{7.6}$$

Later we evaluate the function $\tilde{\varphi}(z)$ for complex parameters z. A visual summary of various quantities that are introduced for proofs is given in Figure 1 below.

7.2. Moment estimates

The key step of the proof will be, just as in the case $s = \infty$, to show that the Fourier transform $\chi_h(t)$ of μ_h is close to $\exp(-t^2/2)$, for small t, so that we can apply Zolotarev's lemma. A new feature compared to $s = \infty$ is that we have to deal with truncation errors. In order to estimate them, it is helpful to have bounds on the quantities $\tilde{m}_j$ and on truncated moments.

Lemma 7.1. *Let $\tilde{m}_k$ be given by* (7.1). *We have for all $k \in \mathbb{N}$,*

$$\frac{1}{k!}|\tilde{m}_k| \leq \max\Big(\frac{1}{\sqrt{k/(2\mathrm{e})}^k}, \frac{1}{(\Delta/\sqrt{\mathrm{e}})^k}\Big) = \begin{cases} 1/\sqrt{k/(2\mathrm{e})}^k, & k \leq 2\Delta^2, \\ 1/(\Delta/\sqrt{\mathrm{e}})^k, & k > 2\Delta^2. \end{cases}$$

Moreover the moments $m_k := \mathbb{E}[X^k]$ satisfy $m_k = \tilde{m}_k$ for $k = 0, 1, \ldots, s+2$.

Remark. For small k, the $\tilde{m}_k$'s satisfy a bound inherited from the moments of Gaussian variables. For large k, the $\tilde{m}_k$'s have a behavior closer to the moments of a random variable that has exponential moments $\mathbb{E}[\exp(tY)]$ up to order $|t| < \Delta$, e.g., an exponential variable $Y \sim \mathrm{Exp}(\Delta)$. Indeed, for $Z \sim \mathcal{N}(0,1)$ as $k \to \infty$,

$$\frac{1}{k!}\mathbb{E}\big[|Z|^k\big] = \frac{1}{k!}\frac{2^{k/2}\Gamma(\frac{k+1}{2})}{\sqrt{\pi}} \sim \frac{\sqrt{2}}{k!}\Big(\frac{k}{\mathrm{e}}\Big)^{k/2} \sim \frac{1}{\sqrt{\pi k}(k/\mathrm{e})^{k/2}} \leq \frac{1}{\sqrt{k/(2\mathrm{e})}^k},$$

where we have used Stirling's formula $\Gamma(x+1) \sim \sqrt{2\pi x}(x/\mathrm{e})^x$ for the Gamma function and for the factorial $k! = \Gamma(k+1)$. On the other hand the moments of an exponential random variable $Y \sim \mathrm{Exp}(\Delta)$ are given by

$$\frac{1}{k!}\,\mathbb{E}\big[Y^k\big] = \frac{1}{\Delta^k} \leq \frac{1}{(\Delta/\sqrt{\mathrm{e}})^k}.$$

Proof of Lemma 7.1 . We have

$$\frac{1}{k!}\,\tilde{m}_k = \sum_{r=1}^{\lfloor k/2 \rfloor} \frac{1}{r!} \sum_{\substack{j_1+\cdots+j_r=k \\ j_\ell=2,\ldots,s}} \frac{\kappa_{j_1}\cdots\kappa_{j_r}}{j_1!\cdots j_r!}. \tag{7.7}$$

Hence

$$\left|\frac{1}{k!}\tilde{m}_k\right| \leq \sum_{r=1}^{\lfloor k/2 \rfloor} \frac{1}{r!} \sum_{\substack{j_1+\cdots+j_r=k \\ j_\ell=2,\ldots,s}} \left(\prod_{\ell=1}^{r} \frac{1}{j_\ell(j_\ell-1)}\right)\frac{1}{\Delta^{k-2r}}$$

$$\leq \sum_{r=1}^{\lfloor k/2 \rfloor} \frac{1}{r!}\left(\sum_{j=2}^{s} \frac{1}{j(j-1)}\right)^r \frac{1}{\Delta^{k-2r}}.$$

Using $\sum_{j=2}^{\infty} \frac{1}{j(j-1)} = \int_0^1 \bigl(-\log(1-t)\bigr)\mathrm{d}t = 1$, we obtain for $\Delta^2 \leq k/2$

$$\left|\frac{1}{k!}\tilde{m}_k\right| \leq \frac{1}{\Delta^k} \sum_{r=1}^{\lfloor k/2 \rfloor} \frac{1}{r!}\Delta^{2r} \leq \frac{1}{\Delta^k}\,\mathrm{e}^{\Delta^2} \leq \frac{1}{\Delta^k}\,\mathrm{e}^{k/2}.$$

For $\Delta^2 \geq k/2$, we estimate instead

$$\left|\frac{1}{k!}\tilde{m}_k\right| \leq \sum_{r=1}^{\lfloor k/2 \rfloor} \frac{1}{r!}\frac{1}{\sqrt{k/2}^{\,k-2r}} \leq \frac{1}{\sqrt{k/2}^{\,k}}\,\mathrm{e}^{k/2}.$$

The moments $\mathbb{E}[X^k]$, $k \leq s+2$, are given by a formula similar to (7.7), but in theory the range of summation indices is now $j_\ell = 2, \ldots, s+2$. However, from $j_1 + \cdots + j_r = k$ and $j_\ell \geq 2$ we get that all indices must be smaller or equal to $k-2 \leq s$, so we are back to Eq. (7.7). $\qquad\square$

Let us quickly explain how Lemma 7.1 affects the parameter choices and the bounds on truncation errors. For $z \in \mathbb{C}$ bounded away from $\Delta/\sqrt{\mathrm{e}}$ and $\sqrt{s/(2\mathrm{e})}$, the remainder term $\tilde{r}(z)$ from (7.3) is bounded by

$$\begin{aligned}
|\tilde{r}(z)| &\leq \sum_{j=s+1}^{\infty} \frac{|z|^j}{(\Delta/\sqrt{\mathrm{e}})^j} + \sum_{j=s+1}^{\infty} \frac{|z|^j}{\sqrt{s/(2\mathrm{e})}^{\,j}} \\
&= O\Bigl(\Bigl(\frac{|z|}{\Delta/\sqrt{\mathrm{e}}}\Bigr)^{s+1}\Bigr) + O\Bigl(\Bigl(\frac{|z|}{\sqrt{s/(2\mathrm{e})}}\Bigr)^{s+1}\Bigr),
\end{aligned}$$

where we have estimated $\max(x,y) \leq x+y$ for $x,y \geq 0$. In the bound there is nothing to be gained from having s larger than $2\Delta^2$. Indeed, if $s > 2\Delta^2$, then it is the first term that dominates, and it corresponds to the bound obtained for $s = 2\Delta^2$. This is why, in assumption (S^*), we do not bother with $s > 2\Delta^2$. It will be convenient to work with

$$s \leq 2\Delta^2, \qquad |z| < a := \sqrt{\frac{s}{4\mathrm{e}}}.$$

Then for all $k \geq s$, we have $\sqrt{\frac{k}{2\mathrm{e}}} \geq \sqrt{\frac{s}{2\mathrm{e}}} = \sqrt{2}\,a$ and $\frac{\Delta}{\sqrt{2}} \geq \frac{1}{\sqrt{\mathrm{e}}}\sqrt{\frac{s}{2}} = \sqrt{2}a$ by the assumption $s \leq 2\Delta^2$, hence

$$\min\Bigl(\sqrt{\frac{k}{2\mathrm{e}}}, \frac{\Delta}{\sqrt{2}}\Bigr) \geq \sqrt{2}a \quad \text{for all } k \geq s$$

and

$$\frac{1}{k!}|\tilde{m}_k| \leq \frac{1}{(\sqrt{2}a)^k} \quad \text{for all } k \geq s. \tag{7.8}$$

It is in this form that Lemma 7.1 is used later on.

In Lemma 7.8 below we need to estimate sums of truncated moments, which are of the type $\sum_{k=0}^{s} \frac{z^k}{k!} \mathbb{E}[|X|^k \mathbb{1}_{\{|X| \geq b\}}]$, for some additional truncation parameter $b \geq 1$. Because of $\mathbb{E}[X^k] = \tilde{m}_k$ for $k = 2, \ldots, s+2$, the bound from Lemma 7.1 extends to the moments $\mathbb{E}[X^k]$, $k \leq s$, and we get for $k \leq s/2$:

$$\mathbb{E}\big[|X|^k \mathbb{1}_{\{|X| \geq b\}}\big] \leq \frac{1}{b^k} \mathbb{E}\big[X^{2k}\big] \leq \frac{1}{b^k} \frac{(2k)!}{\min\big(k/e, \Delta^2/e\big)^k} = \frac{1}{b^k} \frac{(2k)!}{(k/e)^k}, \qquad (7.9)$$

where we have used $2k \leq s \leq 2\Delta^2$. By Stirling's formula,

$$\frac{(2k)!}{k!(k/e)^k} = (1 + o(1))4^k \sqrt{2} \qquad (k \to \infty),$$

so we deduce

$$\frac{1}{k!} \mathbb{E}\big[|X|^k \mathbb{1}_{\{|X| \geq b\}}\big] \leq C \frac{4^k}{b^k} \qquad (7.10)$$

for $k \leq s/2$, with $C \geq \sqrt{2}$ some constant independent of s. This bound has the drawback of not being small for $k = 0$, moreover we need to proceed differently for $k > s/2$. Therefore the bound (7.10) is complemented by the following lemma.

Lemma 7.2. *For $30 \leq s \leq 2\Delta^2$, $a := \sqrt{s/(4e)}$ and $b := 4a$, we have*

$$\frac{1}{k!} \mathbb{E}\big[|X|^k \mathbb{1}_{\{|X| \geq b\}}\big] \leq \begin{cases} \sqrt{2}/a^{\lfloor b \rfloor}, & 0 \leq k \leq b, \\ \sqrt{2}/a^k, & b < k \leq s. \end{cases}$$

Note that $b = \sqrt{4s/e} \geq 1$ and $b = \sqrt{s} \frac{2}{\sqrt{e}} < \frac{s}{2}$ for $s \geq 30$. Set

$$m_k(b) := \mathbb{E}\big[|X|^k \mathbb{1}_{\{|X| \geq b\}}\big].$$

Proof. For the proof, we distinguish the cases $0 \leq k \leq s/2$ and $k > s/2$. For $1 \leq k \leq s/2$, we proceed as in (7.9). With Stirling's formula [96]

$$e^{1/(12n+1)} \sqrt{2\pi n} (n/e)^n \leq n! \leq e^{1/(12n)} \sqrt{2\pi n} (n/e)^n,$$

we obtain

$$\frac{1}{b^k} \frac{(2k)!}{k!(k/e)^k} \leq \frac{1}{a^k} \sqrt{2} e^{\frac{1}{24k} - \frac{1}{12k+1}} \leq \frac{\sqrt{2}}{a^k}$$

since for $k \geq 1$ we have $\frac{1}{24k} - \frac{1}{12k+1} \leq 0$. It follows that (7.10) holds true with $C = \sqrt{2}$, for $k = 1, \ldots, s/2$. This proves in particular the assertion of Lemma 7.2 for $b < k \leq \frac{s}{2}$. For $0 \leq k \leq \lfloor b \rfloor$, using $b \geq 1$, we have

$$\frac{m_k(b)}{k!} \leq \frac{1}{k!} \mathbb{E}\Big[|X|^k \frac{|X|^{\lfloor b \rfloor - k}}{b^{\lfloor b \rfloor - k}} \mathbb{1}_{\{|X| \geq b\}}\Big] \leq \frac{1}{\lfloor b \rfloor!} \mathbb{E}\Big[|X|^{\lfloor b \rfloor} \mathbb{1}_{\{|X| \geq b\}}\Big] = \frac{1}{\lfloor b \rfloor!} m_{\lfloor b \rfloor}(b).$$

We combine this with the bound (7.10) with $C = \sqrt{2}$—which is applicable because $b \leq s/2$—and obtain the first inequality in the lemma.

For $s/2 \leq k \leq s$ and k even, we estimate $m_k(b) \leq \mathbb{E}[|X|^k] = \mathbb{E}[X^k] = \tilde{m}_k$ and by Lemma 7.1,

$$\frac{1}{k!}m_k(b) \leq \frac{1}{[\min(\sqrt{k/(2e)}, \Delta^2/\sqrt{e})]^k} = \frac{1}{\sqrt{k/(2e)}^k} \leq \frac{1}{\sqrt{s/(4e)}^k} = \frac{1}{a^k}.$$

If k is odd, we handle $\mathbb{E}[|X|^k] \neq \mathbb{E}[X^k]$ by bounding the expectation by an even moment:

$$m_k(b) = \mathbb{E}[|X|^k \mathbb{1}_{\{|X| \geq b\}}] \leq \frac{1}{b}\mathbb{E}[|X|^{k+1}] = \frac{1}{b}\mathbb{E}[X^{k+1}] = \frac{1}{b}\tilde{m}_{k+1}$$

hence

$$\frac{1}{k!}m_k(b) \leq \frac{k+1}{b}\frac{\tilde{m}_{k+1}}{(k+1)!} \leq \frac{k+1}{b}\frac{1}{\sqrt{(k+1)/(2e)}^{k+1}}.$$

Now

$$\frac{k+1}{\sqrt{k+1}^{k+1}} = (k+1)^{-\frac{k-1}{2}} \leq \left(\frac{s}{2}+1\right)^{-(k-1)/2}$$

hence

$$\begin{aligned}
\frac{1}{k!}m_k(b) &\leq \frac{s/2+1}{4\sqrt{s/(2e)}}\frac{1}{\sqrt{(s/2+1)/(2e)}^{k+1}} \\
&= \frac{s+2}{8\sqrt{s/(4e)}}\frac{1}{\sqrt{(s+2)/(4e)}}\frac{1}{\sqrt{(s+2)/(4e)}^k} \\
&= \frac{4e\sqrt{s+2}}{8\sqrt{s}}\frac{1}{\sqrt{(s+2)/(4e)}^k} \leq \frac{e}{2}\sqrt{1+\frac{2}{s}}\frac{1}{a^k} \leq \frac{\sqrt{2}}{a^k} \text{ for } s \geq 30,
\end{aligned}$$

since $\frac{e}{2}\sqrt{1+\frac{2}{s}} \leq 1.41 \leq \sqrt{2}$ for $s \geq 30$. $\qquad\square$

7.3. Bounds on tilt parameters

Remember the choice $s \leq 2\Delta^2$, $a = \sqrt{s/(4e)} \leq \Delta/\sqrt{2e}$. From now on θ designates a generic constant in $[-1, 1]$, not always the same, possibly dependent on other constants in statements; for example, $a = b + \theta c$ means "there exists $\theta = \theta(a, b, c) \in [-1, 1]$ such that $a = b + \theta c$. When the quantities involved are complex, θ is a complex number of modulus smaller than or equal to 1.

Lemma 7.3. *Assume that $s \leq 2\Delta^2$. Then for all $x \in [0, \frac{2}{3}a]$, the equation $\tilde{\varphi}'(h) = x$ has a unique solution $h \in [0, a]$. In addition, if $x = \frac{2}{3}a\delta$ with $\delta \in [0, 1]$, then the solution h is in $[0, \delta a]$ and*

$$x = h\left(1 + \theta\frac{\delta}{3}\right), \quad \tilde{\sigma}_h^2 = \varphi''(h) = 1 + \theta\, 0.751\, \delta > 0.$$

Note that the polynomial equation $\tilde{\varphi}'(h) = x$ may have additional solutions $h > \frac{2}{3}a$. The constant 0.751 replaces the constant 0.75 from [101, Eq. (2.21)], which we were not able to reproduce.

Proof. Set $q := 1/\sqrt{2e}$ so that $a \leq q\Delta < \Delta/2$. Fix $\delta \in [0,1]$. For $h \in [0, \delta a]$, we have $h/\Delta \leq q\delta$ and

$$|\tilde{\varphi}''(h) - 1| = \Big|\sum_{j=3}^{s} \frac{\kappa_j}{(j-2)!} h^{j-2}\Big| \leq \sum_{j=3}^{\infty} \Big(\frac{h}{\Delta}\Big)^{j-2} = \frac{h/\Delta}{1 - h/\Delta} \leq \frac{q}{1-q}\delta \leq 0.751\delta.$$

$$(7.11)$$

It follows that $\tilde{\varphi}''(h) > 0$ on $[0, \delta a]$ and $\tilde{\varphi}'(\cdot)$ is a monotone increasing bijection from $[0, \delta a]$ onto $[0, \tilde{\varphi}'(\delta a)]$. Next we bound

$$|\tilde{\varphi}'(h) - h| \leq h \sum_{k=3}^{s} \frac{1}{k-1} \Big(\frac{h}{\Delta}\Big)^{k-2} \leq h\Big(\frac{h/\Delta}{2} + \frac{1}{3}\frac{(h/\Delta)^2}{1 - h/\Delta}\Big)$$

$$= h\frac{h}{\Delta}\Big(\frac{1}{2} + \frac{1}{3}\frac{h/\Delta}{1 - h/\Delta}\Big) \leq h\delta q\Big(\frac{1}{2} + \frac{1}{3}\frac{q}{1-q}\Big) \leq \frac{1}{3}\delta h. \qquad (7.12)$$

It follows in particular that $\tilde{\varphi}'(\delta a) \geq \frac{2}{3}\delta a$. Therefore, for x to be in the image $[0, \tilde{\varphi}'(\delta a)]$ it is sufficient that $0 \leq x \leq \frac{2}{3}\delta a$. Thus $\tilde{\varphi}'$ is a monotone increasing bijection from $[0, \delta a]$ onto an interval containing $\frac{2}{3}\delta a$. Specializing to $\delta = 1$ we see that for every $x \in [0, \frac{2}{3}a]$ there is a unique $h \in [0, a]$ such that $\tilde{\varphi}'(h) = x$. If in addition $x = \frac{2}{3}\delta a$ with $\delta \in [0,1]$, then the solution h must be in $[0, \delta a]$ and the proof is concluded with (7.11) and (7.12). $\quad\square$

Lemma 7.4. *We have $\tilde{\varphi}(h) \geq 0$ for all $h \in [0, a]$.*

Proof. We have

$$h \leq a = \sqrt{\frac{s}{4e}} \leq \frac{\Delta}{\sqrt{2e}} < \frac{\Delta}{2}$$

and

$$\tilde{\varphi}(h) = h^2\Big(\frac{1}{2} + \sum_{j=3}^{s} \frac{\kappa_j}{j!} h^{j-2}\Big) \geq h^2\Big(\frac{1}{2} - \sum_{j=3}^{\infty} \frac{1}{j(j-1)}\Big(\frac{h}{\Delta}\Big)^{j-2}\Big)$$

$$\geq h^2\Big(\frac{1}{2} - \frac{1}{6}\frac{h/\Delta}{1 - h/\Delta}\Big) = h^2\frac{3 - 4h/\Delta}{6(1 - h/\Delta)} \geq 0. \qquad\square$$

Lemma 7.5. *For $0 \leq h \leq \delta a$ with $\delta \leq 1$, we have $|\tilde{r}(h)| \leq \frac{(\delta/\sqrt{2})^{s+1}}{1 - 1/\sqrt{2}}$.*

Proof. By the definition of $\tilde{r}(h)$ in (7.3), the bound (7.8) yields

$$|\tilde{r}(h)| \leq \sum_{k=s+1}^{\infty} \frac{h^k}{(\sqrt{2}a)^k} \leq \frac{(\delta/\sqrt{2})^{s+1}}{1 - \delta/\sqrt{2}}. \qquad\square$$

Lemma 7.6. *For $s \geq 30$ and $h \in [0, \delta a]$, the function g_h given in (7.2) is monotone increasing on $[0, \infty)$ and satisfies $g_h(u) \geq 1$ for all $u \in \mathbb{R}_+$.*

Proof. Since the assertion of the lemma is trivially true for $h = 0$, we assume $h > 0$. We start from

$$g_h'(u) = h\sum_{k=0}^{s-1} \frac{(hu)^k}{k!} + 2u\tilde{r}(h) = h\Big(\exp_{s-1}(hu) + 2hu\frac{\tilde{r}(h)}{h^2}\Big).$$

By the definition of $\tilde{r}(h)$ in (7.3) together with (7.8), i.e. $|\tilde{m}_k| \leq (\sqrt{2}a)^{-k}$ for $k \geq s$, and $h \in (0, \delta a]$

$$
\begin{aligned}
\frac{|\tilde{r}(h)|}{h^2} &\leq \sum_{k=s+1}^{\infty} \frac{h^{k-2}}{(\sqrt{2}a)^k} \leq \frac{1}{2a^2} \sum_{k=s-1}^{\infty} \frac{\delta^k}{\sqrt{2}^k} = \frac{1}{2a^2} \frac{1}{1-\frac{\delta}{\sqrt{2}}} \left(\frac{\delta}{\sqrt{2}}\right)^{s-1} \\
&\leq 2\left(\frac{\delta}{\sqrt{2}}\right)^{s-1}.
\end{aligned}
\tag{7.13}
$$

For the last inequality we have used that for $s \geq 30$, we have

$$
2a^2\left(1 - \frac{\delta}{\sqrt{2}}\right) \geq \frac{s}{2\mathrm{e}}\left(1 - \frac{1}{\sqrt{2}}\right) \geq \frac{1}{2}.
$$

Therefore
$$
\frac{|\tilde{r}(h)|}{h^2} \leq 2\left(\frac{1}{\sqrt{2}}\right)^{s-1} \leq \frac{1}{2}.
$$

We conclude

$$
g_h'(u) \geq h\left(\exp_{s-1}(hu) - hu\right) = h\left(1 + \sum_{k=2}^{s-1} \frac{(hu)^k}{k!}\right) \geq 0
$$

and g_h is monotone increasing. In particular, $g_h(u) \geq g_h(0) = 1$ for all $u \in \mathbb{R}_+$. □

7.4. *Fourier transform of the tilted measure: a first bound*

The Fourier transform of the signed measure μ_h is

$$
\chi_h(t) = \mathrm{e}^{-\tilde{\varphi}(h)} \mathbb{E}\left[g_h(X)\mathrm{e}^{\mathrm{i}t(X-x)/\tilde{\sigma}_h}\right].
$$

Motivated by Eq. (6.1), we set

$$
\tilde{\chi}_h(t) := \exp\left(\tilde{\varphi}(h + \mathrm{i}t/\tilde{\sigma}_h) - \tilde{\varphi}(h) - \mathrm{i}tx/\tilde{\sigma}_h\right).
$$

With the analogue of (7.5) for $z = h + \mathrm{i}t/\tilde{\sigma}_h$ instead of h we see

$$
\chi_h(t) = \tilde{\chi}_h(t) + \mathrm{e}^{-\tilde{\varphi}(h) - \mathrm{i}tx/\tilde{\sigma}_h} \mathbb{E}\left[g_h(X)\mathrm{e}^{\mathrm{i}tX/\tilde{\sigma}_h} - g_{h+\mathrm{i}t/\tilde{\sigma}_h}(X)\right]. \tag{7.14}
$$

Eq. (7.14) is a substitute for Eq. (6.1). The term $\tilde{\chi}_h(t)$ is easily treated with the Taylor series of $\tilde{\varphi}$. Remember the choice $s \leq 2\Delta^2$, $a = \sqrt{s/(4\mathrm{e})} \leq \Delta/\sqrt{2\mathrm{e}}$.

Lemma 7.7. *Assume that* $s \leq 2\Delta^2$. *Fix* $0 < \delta < \delta_2 < 1$, *set* $a := \sqrt{s/(4\mathrm{e})}$ *and* $T := (\delta_2 - \delta)a\tilde{\sigma}_h$, *see Figure 1. Then for all* $h \in [0, \delta a]$ *and* $t \in [-T, T]$ *we have*

$$
\left|\tilde{\chi}_h(t) - \exp(-t^2/2)\right| \leq \frac{|t|}{T}\left(\mathrm{e}^{-t^2/4} - \mathrm{e}^{-t^2/2}\right).
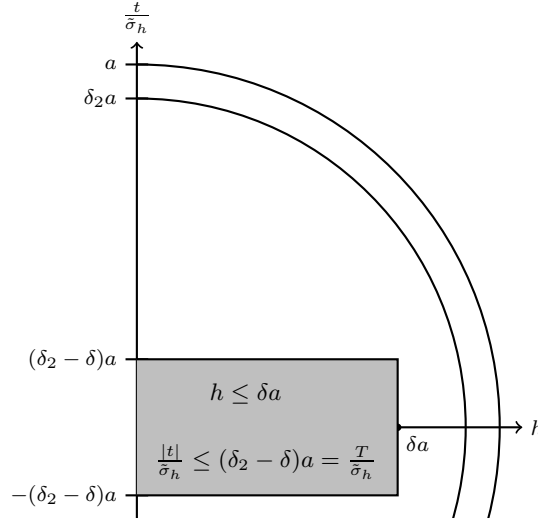$$

FIG 1. *A summary of parameter choices in Lemma 7.7. The function $\tilde{\varphi}(z)$ is evaluated at $z = h + \mathrm{i}t/\tilde{\sigma}_h$ with $0 \le h \le \delta a$ and $|t| \le (\delta_2 - \delta)a\tilde{\sigma}_h = T$ (shaded rectangle), which is contained in the disk $|z| \le \delta_2 a$. The parameter $a = \sqrt{s/(4\mathrm{e})}$ enters the moment bound (7.8).*

Proof. Let $z := h + \mathrm{i}t/\tilde{\sigma}_h$. Then $|z| \le \delta_2 a$. A third order Taylor expansion of $\tilde{\varphi}$ yields, for some $\theta \in [0, 1]$,

$$\tilde{\varphi}\big(h + \mathrm{i}\tfrac{t}{\tilde{\sigma}_h}\big) - \tilde{\varphi}(h) - \mathrm{i}\tfrac{tx}{\tilde{\sigma}_h} = -\frac{1}{2}t^2 + \frac{1}{3!}\tilde{\varphi}'''\big(h + \mathrm{i}\theta\tfrac{t}{\tilde{\sigma}_h}\big)\big(\mathrm{i}\tfrac{t}{\tilde{\sigma}_h}\big)^3.$$

In $|z| < \Delta$, the third derivative is bounded as

$$\Big|\tilde{\varphi}'''(z)\Big| \le \sum_{j=3}^{s} \frac{|\kappa_j|}{(j-3)!}|z|^{j-3} \le \sum_{j=3}^{s} \frac{(j-2)}{\Delta^{j-2}}|z|^{j-3} \le \frac{\mathrm{d}}{\mathrm{d}u}\Big(\sum_{k=0}^{\infty}\big(\tfrac{u}{\Delta}\big)^k\Big)\Big|_{u=|z|}$$

$$= \frac{1}{\Delta}\frac{1}{(1-|z|/\Delta)^2}.$$

Set $L := 3\Delta(1 - |z|/\Delta)^2\tilde{\sigma}_h^3$, then

$$\tilde{\chi}_h(t) = \exp\Big(-\frac{1}{2}t^2\Big(1 + \frac{\theta}{3(1-|z|/\Delta)^2\tilde{\sigma}_h^3}\frac{t}{\Delta}\Big)\Big) = \exp\Big(-\frac{1}{2}t^2\Big(1 + \theta\frac{t}{L}\Big)\Big)$$

for some $\theta \in \mathbb{C}$ with $|\theta| \le 1$. Assuming that $|t| \le T \le L/2$, we have

$$\Big|\mathrm{e}^{-\frac{1}{2}t^2(1+\theta\frac{t}{L})} - \mathrm{e}^{-\frac{1}{2}t^2}\Big| \le \mathrm{e}^{-\frac{1}{2}t^2}\Big(\mathrm{e}^{|t|^3/(2L)} - 1\Big) = \mathrm{e}^{-\frac{1}{2}t^2}\sum_{k=1}^{\infty}\frac{1}{k!}\Big(\frac{|t|^3}{2L}\Big)^k$$

$$= \mathrm{e}^{-\frac{1}{2}t^2}|t|\sum_{k=1}^{\infty}\frac{1}{k!}\Big(\frac{t^2}{4}\Big)^k\frac{(4|t|)^{k-1}}{(2L)^{k-1}}\frac{4}{2L} \le \frac{|t|}{T}\mathrm{e}^{-\frac{1}{2}t^2}\big(\mathrm{e}^{\frac{1}{4}t^2} - 1\big)$$

$$= \frac{|t|}{T}\left(\mathrm{e}^{-\frac{1}{4}t^2} - \mathrm{e}^{-\frac{1}{2}t^2}\right).$$

It remains to check $T \leq L/2$. To that aim we use the lower bound for $\tilde{\sigma}_h^2$ from Lemma 7.3 together with $|z| \leq \delta_2 a$ and we evaluate

$$\frac{T}{L/2} = \frac{(\delta_2 - \delta)a\tilde{\sigma}_h}{3\Delta(1 - |z|/\Delta)^2\tilde{\sigma}_h^3} \leq \frac{\delta_2 - \delta}{1 - 0.751\,\delta}\,\frac{a/\Delta}{(1 - \delta_2 a/\Delta)^2} \leq \frac{\delta_2 - \delta}{1 - 0.751\,\delta} < 1.$$

The last inequality is equivalent to $\delta_2 < 1 + (1 - 0.751)\delta$, it holds true because $\delta_2 < 1$.   □

Lemma 7.8. *Assume that $30 \leq s \leq 2\Delta^2$. Fix $\delta_2 \in (0,1)$, set $a := \sqrt{s/(4\mathrm{e})}$ and further assume that $a^{-1} \leq \delta_2^2$. Then for all $h \geq 0$ and $t \in \mathbb{R}$ with $|h + \mathrm{i}\frac{t}{\tilde{\sigma}_h}| \leq \delta_2 a$, we have*

$$|\chi_h(t) - \tilde{\chi}_h(t)| \leq 4\sqrt{2}\,\frac{\delta_2^{\lfloor 4a \rfloor}}{1 - \delta_2}.$$

Proof. We will use (7.14). By the definition of g_h, we have

$$g_h(X)\mathrm{e}^{\mathrm{i}tX/\tilde{\sigma}_h} - g_{h+\mathrm{i}t/\tilde{\sigma}_h}(X)$$
$$= \left(\exp_s(hX)\mathrm{e}^{\mathrm{i}tX/\tilde{\sigma}_h} - \exp_s((h + \mathrm{i}\tfrac{t}{\tilde{\sigma}_h})X)\right) + X^2\left(\tilde{r}(h)\mathrm{e}^{\mathrm{i}tX/\tilde{\sigma}_h} - \tilde{r}(h + \mathrm{i}\tfrac{t}{\tilde{\sigma}_h})\right)$$
$$\tag{7.15}$$

By (7.14) we have to evaluate the expected value of (7.15). Set $z := h + \mathrm{i}t/\tilde{\sigma}_h$; notice $|h| \leq |z|$. With Lemma 7.1, the expectation of the second term in (7.15) is bounded in absolute value by

$$I_2 := \mathbb{E}\left[X^2\left|\tilde{r}(h)\mathrm{e}^{\mathrm{i}tX/\tilde{\sigma}_h} - \tilde{r}(h + \mathrm{i}t/\tilde{\sigma}_h)\right|\right] \leq \mathbb{E}[X^2]\left(|\tilde{r}(h)| + |\tilde{r}(h + \mathrm{i}t/\tilde{\sigma}_h)|\right)$$
$$\leq 2\sum_{k=s+1}^{\infty} \frac{1}{k!}|\tilde{m}_k|\,|z|^k \leq 2\sum_{k=s+1}^{\infty} \frac{(\delta_2 a)^k}{(\sqrt{2}a)^k}, \tag{7.16}$$

where we have used $\mathbb{E}[X^2] = 1$, $|z| \leq \delta_2 a$ and $\frac{1}{k!}|\tilde{m}_k| \leq \frac{1}{(\sqrt{2}a)^k}$ for $k \geq s$ as derived in (7.8). For the first term in (7.15), we remember $\exp_s(y) = \exp(y) - \sum_{k=s+1}^{\infty} \frac{1}{k!}y^k$ and deduce

$$\exp_s(hX)\mathrm{e}^{\mathrm{i}tX/\tilde{\sigma}_h} - \exp_s((h + \mathrm{i}\tfrac{t}{\tilde{\sigma}_h})X) = -\sum_{k=s+1}^{\infty} \frac{X^k}{k!}\left(\mathrm{e}^{\mathrm{i}tX/\tilde{\sigma}_h}h^k - (h + \mathrm{i}\tfrac{t}{\tilde{\sigma}_h})^k\right).$$

The intuition is of course that this term should be small, however taking expected values we obtain moments $\mathbb{E}[X^k]$ with $k \geq s+1$, over which we have little control. Therefore we introduce an additional truncation parameter $b := 4a > 0$ and use the previous expression for $|X| \leq b$ only: we estimate, with the help of $k! \geq (k/\mathrm{e})^k \geq (s/\mathrm{e})^k$ for $k \geq s$ as well as $b/(s/\mathrm{e}) \leq 1/a$ and $|z|/a \leq \delta_2$

$$I_1^{(1)} := \left|\mathbb{E}\left[\mathbb{1}_{\{|X| \leq b\}}\left(\exp_s(hX)\mathrm{e}^{\mathrm{i}tX/\tilde{\sigma}_h} - \exp_s((h + \mathrm{i}\tfrac{t}{\tilde{\sigma}_h})X)\right)\right]\right|$$

$$\leq 2 \sum_{k=s+1}^{\infty} \frac{(b|z|)^k}{k!} \leq 2 \sum_{k=s+1}^{\infty} \frac{(b|z|)^k}{(s/\mathrm{e})^k} \leq 2 \sum_{k=s+1}^{\infty} \delta_2^k. \tag{7.17}$$

For $|X| > b$, we use instead $|\exp_s(y)| \leq \sum_{k=0}^{s} |y|^k/k!$, which yields

$$I_1^{(2)} := \left| \mathbb{E}\left[\mathbb{1}_{\{|X|>b\}} \left(\exp_s(hX)\mathrm{e}^{\mathrm{i}tX/\tilde{\sigma}_h} - \exp_s((h+\mathrm{i}\tfrac{t}{\tilde{\sigma}_h})X) \right) \right] \right|$$

$$\leq 2 \sum_{k=0}^{s} \frac{|z|^k}{k!} \mathbb{E}\left[|X|^k \mathbb{1}_{\{|X|\geq b\}} \right].$$

Using the bound on the truncated moments provided by Lemma 7.2 and $|z|/a \leq \delta_2$, we can further estimate

$$\frac{1}{2} I_1^{(2)} \leq \sum_{k=0}^{\lfloor b \rfloor} \sqrt{2} \frac{|z|^k}{a^{\lfloor b \rfloor}} + \sum_{k=\lfloor b \rfloor+1}^{s} \sqrt{2} \frac{|z|^k}{a^k} \leq \sum_{k=0}^{\lfloor b \rfloor} \sqrt{2} \frac{|z|^k}{a^{\lfloor b \rfloor}} + \sum_{k=\lfloor b \rfloor+1}^{s} \sqrt{2} \delta_2^k. \tag{7.18}$$

The condition $a^{-1} \leq \delta_2^2$ allows us to treat the sum over $0 \leq k \leq \lfloor b \rfloor$ using $|z|^k/a^{\lfloor b \rfloor} \leq \delta_2^k/a^{\lfloor b \rfloor - k} \leq \delta_2^{2\lfloor b \rfloor - k}$. We get

$$I_1^{(2)} \leq 2\sqrt{2}\left(\sum_{k=0}^{\lfloor b \rfloor} \delta_2^{2\lfloor b \rfloor - k} + \sum_{k=\lfloor b \rfloor+1}^{s} \delta_2^k \right) \leq 2\sqrt{2}\left(\sum_{\ell=\lfloor b \rfloor}^{2\lfloor b \rfloor} \delta_2^{\ell} + \sum_{k=\lfloor b \rfloor+1}^{s} \delta_2^k \right)$$

$$\leq 4\sqrt{2} \sum_{k=\lfloor b \rfloor}^{s} \delta_2^k,$$

where we have used again $2b \leq s$. Altogether

$$I_1^{(2)} + I_1^{(1)} + I_2 \leq 4\sqrt{2} \sum_{k=\lfloor b \rfloor}^{s} \delta_2^k + 2 \sum_{k=s+1}^{\infty} \delta_2^k + 2 \sum_{k=s+1}^{\infty} \left(\frac{\delta_2}{\sqrt{2}} \right)^k \leq 4\sqrt{2} \frac{\delta_2^{\lfloor b \rfloor}}{1 - \delta_2}.$$

The proof is concluded with $|\chi_h(t) - \tilde{\chi}_h(t)| \leq \mathrm{e}^{-\tilde{\varphi}(h)}(I_1^{(1)} + I_1^{(2)} + I_2)$ and the lower bound $\tilde{\varphi}(h) \geq 0$ from Lemma 7.4. □

Lemma 7.9. *Assume that $30 \leq s \leq 2\Delta^2$. Fix $0 < \delta < \delta_2 < 1$, set $a := \sqrt{s/(4\mathrm{e})}$ and $T := (\delta_2 - \delta)a\tilde{\sigma}_h$. Further assume that $a^{-1} \leq \delta_2^2$. Then for all $h \in [0, \delta a]$ and $t \in \mathbb{R}$ with $|t| \leq T$, we have*

$$\left| \chi_h(t) - \exp(-t^2/2) \right| \leq \frac{|t|}{T} \left(\mathrm{e}^{-t^2/4} - \mathrm{e}^{-t^2/2} \right) + 4\sqrt{2} \frac{\delta_2^{\lfloor 4a \rfloor}}{1 - \delta_2}.$$

Proof. The lemma is an immediate consequence from the Lemmas 7.7 and 7.8. □

We want to plug this bound into Zolotarev's lemma. The first part $\tilde{\chi}_h(t)$ is treated exactly as it was for $s = \infty$. The second term is problematic: because of $|z| \geq |h|$, it does not go to zero as $t \to 0$ and the integral of the error estimate against $\mathrm{d}t/t$ diverges. So another bound is needed for very small t.

7.5. Fourier transform of the tilted measure: a second bound.

An alternative bound for very small t is derived with a Taylor expansion of the Fourier transform,

$$\chi_h(t) = 1 + \mathrm{it} \int_{\mathbb{R}} y \, \mathrm{d}\mu_h(y) + \theta \frac{t^2}{2} \int_{\mathbb{R}} y^2 \mathrm{d}|\mu_h|(y), \tag{7.19}$$

see Eq. (4.14) in [35, Chapter VI.4]. When $s = \infty$, the mean and second moment of the tilted normalized measure are 0 and 1, respectively. For finite s, we have instead the following lemma.

Lemma 7.10. *Assume that s is even and satisfies $30 \leq s \leq 2\Delta^2$. Further assume that $x \in [0, \frac{2}{3}\delta a]$ with $\delta \in (0,1)$ and $a = \sqrt{s/(4\mathrm{e})}$. Then*

$$\int_{\mathbb{R}} y \mathrm{d}\mu_h(y) = \theta \, 0.0023 \, \delta^{s+1}, \qquad \int_{\mathbb{R}} y^2 \mathrm{d}|\mu_h|(y) = 1 + \theta \, 0.07724 \, \delta^{s+1}.$$

Proof. First we check that the order of magnitude $O(\delta^{s+1})$ is correct. We evaluate

$$\int_{\mathbb{R}} y \, \mathrm{d}\mu_h(y) = \mathrm{e}^{-\tilde{\varphi}(h)} \mathbb{E}\Big[g_h(X)\frac{X-x}{\tilde{\sigma}_h}\Big] = \tilde{\sigma}_h^{-1}\Big(\mathrm{e}^{-\tilde{\varphi}(h)}\mathbb{E}\big[Xg_h(X)\big] - x\Big). \tag{7.20}$$

By the definition of g_h and in view of $\mathbb{E}[X^k] = m_k = \tilde{m}_k$ for $k \leq s+2$ (see Lemma 7.1), we have

$$\mathbb{E}\big[Xg_h(X)\big] = \sum_{j=0}^{s} \frac{h^j}{j!}m_{j+1} + m_3\tilde{r}(h) = \frac{\mathrm{d}}{\mathrm{d}h}\mathrm{e}^{\tilde{\varphi}(h)} - \sum_{j=s+1}^{\infty} \frac{h^j}{j!}\tilde{m}_{j+1} + m_3\tilde{r}(h).$$

In view of $\tilde{\varphi}'(h) = x$, the first term is cancelled by $-x$ in (7.20), we find

$$\int_{\mathbb{R}} y \, \mathrm{d}\mu_h(y) = \tilde{\sigma}_h^{-1}\mathrm{e}^{-\tilde{\varphi}(h)}\Big(-\sum_{j=s+1}^{\infty} \frac{h^j}{j!}\tilde{m}_{j+1} + m_3\tilde{r}(h)\Big) = O(\delta^{s+1}), \tag{7.21}$$

where we have used (7.8) and $h \leq \delta a$. For the second moment, we need to evaluate

$$\int_{\mathbb{R}} y^2 \, \mathrm{d}|\mu_h|(y) = \mathrm{e}^{-\tilde{\varphi}(h)}\mathbb{E}\Big[|g_h(X)|\Big(\frac{X-x}{\tilde{\sigma}_h}\Big)^2\Big].$$

Since s is even, the truncated exponential is positive by Lemma 4.4 and the triangle inequality yields

$$\exp_s(u) - u^2|\tilde{r}(h)| \leq |g_h(u)| \leq \exp_s(u) + u^2|\tilde{r}(h)| \qquad (u \in \mathbb{R})$$

hence

$$J_1 - J_2 \leq \int_{\mathbb{R}} y^2 \, \mathrm{d}|\mu_h|(y) \leq J_1 + J_2 \tag{7.22}$$

with

$$J_1 := \frac{\exp(-\tilde{\varphi}(h))}{\tilde{\sigma}_h^2} \mathbb{E}\Big[\exp_s(hX)(X-x)^2\Big],$$

$$J_2 := \frac{\exp(-\tilde{\varphi}(h))}{\tilde{\sigma}_h^2} \mathbb{E}\Big[X^2|\tilde{r}(h)|\,(X-x)^2\Big].$$

We evaluate

$$J_1 = \tilde{\sigma}_h^{-2}\mathrm{e}^{-\tilde{\varphi}(h)}\mathbb{E}\Big[\exp_s(hX)(X^2 - 2xX + x^2)\Big]$$

$$= \tilde{\sigma}_h^{-2}\mathrm{e}^{-\tilde{\varphi}(h)}\sum_{k=0}^{s}\frac{h^k}{k!}(m_{k+2} - 2xm_{k+1} + x^2 m_k)$$

$$= \tilde{\sigma}_h^{-2}\mathrm{e}^{-\tilde{\varphi}(h)}\Big\{\Big((\mathrm{e}^{\tilde{\varphi}})''(h) - 2x(\mathrm{e}^{\tilde{\varphi}})'(h) + x^2 \mathrm{e}^{\tilde{\varphi}(h)}\Big)$$

$$- \sum_{k=s+1}^{\infty}\frac{h^k}{k!}\big(\tilde{m}_{k+2} - 2x\tilde{m}_{k+1} + x^2\tilde{m}_k\big)\Big\}.$$

The terms involving $\tilde{\varphi}(h)$ and its derivatives combine to

$$\tilde{\sigma}_h^{-2}\mathrm{e}^{-\tilde{\varphi}(h)}\Big((\mathrm{e}^{\tilde{\varphi}})''(h) - 2x(\mathrm{e}^{\tilde{\varphi}})'(h) + x^2\mathrm{e}^{\tilde{\varphi}(h)}\Big)$$

$$= \frac{1}{\tilde{\sigma}_h^2}\big(\tilde{\varphi}''(h) + \tilde{\varphi}'(h)^2 - 2x\tilde{\varphi}'(h) + x^2\big) = 1$$

(remember $x = \tilde{\varphi}'(h)$ and $\tilde{\sigma}_h^2 = \tilde{\varphi}''(h)$), consequently

$$J_1 = 1 - \tilde{\sigma}_h^{-2}\mathrm{e}^{-\tilde{\varphi}(h)}\sum_{k=s+1}^{\infty}\frac{h^k}{k!}\big(\tilde{m}_{k+2} - 2x\tilde{m}_{k+1} + x^2\tilde{m}_k\big) = 1 + O(\delta^{s+1}). \quad (7.23)$$

For J_2, we evaluate by Lemma 7.5

$$J_2 = \tilde{\sigma}_h^{-2}\mathrm{e}^{-\tilde{\varphi}(h)}|\tilde{r}(h)|(m_4 - 2xm_3 + x^2) = O(\delta^{s+1}). \quad (7.24)$$

It follows that $\int_{\mathbb{R}} y^2 \mathrm{d}|\mu_h|(y) = 1 + O(\delta^{s+1})$.

Next we bound the constants in front of δ^{s+1}. By Lemmas 7.3 and 7.4, we have $\tilde{\sigma}_h \geq 1 - 0.751 = 0.249$ and $\tilde{\varphi}(h) \geq 0$. Consequently

$$0 < \tilde{\sigma}_h^{-1}\exp(-\tilde{\varphi}(h)) \leq 2.0041, \quad 0 < \tilde{\sigma}_h^{-2}\exp(-\tilde{\varphi}(h)) \leq 4.017. \quad (7.25)$$

Moreover, we can estimate the third and fourth moment by

$$|m_3| = |\kappa_3| \leq \frac{1}{\Delta} \leq \frac{1}{a\sqrt{2\mathrm{e}}}$$

by the cumulant bound and $\Delta^2 \geq \frac{s}{2} = 2\mathrm{e}\,a^2$ and

$$|m_4| = |\kappa_4 + 3\kappa_2^2| \leq \frac{2}{\Delta^2} + 3 \leq 3 + \frac{2}{15}$$

for $\Delta^2 \geq \frac{s}{2} \geq 15$. For $\tilde{r}(h)$ and the power series in h, we use (7.8) and $h \leq \delta a$, set $u = h/(\sqrt{2}a)$ and obtain

$$|\tilde{r}(h)| \leq \frac{u^{s+1}}{1-u},$$

$$\sum_{k=s+1}^{\infty} \frac{h^k}{k!}|\tilde{m}_{k+1}| \leq \frac{1}{\sqrt{2}a}\frac{\mathrm{d}}{\mathrm{d}u}\sum_{k=s+2}^{\infty} u^k = \frac{u^{s+1}}{\sqrt{2}a}\Big(\frac{(s+2)}{1-u} + \frac{u}{(1-u)^2}\Big),$$

$$\sum_{k=s+1}^{\infty} \frac{h^k}{k!}|\tilde{m}_{k+2}| \leq \frac{1}{2a^2}\frac{\mathrm{d}^2}{\mathrm{d}u^2}\sum_{k=s+3}^{\infty} u^k$$

$$= \frac{u^{s+1}}{2a^2}\Big(\frac{(s+3)(s+2)}{1-u} + 2\frac{(s+3)u}{(1-u)^2} + 2\frac{u^2}{(1-u)^3}\Big).$$

We use the notation

$$U_j := \sup_{u \in [0,\frac{1}{\sqrt{2}}]} \frac{u^{j-1}}{(1-u)^j} = \begin{cases} 2 + \sqrt{2}, & j = 1 \\ 4 + 3\sqrt{2}, & j = 2 \\ 10 + 7\sqrt{2}, & j = 3 \end{cases},$$

where the latter representation follows from monotonicity. With this notation we have, in view of (7.21), (7.24), (7.23) together with $u^{s+1} \leq (\delta/\sqrt{2})^{s+1}$, (7.25) and $x \leq \frac{2}{3}a = \frac{2}{3}\sqrt{\frac{s}{4e}} \leq \frac{\sqrt{2}}{3\sqrt{e}}\Delta$ resp. $x^2 \leq \frac{s}{9e}$

$$\Big|\int_{\mathbb{R}} y\,\mathrm{d}\mu_h(y)\Big| \leq 2.0041(\delta/\sqrt{2})^{s+1}\Big(\frac{1}{\sqrt{2}a}((s+2)U_1 + U_2) + \frac{1}{a\sqrt{2e}}U_1\Big),$$

$$J_2 \leq 4.017(\delta/\sqrt{2})^{s+1}U_1(3 + \frac{2}{15} + 2x\frac{1}{\Delta} + x^2),$$

$$\leq 4.017(\delta/\sqrt{2})^{s+1}U_1(3 + \frac{2}{15} + \frac{2\sqrt{2}}{3\sqrt{e}} + \frac{s}{9e})$$

$$|J_1 - 1| \leq 4.017(\delta/\sqrt{2})^{s+1}\Big(\frac{1}{2a^2}(U_1(s+3)(s+2) + 2(s+3)U_2 + 2U_3)$$

$$+ \frac{4}{3\sqrt{2}}((s+2)U_1 + U_2) + \frac{s}{9e}U_1\Big).$$

From $a = \sqrt{\frac{s}{4e}} \geq \sqrt{\frac{30}{4e}}$ together with the representations of $U_j, j = 1,2,3$ and estimate of the form $\frac{s+2}{\sqrt{2}^{s+1}} \leq \frac{32}{\sqrt{2}^{31}}$ resp. $\frac{(s+3)(s+2)}{\sqrt{2}^{s+1}} \leq \frac{33\cdot32}{\sqrt{2}^{31}}$ we immediately get

$$\Big|\int_{\mathbb{R}} y\,\mathrm{d}\mu_h(y)\Big| \leq \delta^{s+1}0.0023$$

$$J_2 \leq \delta^{s+1}0.0015$$

$$|J_1 - 1| \leq \delta^{s+1}0.07577$$

and hence

$$\Big|\int_{\mathbb{R}} x^2\mathrm{d}|\mu_h|(x) - 1\Big| \leq |J_1 - 1| + |J_2| \leq 0.07724\delta^{s+1}. \qquad \square$$

Lemma 7.11. *Under the assumptions of Lemma 7.10, we have*

$$\left|\chi_h(t) - \exp(-\tfrac{1}{2}t^2)\right| \leq \delta^{s+1}\, 0.0023\, |t| + (1 + 0.0387\, \delta^{s+1})t^2$$

for all $t \in \mathbb{R}$.

Proof. We combine Eq. (7.19), Lemma 7.10 and $|\exp(-\tfrac{1}{2}t^2) - 1| \leq \tfrac{1}{2}t^2$

$$\left|\chi_h(t) - \exp(-\tfrac{1}{2}t^2)\right| = \left|1 + \theta t\, 0.0023\, \delta^{s+1} + \theta\frac{t^2}{2}\left(1 + \theta\, 0.0773\, \delta^{s+1}\right) - \mathrm{e}^{-t^2/2}\right|$$

$$\leq 0.0023\, \delta^{s+1}|t| + \frac{t^2}{2}(1 + 0.0773\, \delta^{s+1}) + \frac{t^2}{2}. \qquad \square$$

Combining Lemmas 7.9 and 7.11, we get

$$\left|\chi_h(t) - \exp(-\tfrac{1}{2}t^2)\right| \leq \min\Big(\frac{t}{T}\big(\mathrm{e}^{-t^2/4} - \mathrm{e}^{-t^2/2}\big)$$

$$+ 4\sqrt{2}\,\frac{\delta_2^{\lfloor 4a\rfloor}}{1 - \delta_2},\, \delta^{s+1}\, 0.0023\, |t| + (1 + 0.0387\, \delta^{s+1})t^2. \quad (7.26)$$

Notice that this upper bound is integrable against $\mathrm{d}t/t$. The bound is valid if the assumptions of both lemmas are met. For future reference we summarize these assumptions, see Figure 1.

Assumption 7.12. The quantities $s \in \mathbb{N}$, $x \geq 0$, $\delta, \delta_2 \in [0,1]$ satisfy the following:

- $30 \leq s \leq 2\Delta^2$ and s is even. We set $a := \sqrt{\frac{s}{4\mathrm{e}}}$.
- $x \in [0, \frac{2}{3}\delta a]$.
- $0 < \delta < \delta_2 < 1$ and $a^{-1} \leq \delta_2^2$. We set $T := (\delta_2 - \delta)a\tilde{\sigma}_h$, with $h \in [0, \delta a]$ the solution of $\tilde{\varphi}'(h) = x$.
- Fourier transforms are evaluated at $t \in [-T, T]$.

7.6. Normal approximation for the tilted measure

Armed with the bound (7.26) and Lemma 4.1 we can bound the Kolmogorov distance between the normal law and the tilted signed measure μ_h. Set

$$D_h := \sup_{y \in \mathbb{R}}\left|\mu_h\big((-\infty, y]\big) - \mathbb{P}(Z \leq y)\right|.$$

Lemma 7.13. *Fix $\delta \in (0, 1)$. Then, there exists $C(\delta) > 0$ and $s_\delta \in \mathbb{N}$ such that for all $h \in (0, \delta a)$ and all even $s \geq s_\delta$ with $s \leq 2\Delta^2$, we have*

$$D_h \leq \frac{C(\delta)}{\sqrt{s}}. \qquad (7.27)$$

Proof. Fix $\delta_2 \in (\delta, 1)$. Clearly for sufficiently large s, the condition $\delta_2^2 > a^{-1}$ in Assumption 7.12 is satisfied. We apply Lemma 4.1 to $\nu = \mu_h$ (the signed tilted measure), $\mu = \mathcal{N}(0,1)$ the standard normal distribution, $q = 1/\sqrt{2\pi}$, and $\varepsilon = 1/T$ with T of the order of $\sqrt{s}$, which gives

$$D_h \leq \frac{C_1}{\sqrt{2\pi}} \frac{1}{T} + \frac{C_2}{\pi} \int_0^T \Big(1 - \frac{t}{T}\Big) \big|\chi_h(t) - \mathrm{e}^{-t^2/2}\big| \frac{\mathrm{d}t}{t} + C_3 \mu_h^-(\mathbb{R}) \qquad (7.28)$$

with C_1 and C_2 the numerical constants defined in Eq. (4.3) and

$$C_3 = \frac{C_1}{3.55} \approx 1.282.$$

The third term in (7.28) is missing in [101] because of an erroneous application of Zolotarev's lemma, see the remark after Lemma 4.1.

For T of the order of $\sqrt{s}$, clearly the first term on the right-hand side of (7.28) is of the order of $1/\sqrt{s}$. The measure μ_h was defined in (7.4), its negative part is given by

$$\mu_h^-(\mathbb{R}) = \mathrm{e}^{-\tilde{\varphi}(h)} \mathbb{E}\big[(g_h)_-(X)\big] \leq \mathrm{e}^{-\tilde{\varphi}(h)} \mathbb{E}\big[X^2 \tilde{r}(h)\big] = \mathrm{e}^{-\tilde{\varphi}(h)} \tilde{r}(h),$$

where we used the positivity of the truncated exponential shown in Lemma 4.4. Lemma 7.4 and 7.5 give

$$\mu_h^-(\mathbb{R}) \leq \frac{(\delta/\sqrt{2})^{s+1}}{1 - 1/\sqrt{2}} = O\Big(\Big(\frac{1}{\sqrt{2}}\Big)^s\Big), \qquad (7.29)$$

which decays exponentially fast as $s \to \infty$ and is negligible compared to $1/\sqrt{s}$.

In the integrand in the second term in (7.28) we bound $(1 - |t|/T) \leq 1$, split the domain of integration into two subintervals $[0, t_0]$ and $[t_0, T]$ for some $t_0 > 0$, and apply (7.26). This yields two contributions. The first one is bounded by

$$\int_0^{t_0} \big|\chi_h(t) - \mathrm{e}^{-t^2/2}\big| \frac{\mathrm{d}t}{t} \leq \int_0^{t_0} \Big(O(\delta^{s+1}) + (1 + O(\delta^{s+1}))t\Big) \mathrm{d}t$$

$$= t_0 O(\delta^{s+1}) + \frac{t_0^2}{2}(1 + O(\delta^{s+1})).$$

Using Lemma 7.9 we find that the second contribution is bounded by

$$\int_{t_0}^T \big|\chi_h(t) - \mathrm{e}^{-t^2/2}\big| \frac{\mathrm{d}t}{t} \leq \frac{2}{T} \int_{t_0}^\infty \mathrm{e}^{-t^2/4} \mathrm{d}t + \int_{t_0}^T O(\delta_2^{\lfloor 4a \rfloor}) \frac{\mathrm{d}t}{t} \leq \frac{C}{T} + O(\delta_2^{\lfloor 4a \rfloor}) \log \frac{T}{t_0}.$$

Choose $t_0 = \delta_2^{\lfloor 4a \rfloor/2}$, then we get in the limit $s \to \infty$ at fixed δ_2

$$\int_0^T \big|\chi_h(t) - \mathrm{e}^{-t^2/2}\big| \frac{\mathrm{d}t}{t} = O\Big(\frac{1}{\sqrt{s}}\Big) + O\big(\delta_2^{\lfloor 4a \rfloor}\big(s + \log \delta_2 + \log s\big)\big) = O\Big(\frac{1}{\sqrt{s}}\Big).$$
$$(7.30)$$

We combine Eqs. (7.28), (7.29), (7.30) and arrive at $D_h = O(1/\sqrt{s})$. Precise bounds depend on δ and δ_2 but not on the exact value of $h \in (0, \delta)$ or $x = \tilde{\varphi}'(h)$ and the proof is complete. $\qquad \square$

Lemma 7.13 is all that is needed for the proof of Theorem 2.2. For the reader interested in numerical values, we add two further bounds, following [101, Equations (2.48) and (2.55)]. Set

$$t_0^2 := \frac{4\sqrt{2}\delta_2^{\lfloor 4a \rfloor}}{1 - \delta_2}. \tag{7.31}$$

The main idea for the following lemma is then to split the domain of integration $[0, T]$ into the intervals $[0, t_0]$ and $[t_0, T]$.

Lemma 7.14. *Under Assumption 7.12, we have*

$$\int_0^T \big|\chi_h(t) - e^{-t^2/2}\big|\frac{dt}{t} \leq \frac{\sqrt{2\pi}}{T}\frac{\sqrt{2} - 1}{2} + t_0^2\Big(0.5203 + \max(\log\frac{T}{t_0}, 0)\Big). \tag{7.32}$$

Proof. We refine the bounds from the proof of Lemma 7.13 as in [101, Eq. (2.46)]. The fraction with δ_2, which we denote by t_0 according to (7.31), appeared in the error bound on Fourier transforms in Lemma 7.8. The bound (7.26) yields

$$\int_0^{t_0} \big|\chi_h(t) - e^{-t^2/2}\big|\frac{dt}{t} \leq 0.0023\,\delta^{s+1}t_0 + (1 + 0.0387\,\delta^{s+1})\frac{t_0^2}{2}. \tag{7.33}$$

and, if $T \geq t_0$,

$$\int_{t_0}^T \big|\chi_h(t) - e^{-t^2/2}\big|\frac{dt}{t} \leq \frac{1}{T}\int_{t_0}^T (e^{-t^2/4} - e^{-t^2/2})\,dt + t_0^2\int_{t_0}^T \frac{dt}{t}$$

$$\leq \frac{\sqrt{2\pi}}{T}\frac{\sqrt{2} - 1}{2} + t_0^2\log\frac{T}{t_0}. \tag{7.34}$$

To that aim we note, first, that, because of $\delta < \delta_2 < 1$ and $4a \leq 2\sqrt{s} \leq s$, we have

$$\frac{\delta^{s+1}}{t_0} \leq \delta_2^{s+1-\lfloor 4a \rfloor}\sqrt{\frac{1 - \delta_2}{4\sqrt{2}}} \leq 2^{-5/4}$$

Therefore (7.33) is bounded by

$$t_0^2\big(0.0023 \times 2^{-5/4} + \frac{1.0387}{2}\big) \leq t_0^2\,0.5203. \tag{7.35}$$

If $T \geq t_0$, the inequality (7.32) follows by adding up (7.33) and (7.34) and combining with (7.35). If $T \leq t_0$, then the integral from zero to T is bounded by the integral from zero to t_0 and (7.33) holds true because the right-hand side of (7.32) is larger than the right-hand side of (7.33). We insert the bound into (7.28), combine with (7.29), and obtain the lemma. □

Lemma 7.15. *Let s be an even integer with $30 \leq s \leq 2\Delta^2$. Fix $\delta \in [0, 1)$ and $x \in [0, \frac{2}{3}\delta a]$ and define $h \in [0, \delta a]$ by $\tilde{\varphi}'(h) = x$. Define*

$$\delta_2 := 1 - \frac{1 - \delta}{2s^{1/4}}.$$

Then Assumption 7.12 is satisfied and

$$D_h \leq \frac{22.52}{(1-\delta)\sqrt{s}}\left(0.8 + s\exp\left(-\frac{1}{2}(1-\delta)s^{1/4}\right)\right).$$

Proof. The only conditions to be checked in Assumption 7.12 are $\delta_2^2 \geq a^{-1}$ and $0 < \delta < \delta_2 < 1$. The second inequality follows from

$$0 < 1 - \delta_2 = \frac{1-\delta}{2s^{1/4}} < 1 - \delta.$$

The first inequality holds true because

$$\sqrt{a}\,\delta_2 \geq \left(\frac{s}{4\mathrm{e}}\right)^{1/4}\left(1 - \frac{1}{2s^{1/4}}\right) = (4\mathrm{e})^{-1/4}\left(s^{1/4} - 0.5\right).$$

For $s = 30$, the last expression is approximately equal to 1.0135 hence in particular larger than 1, which by monotonicity extends to all of $s \geq 30$. Thus Assumption 7.12 holds true and we may apply the bounds from Lemma 7.13 and Lemma 7.14.

We want to extract from (7.28) and (7.32) a bound expressed directly in terms of s, δ, and numerical constants. In order to bound $1/T$ with $T = (\delta_2 - \delta)a\tilde{\sigma}_h$, we remember the bound $\tilde{\sigma}_h^2 = \tilde{\varphi}''(h) \geq 1 - 0.751$ from Lemma 7.3 and bound

$$T \geq \sqrt{0.249}\sqrt{\frac{s}{4\mathrm{e}}}\left(1 - \frac{1}{2s^{1/4}}\right)(1-\delta) \geq 0.118\,(1-\delta)\sqrt{s} \qquad (7.36)$$

for $s \geq 30$. In order to bound t_0^2 defined in (7.31), we note, first, $4a = 2\sqrt{s/\mathrm{e}} \geq \sqrt{s}$, hence $\lfloor 4a \rfloor \geq \sqrt{s} - 1$ and, second,

$$\delta_2^{\sqrt{s}} \leq \exp\left(-(1-\delta)s^{1/4}/2\right),$$

where we have used the standard inequality $1 - x \leq \exp(-x)$. As a consequence,

$$t_0^2 \leq \frac{4\sqrt{2}}{\delta_2(1-\delta_2)}\delta_2^{\sqrt{s}} \leq \frac{4\sqrt{2}}{1 - 1/(2s^{1/4})}\frac{2s^{1/4}}{1-\delta}\exp\left(-(1-\delta)s^{1/4}/2\right).$$

It remains to take care of $t_0^2 \max(\log(T/t_0), 0)$ in (7.32). We bound $\log(T/t_0)$ by a constant times $s^{1/4}$. In view of Lemma 7.3 and $\delta_2 - \delta \leq 1$ we have

$$\log T \leq \log(a\tilde{\sigma}_h) \leq \frac{1}{2}\log 1.751 + \frac{1}{2}\log\frac{s}{4\mathrm{e}}.$$

It is easily checked that $x^{-1}\log x$ is maximal at $x = \mathrm{e}$ hence $\log x \leq x/\mathrm{e}$ and $\log x = 4\log x^{1/4} \leq 4x^{1/4}/\mathrm{e}$, therefore

$$\log T \leq \frac{1}{2}\log\left(\frac{1.751}{4\mathrm{e}}s\right) \leq \frac{2}{\mathrm{e}}\left(\frac{1.751}{4\mathrm{e}}\right)^{1/4}s^{1/4} \leq 0.47\,s^{1/4}.$$

Furthermore, by the definition of t_0

$$-\log t_0 \leq \frac{1}{2}\Big(-4a\log\delta_2 - \log(4\sqrt{2}) + \log(1-\delta_2)\Big) \leq -2a\log\delta_2 = -\sqrt{\frac{s}{e}}\log\delta_2.$$

It is easily checked that $x(\log x - 1)$ is minimal at $x = 1$, from which one deduces the inequality $\log x \geq 1 - 1/x$. We apply the inequality to $\log\delta_2$, use $(1 - 1/(2s^{1/4}))^{-1} \leq 1.272$ for $s \geq 30$, and find

$$-\log\delta_2 \leq \frac{1-\delta_2}{\delta_2} \leq \frac{1-\delta}{2s^{1/4}}\frac{1}{1-1/(2s^{1/4})} \leq 0.64\,s^{-1/4}.$$

It follows that

$$\log T - \log t_0 \leq \Big(0.47 + \frac{0.64}{\sqrt{e}}\Big)s^{1/4} \leq 0.86\,s^{1/4}$$

and for $s \geq 30$

$$0.5203 + \max\Big(\log\frac{T}{t_0}, 0\Big) \leq 1.0805\,s^{1/4}.$$

We insert this bound into (7.32), combine with (7.36) and (7.28) for C_1 and C_2 defined by (4.3), and find that $D_h \leq A + B + C$ with

$$A = \frac{1}{T}\Big(\frac{C_1}{\sqrt{2\pi}} + \frac{C_2}{\pi}\sqrt{2\pi}\frac{\sqrt{2}-1}{2}\Big)$$
$$\leq \frac{1}{(1-\delta)\sqrt{s}}\frac{1}{0.118}\Big(\frac{4.551}{\sqrt{2\pi}} + \frac{1.564}{\sqrt{2\pi}}(\sqrt{2}-1)\Big)$$
$$\leq \frac{18}{(1-\delta)\sqrt{s}}.$$

and

$$B = \frac{C_2}{\pi}t_0^2 \times 1.0805\,s^{1/4}$$
$$\leq \frac{4.551}{\pi} \times 1.0805 \times \frac{8\sqrt{2}}{1-1/(2\times 30^{1/4})}\frac{\sqrt{s}}{1-\delta}\exp\Big(-\frac{1}{2}(1-\delta)s^{1/4}\Big)$$
$$\leq 22.5193\frac{\sqrt{s}}{1-\delta}\exp\Big(-\frac{1}{2}(1-\delta)s^{1/4}\Big).$$

and finally for $s \geq 30$

$$C = C_3\mu_h^-(\mathbb{R}) \leq 4.38\Big(\frac{\delta}{\sqrt{2}}\Big)^{s+1} \leq 10^{-4}\,\delta^{s+1}.$$

Notice
$$\delta^{s+1} = \delta\big(1-(1-\delta)\big)^s \leq e^{-(1-\delta)s} \leq \sqrt{s}\,e^{-(1-\delta)s^{1/4}/2}.$$

The lemma follows by adding up the bounds for A, B, and C. $\qquad\square$

7.7. Undoing the tilt

The relation (7.6) yields

$$\mathbb{P}(X \geq x) = \mathrm{e}^{\tilde{\varphi}(h)} \int_0^\infty g_h(\tilde{\sigma}_h y + x)^{-1} \mathrm{d}\mu_h(y),$$

which we want to bring into the form

$$\mathbb{P}(X \geq x) = \mathrm{e}^{\tilde{\varphi}(h) - hx} \Big(\mathbb{E}\big[\mathrm{e}^{-h\tilde{\sigma}_h Z} \mathbb{1}_{\{Z \geq 0\}}\big] + \text{error term} \Big).$$

Let

$$I_1(h) := \int_0^\infty g_h(\tilde{\sigma}_h y + x)^{-1} \mathrm{d}\mu_h(y) - \mathbb{E}\Big[g_h(\tilde{\sigma}_h Z + x)^{-1} \mathbb{1}_{\{Z \geq 0\}} \Big],$$
$$I_2(h) := \mathbb{E}\big[g_h(x + \tilde{\sigma}_h Z)^{-1} \mathbb{1}_{\{Z \geq 0\}} \big]$$

so that

$$\mathbb{P}(X \geq x) = \mathrm{e}^{\tilde{\varphi}(h)} \big(I_1(h) + I_2(h) \big). \tag{7.37}$$

Lemma 7.16 below says that $I_1(h)$ is small, Lemma 7.17 says that $I_2(h)$ is approximately equal to $\mathbb{E}[\exp(-hx - h\tilde{\sigma}_h Z) \mathbb{1}_{\{Z \geq 0\}}]$.

Lemma 7.16. *For $s \geq 30$, $h \in [0, \delta a]$ and $x < \sqrt{s}/(3\sqrt{\mathrm{e}})$ let D_h be the Kolmogorov distance between μ_h and the normal distribution $\mathcal{N}(0, 1)$. Then*

$$|I_1(h)| \leq 2.0004 \, D_h \exp(-hx).$$

Proof. We apply Lemma 4.3 to $f(y) = 1/g_h(\tilde{\sigma}_h y + x)$, which is monotone decreasing by Lemma 7.6, and get

$$|I_1(h)| \leq 2g_h(x)^{-1} \sup_{y \in \mathbb{R}} \big| \mu_h\big((-\infty, y]\big) - \mathbb{P}(Z \leq y) \big| = 2g_h(x)^{-1} D_h.$$

We have for all $u \geq 0$ and suitable $\theta = \theta_{h,u} \in [-1, 1]$

$$g_h(x) = \exp_s(hx) + \frac{1}{2}(hx)^2 \frac{2\tilde{r}(h)}{h^2} = \Big(1 + \theta \frac{2\tilde{r}(h)}{h^2} \Big) \exp_s(hx).$$

A Taylor expansion for the exponential shows $\exp(hx) = \exp_s(hx) + \theta \frac{(hx)^{s+1}}{(s+1)!}$ hence

$$\exp_s(hx) = \mathrm{e}^{hx} \Big(1 - \theta \frac{(hx)^{s+1}}{(s+1)!} \Big).$$

Under our assumptions on h and x we have

$$0 \leq hx \leq \delta \sqrt{\frac{s}{4\mathrm{e}}} \frac{\sqrt{s}}{3\sqrt{\mathrm{e}}} = \frac{s}{\mathrm{e}} \frac{\delta}{6}.$$

We estimate with the aid of Stirling

$$\frac{(hx)^{s+1}}{(s+1)!} \leq \frac{(s/\mathrm{e})^{s+1}}{(s+1)!} \Big(\frac{\delta}{6} \Big)^{s+1} \leq \Big(\frac{\delta}{6} \Big)^{s+1}.$$

Altogether, combining also with (7.13) and $s \geq 30$,

$$g_h(x)^{-1} \leq \frac{\exp(-hx)}{(1 - (\delta/\sqrt{2})^{s-5})(1 - (\delta/6)^{s+1})} \leq 1.0002\, \mathrm{e}^{-hx} \tag{7.38}$$

and the lemma follows. □

Lemma 7.17. *Under Assumption 7.12, we have*

$$I_2(h) = \mathrm{e}^{-hx}\left(\frac{\mathbb{E}\big[\exp(-h\tilde{\sigma}_h Z)\mathbb{1}_{\{Z \geq 0\}}\big]}{(1 + \theta(\delta/\sqrt{2})^{s-5})(1 + \theta\delta^{s+1})} + \theta\frac{1.0002}{\sqrt{2\pi}}\,\mathrm{e}^{-s/8}\right).$$

Proof. We split the expectations in two contributions, one belonging to $Z \leq b_1$ for some well-chosen truncation parameter b_1 and another contribution belonging to $Z > b_1$. On the event $Z \leq b_1$, we are going to estimate $g_h(x + \tilde{\sigma}_h Z)^{-1}$ in a way similar to Lemma 7.16 (where we had looked at $g_h(x)^{-1}$ only). The event $Z > b_1$ is shown to give a negligible contribution. Set

$$b_1 := \frac{b - x}{\tilde{\sigma}_h} \geq \frac{5}{3}a.$$

The bound uses $\tilde{\sigma}_h \leq 2$ (see Lemma 7.3), $b = 4a$ and $x \leq \frac{2}{3}a$. For $Z = y \geq b_1$, we note $0 \leq g_h(x + \tilde{\sigma}_h Z)^{-1} \leq g_h(x)^{-1}$ and bound $g_h(x)^{-1}$ as in (7.38). This yields

$$0 \leq \mathbb{E}\big[g_h(x + \tilde{\sigma}_h Z)^{-1}\mathbb{1}_{\{Z \geq b_1\}}\big] \leq 1.0002\,\exp(-hx)\mathbb{P}(Z \geq b_1).$$

Next consider $Z = y \in [0, b_1]$. Then $u := x + \tilde{\sigma}_h y \leq b$, moreover

$$hu \leq hb \leq (\delta a)(4a) = \delta\frac{s}{\mathrm{e}}$$

and

$$\frac{(hu)^{s+1}}{(s+1)!} \leq \frac{(s/\mathrm{e})^{s+1}}{(s+1)!}\delta^{s+1} \leq \delta^{s+1}.$$

Proceeding as in the proof of Lemma 7.16, we see that

$$\frac{\exp(-hu)}{(1 + (\delta/\sqrt{2})^{s-5})(1 + \delta^{s+1})} \leq g_h(u)^{-1} \leq \frac{\exp(-hu)}{(1 - (\delta/\sqrt{2})^{s-5})(1 - \delta^{s+1})}.$$

We substitute $u = x + \tilde{\sigma}_h Z$, take expectations, and find

$$\mathbb{E}\big[g_h(x + \tilde{\sigma}_h Z)^{-1}\mathbb{1}_{\{Z \in [0, b_1]\}}\big] = \frac{\mathbb{E}\big[\exp(-hx - h\tilde{\sigma}_h Z)\mathbb{1}_{\{Z \in [0, b_1]\}}\big]}{(1 + \theta(\delta/\sqrt{2})^{s-5})(1 + \theta\delta^{s+1})}.$$

On the right-hand side, replacing the constraint $Z \in [0, b_1]$ by $Z \geq 0$ we obtain an upper bound. For a lower bound we note

$$\mathbb{E}\big[\exp(-h\tilde{\sigma}_h Z)\mathbb{1}_{\{Z \in [0, b_1]\}}\big] \geq \mathbb{E}\big[\exp(-h\tilde{\sigma}_h Z)\mathbb{1}_{\{Z \geq 0\}}\big] - \mathbb{P}(Z \geq b_1).$$

Combining everything, we get

$$I_2(h) \leq \frac{\mathbb{E}\big[\exp(-hx - h\tilde{\sigma}_h Z)\mathbb{1}_{\{Z \geq 0\}}\big]}{(1 + \theta(\delta/\sqrt{2})^{s-5})(1 + \theta\delta^{s+1})} + 1.0002\,\exp(-hx)\,\mathbb{P}(Z \geq b_1)$$

and

$$I_2(h) \geq \frac{\mathbb{E}\big[\exp(-hx - h\tilde{\sigma}_h Z)\mathbb{1}_{\{Z \geq 0\}}\big]}{(1 + \theta(\delta/\sqrt{2})^{s-5})(1 + \theta\delta^{s+1})} - 1.0002\,\exp(-hx)\mathbb{P}(Z \geq b_1).$$

Finally we estimate

$$\mathbb{P}(Z \geq b_1) \leq \frac{1}{\sqrt{2\pi}} \int_{b_1}^{\infty} \frac{y}{b_1} \mathrm{e}^{-y^2/2} \mathrm{d}y \leq \frac{1}{b_1\sqrt{2\pi}} \mathrm{e}^{-b_1^2/2} \leq \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-s/8}.$$

In the last inequality we have used $b_1 \geq 1$ and $4b_1^2 \geq \frac{100}{9}a^2 = \frac{100}{9}\frac{s}{4\mathrm{e}} \simeq 1.02\,s \geq\ s$. $\qquad\square$

7.8. Cramér-Petrov series

The function $\tilde{L}(x)$ defined in (2.7) is for $h = t(x)$ equal to

$$\tilde{L}(x) = \check{\varphi}(h) - hx + \frac{1}{2}x^2.$$

We can now estimate the absolute value of $\tilde{L}$.

Lemma 7.18. *Suppose* $0 \leq x < \sqrt{s}/(3\sqrt{\mathrm{e}})$ *with* $s \leq 2\Delta^2$. *Then* $|\tilde{L}(x)| \leq 1.2\,x^3/\Delta$.

We emphasize that the lemma does not need the assumption $s \geq 30$ or s even.

Proof. To conclude, we follow [101, p. 31] to bound $\tilde{L}(x)$, starting from

$$\tilde{L}(x) = \frac{1}{2}h^2 + \sum_{j=3}^{s} \frac{\kappa_j}{j!}h^j - hx + \frac{1}{2}x^2 = \frac{1}{2}(h - x)^2 + \sum_{j=3}^{s} \frac{\kappa_j}{j!}h^j,$$

which gives

$$|\tilde{L}(x)| \leq x^2\Big(\frac{1}{2}\Big(\frac{h}{x} - 1\Big)^2 + \frac{h^2}{x^2}\sum_{j=3}^{s} \frac{h^{j-2}}{j(j-1)\Delta^{j-2}}\Big).$$

Set $s' := 2\Delta^2$, $a' := \sqrt{s'/(4\mathrm{e})}$, and $\delta' := x/(\frac{2}{3}a')$. Because of $\frac{2}{3}a' \geq \frac{2}{3}a = \sqrt{s}/(3\sqrt{\mathrm{e}})$, we have $\delta' < 1$. It is easily checked that Lemma 7.3 also holds true for primed quantities. In particular, the solution h of $\check{\varphi}'(h) = x$ satisfies $h \leq \delta'a' = \delta'\Delta/\sqrt{2\mathrm{e}}$. It follows that

$$\sum_{j=3}^{s} \frac{h^{j-2}}{j(j-1)\Delta^{j-2}} \leq \delta' \sum_{j=3}^{\infty} \frac{(1/\sqrt{2\mathrm{e}})^{j-2}}{j(j-1)} \leq 0.0924\,\delta',$$

where we used a numerical evaluation of the series. By Lemma 7.3 applied to primed variables, $x = h(1 + \theta\delta'/3)$, therefore $(h/x)^2 \leq 9/4$ and

$$\left(\frac{h}{x} - 1\right)^2 \leq \left(\frac{\delta'/3}{1 - \delta'/3}\right)^2 \leq \delta'^2 \left(\frac{1/3}{1 - 1/3}\right)^2 \leq \frac{\delta'}{4}.$$

Altogether

$$|\tilde{L}(x)| \leq Cx^2\delta' = Cx^2 \frac{x}{2\Delta/(3\sqrt{2e})} \leq 1.17\, x^2 \frac{x}{\Delta}$$

with $C = 0.3329$. $\qquad\square$

Lemma 7.19. *Suppose $0 \leq x < \Delta/(3\sqrt{e})$. Then $\tilde{L}(x) = \sum_{j=3}^{\infty} \tilde{\lambda}_j x^j$ with $\tilde{\lambda}_j$ the coefficients given after Definition A.3. The series is absolutely convergent in $|x| \leq 0.3\,\Delta$.*

Proof. Let $\tilde{b}_k$ be the coefficients from Eq. (A.7). By Proposition A.5, the series

$$H(x) = \sum_{k=1}^{\infty} \tilde{b}_k x^k$$

has radius of convergence larger or equal to $0.3\,\Delta$. Because of

$$\frac{2}{3}a = \frac{2}{3}\sqrt{\frac{s}{4e}} \leq \frac{\sqrt{2}}{3\sqrt{e}}\Delta < 0.3\,\Delta,$$

it follows that $H(x)$ is absolutely convergent for all $x \in [0, \frac{2}{3}a]$. The definition of the coefficient $\tilde{b}_k$ ensures that $\tilde{\varphi}'(H(x)) = x$ whenever $H(x)$ is convergent. Since $H(x) \to 0$ when $x \to 0$, Lemma 7.3 guarantees that $H(x) = h(x)$ for all sufficiently small x. Because of $h = h(x) \in [0, \delta a]$ and $\tilde{\varphi}''(h) \neq 0$ on $[0, a]$ (again by Lemma 7.3), the holomorphic inverse function theorem shows that $h(x)$ is analytic in $[0, \frac{2}{3}a)$. Therefore the equality $h(x) = H(x)$ extends to all of $[0, \frac{2}{3}a)$. The series representation for $\tilde{L}(x)$ then follows from the considerations in Appendix A. $\qquad\square$

7.9. Theorem 2.2—Conclusion of the proof

In the evaluation of $\mathbb{E}[\exp(-h\tilde{\sigma}_h Z)\mathbb{1}_{\{Z \geq 0\}}]$ appearing in the approximation for $I_2(h)$ from Lemma 7.17 we would like to replace $h\tilde{\sigma}_h$ by x, and control the error with Lemma 4.2. To that aim we first compare $h\tilde{\sigma}_h$ with x.

Lemma 7.20. *Let $\delta \in [0, 1]$, $x = \frac{2}{3}\delta a$, and $h \in [0, \delta a]$ the solution of $\tilde{\varphi}'(h) = x$. Then*

$$\mathbb{E}\left[e^{-h\tilde{\sigma}_h Z}\mathbb{1}_{\{Z \geq 0\}}\right] = \frac{1}{1 + \theta\, 0.86\,\delta}\, e^{x^2/2}\mathbb{P}(Z \geq x).$$

Proof. We show first

$$\left|\frac{h\tilde{\sigma}_h}{x} - 1\right| \leq 0.86\delta. \tag{7.39}$$

This bound replaces [101, Eq. (2.53)] which looks incorrect. We start as in [101] and note

$$h\tilde{\sigma}_h^2 - x = h\tilde{\varphi}''(h) - \tilde{\varphi}'(h)$$
$$= \sum_{j=2}^{s}\Big(\frac{1}{(j-2)!} - \frac{1}{(j-1)!}\Big)\kappa_j h^{j-1} = \theta\sum_{j=3}^{s}\frac{j-2}{j-1}h\Big(\frac{h}{\Delta}\Big)^{j-2}.$$

The sum is non-negative and can be bounded from above, with $q = 1/\sqrt{2e}$ and $h \le \delta a \le \delta q\Delta$, by h multiplied with

$$\sum_{j=3}^{s}\frac{j-2}{j-1}(q\delta)^{j-2} \le q\delta\Big(\frac{1}{2} + \sum_{k=1}^{s-4}(q\delta)^k\Big) \le q\delta\Big(\frac{1}{2} + \frac{q}{1-q}\Big) \le 0.537\,\delta.$$

Therefore we have $h\tilde{\sigma}_h^2 - x = \theta 0.537\delta h$. Combining with the bound $x = h(1+\theta\frac{\delta}{3})$ from Lemma 7.3, we get

$$\frac{h\tilde{\sigma}_h^2}{x} = 1 + \theta\delta h\frac{0.537}{h(1+\theta\delta/3)} = 1 + \theta\,\delta\frac{0.537}{1-1/3} = 1 + \theta\,0.8055\,\delta.$$

Eq. (2.53) in [101] is the same but with $\tilde{\sigma}_h$ instead of $\tilde{\sigma}_h^2$, which looks like a typo. We have to work a bit more to get rid of the square of the tilted variance. We deduce with the help of the inequality $\sqrt{1+u} \le 1 + \frac{1}{2}u$ that

$$h\tilde{\sigma}_h = \sqrt{hx(1+\theta 0.8055\delta)} = x\sqrt{\frac{1+\theta 0.8055\delta}{1+\theta\delta/3}}$$
$$\le x\Big(1 + \frac{1}{2}\frac{(0.8055+1/3)\delta}{2/3}\Big) \le x(1 + 0.8542\,\delta).$$

For the lower bound, we exploit the concavity of $u \mapsto \sqrt{1-u}$ on $[0,1]$, which implies that the difference quotient $\frac{1}{u}(\sqrt{1-u}-1)$ is monotone decreasing. Let

$$u = \frac{(0.8055+1/3)\delta}{1+\delta/3}, \quad u^* = \frac{0.8055+1/3}{1+1/3} = 0.8542.$$

Then $u \le u^*$ and

$$\frac{h\tilde{\sigma}_h}{x} - 1 = \sqrt{\frac{1+\theta 0.8055\delta}{1+\theta\delta/3}} - 1 \ge \sqrt{1-u} - 1 \ge \frac{\sqrt{1-u^*}-1}{u^*}u.$$

We notice

$$\frac{1-\sqrt{1-u^*}}{u^*} \le 0.7237, \quad \frac{u}{\delta} \le 0.8055 + 1/3 \le 1.139$$

and conclude

$$\frac{h\tilde{\sigma}_h}{x} - 1 \ge -0.7237 \cdot 1.139\,\delta \ge -0.8243\,\delta.$$

The bound (7.39) for $h\tilde{\sigma}_h/x$ follows. In other words, for $\eta := h\tilde{\sigma}_h - x$ we showed $|\eta/x| \le 0.86\,\delta$. Therefore Lemma 4.2 yields

$$\mathbb{E}\big[\mathrm{e}^{-h\tilde{\sigma}_h Z}\mathbb{1}_{\{Z\ge 0\}}\big] = \frac{x}{x+\theta\eta}\mathbb{E}\big[\mathrm{e}^{-xZ}\mathbb{1}_{\{Z\ge 0\}}\big] = \frac{1}{1+\theta\eta/x}\,\mathrm{e}^{x^2/2}\,\mathbb{P}(Z \ge x). \quad \square$$

Now we can turn to the proof of Theorem 2.2. We distinguish the cases $s \ge 30$ and $s \le 29$. The bound on $\tilde{L}(x)$ has already been proven in Lemma 7.18, it holds true for all $s \le 2\Delta^2$.

Proof of Theorem 2.2 when $s \ge 30$. We start from $\mathbb{P}(X \ge x) = \mathrm{e}^{\tilde{\varphi}(h)}(I_1(h) + I_2(h))$, see (7.37). Lemma 7.16 on $I_1(h)$, Lemma 7.15 on the Kolmogorov distance D_h, and the lower bound for $\mathbb{P}(Z \ge x)$ from Lemma 4.2 yield

$$\begin{aligned}
|I_1(h)| &\le \mathrm{e}^{-hx}\frac{45.05}{\sqrt{s}}\Big(0.8 + s\exp\Big(-\frac{1}{2}(1-\delta)s^{1/4}\Big)\Big) \\
&\le \mathrm{e}^{-hx+x^2/2}\mathbb{P}(Z \ge x)\frac{113\,(x+1)}{\sqrt{s}}\Big(0.8 + s\exp\Big(-\frac{1}{2}(1-\delta)s^{1/4}\Big)\Big).
\end{aligned}$$
(7.40)

Lemmas 7.17 and 7.20 yield

$$I_2(h) = C\,\mathrm{e}^{-hx+x^2/2}\mathbb{P}(Z \ge x) + \theta\,\frac{\exp(-hx)}{\sqrt{2\pi}}\,1.0002\,\mathrm{e}^{-s/8} \qquad (7.41)$$

with

$$C := \Big((1+\theta(\delta/\sqrt{2})^{s-5})(1+\theta\delta^{s+1})(1+\theta\,0.86\delta)\Big)^{-1}.$$

In C, we bound $1+\theta\delta^{s+1} \ge 1-\delta$ and, using $s \ge 30$

$$(1+\theta(\delta/\sqrt{2})^{s-5})(1+\theta\,0.86\,\delta) \ge 1-\delta\Big(0.86+2^{-25/2}\Big) =: 1-c\delta$$

with $c \le 0.8602$. The function $\delta \mapsto (1-c\delta)^{-1}$ is convex on $[0,1]$, therefore

$$\frac{1}{\delta}\Big(\frac{1}{1-c\delta}-1\Big) \le \frac{1}{\delta}\Big(\frac{1}{1-c\delta}-1\Big)\Big|_{\delta=1} = \frac{c}{1-c} \le 6.2.$$

We deduce

$$C \le \frac{1+6.2\,\delta}{1-\delta} = 1 + \frac{7.2\,\delta}{1-\delta}.$$

For a lower bound for C, we use $1/(1+x) \ge 1-x$ which gives

$$\begin{aligned}
C &\ge (1-\delta^{s+1})\big(1-(\delta/\sqrt{2})^{s+1}\big)(1-\delta/6) \\
&\ge 1-\delta^{s+1}-(\delta/\sqrt{2})^{31}-\delta/6-\delta^{63}\frac{1}{\sqrt{2}^{31}}\frac{1}{6} \\
&\ge 1-\delta(1+2^{-31/2}+1/6+\frac{1}{\sqrt{2}^{31}}\frac{1}{6}),
\end{aligned}$$

which is certainly larger than $1 - 7.2\,\delta/(1-\delta)$. Thus we have checked

$$C = 1 + \theta\frac{7.2\,\delta}{1-\delta}.$$

We insert this bound into (7.41), combine with the lower bound $\mathbb{P}(Z \geq x) \geq \exp(-x^2)/(\sqrt{2\pi}(x+1))$ from Lemma 4.2, and obtain

$$I_2(h) = \mathrm{e}^{-hx+x^2/2}\mathbb{P}(Z \geq x)\Big(1 + \theta\,\frac{7.2\,\delta}{1-\delta} + \theta\,1.0002\,(x+1)\mathrm{e}^{-s/8}\Big). \qquad (7.42)$$

Because of $x = \frac{2}{3}\delta a = \delta\sqrt{s}/(3\sqrt{e})$ we have

$$7.2\,\delta = \frac{7.2\,x}{\sqrt{s}/(3\sqrt{e})} \leq 35.4\,\frac{x}{\sqrt{s}}.$$

Moreover, $1.0002\,\sqrt{s}\,\mathrm{e}^{-s/8} \leq 1$ for all $s \geq 30$. It follows that

$$I_2(h) = \mathrm{e}^{-hx+x^2/2}\mathbb{P}(Z \geq x)\Big(1 + \theta\,\frac{36.4\,(x+1)}{(1-\delta)\sqrt{s}}\Big). \qquad (7.43)$$

Finally we add up the estimates (7.40) and (7.43) for $I_1(h)$ and $I_2(h)$, remember Eq. (7.37), and obtain

$$\mathbb{P}(X \geq x) = \mathrm{e}^{\tilde{L}(x)}\mathbb{P}(Z \geq x)\Big(1 + \theta\,\frac{x+1}{(1-\delta)\sqrt{s}}\Big(127 + 113\,s\,\mathrm{e}^{-(1-\delta)s^{1/4}/2}\Big)\Big).$$

Because of $\delta = x/(\sqrt{s}/(3\sqrt{e}))$, this completes the proof of the theorem when s is even and $s \geq 30$. $\qquad\square$

Proof of Theorem 2.2 when $s < 30$. If $s \leq 30$, then $x \leq \sqrt{s}/(3\sqrt{e}) \leq 1.2$. By Lemma 7.18 and $s \leq 2\Delta^2$,

$$|\tilde{L}(x)| \leq 1.2\,x^2\frac{x}{\Delta} \leq 1.2\,x^2\frac{1}{3}\sqrt{\frac{s}{\Delta^2\mathrm{e}}} \leq \frac{1.2^3\sqrt{2}}{3\sqrt{e}} \leq 0.5.$$

Using $\mathbb{P}(Z \geq 1.2) \geq 0.115$, we get

$$\mathrm{e}^{\tilde{L}(x)}\mathbb{P}(Z \geq x)\Big(1 + f(\delta,s)\frac{x+1}{\sqrt{s}}\Big) \geq \mathrm{e}^{-0.5}\,0.115\Big(1 + \frac{127}{\sqrt{30}}\Big)$$
$$\geq 1.6 \geq \mathbb{P}(X \geq x).$$

Furthermore

$$\frac{\mathbb{P}(X \geq x)}{\exp(\tilde{L}(x))\mathbb{P}(Z \geq x)} - 1 \geq -1 \geq -\frac{127}{\sqrt{30}} \geq -f(\delta,s)\frac{x+1}{\sqrt{s}}.$$

This completes the proof of the inequality. $\qquad\square$

8. Bounds for Weibull tails. Proof of Theorem 2.3

Recall that in Theorem 2.3 we assume condition $(\mathcal{S}_\gamma)$, i.e. $|\kappa_j(X)| \leq \frac{j!^{1+\gamma}}{\Delta^{j-2}}$ for $j \geq 3$. The main idea is to use Theorem 2.2 (which uses condition $(\mathcal{S}^*)$), i.e. $|\kappa_j(X)| \leq \frac{(j-2)!}{\Delta^{j-2}}, j \in \{3, \ldots, s+2\}$). However, in general $(\mathcal{S}_\gamma)$ does not imply $(\mathcal{S}^*)$, but this is true if Δ in $(\mathcal{S}^*)$ is replaced by some Δ_s. For this purpose, we set for $s \in \mathbb{N}$

$$\Delta_s := \frac{\Delta}{6(s+2)^\gamma}, \tag{8.1}$$

where we choose s depending on Δ and γ later.

Lemma 8.1. *Let $s \geq 4$. If $|\kappa_j| \leq j!^{1+\gamma}/\Delta^{j-2}$ for all $j \geq 3$ then we have $|\kappa_j| \leq (j-2)!/\Delta_s^{j-2}$ for all $j = 3, \ldots, s+2$.*

Proof. We show $j!^{1+\gamma} \leq (j-2)!(6(s+2)^\gamma)^{j-2}$ for $j = 3, \ldots, s+2$ or equivalently,

$$j(j-1)j!^\gamma \leq (6(s+2)^\gamma)^{j-2} \quad (j = 3, \ldots, s+2).$$

The proof is by a finite induction over j. For $j = 3$, the inequality reads $6 \cdot 6^\gamma \leq 6(s+2)^\gamma$ and it holds true because of $s \geq 4$. For the induction step, we note that if the inequality holds true for $j - 1 \geq 3$ with $j - 1 < s + 2$ then

$$\begin{aligned}
j(j-1)j!^\gamma &= \frac{j}{j-2}j^\gamma \left((j-1)(j-2)(j-1)!^\gamma\right) \\
&\leq 2(s+2)^\gamma \left(6(s+2)^\gamma\right)^{j-3} \leq (6(s+2)^\gamma)^{j-2}. \qquad \square
\end{aligned}$$

Set

$$s = s_\gamma := 2\left\lfloor \frac{1}{2}\left(\frac{\Delta^2}{18}\right)^{1/(1+2\gamma)} \right\rfloor - 2. \tag{8.2}$$

Before we can apply Theorem 2.2 with this s and Δ_s we note some relations between s and Δ_s resp. Δ and (below Lemma 8.2) a relation between Δ_s and Δ_γ, where Δ_γ was given in (2.8) and appears in the formulation of Theorem 2.3. Since we will see that the statement Theorem 2.3 is trivially true for Δ sufficiently small, some of the following estimates are stated for large values of Δ only.

Lemma 8.2. *The even integer s defined in (8.2) satisfies $s \leq 2\Delta_s^2$. If $\Delta \geq 20^{1+2\gamma}$, then in addition $s \geq 4$ and*

$$\sqrt{s} \geq \sqrt{0.82}\left(\frac{\sqrt{2}\Delta}{6}\right)^{1/(1+2\gamma)}. \tag{8.3}$$

In [101] it is claimed that the bound $s \geq 4$ and the inequality (8.3) (with 0.95 instead of $\sqrt{0.82} \approx 0.9$, see [101, Eq. (2.64)]) holds true for $\Delta > 10^{1+2\gamma}$, which looks incorrect.

Proof. We note

$$s(s+2)^{2\gamma} \leq (s+2)^{1+2\gamma} \leq \frac{\Delta^2}{18},$$

which gives $s \leq 2\Delta_s^2$. For the lower bound, if $\Delta \geq 20^{1+2\gamma}$, then $\left(\frac{\Delta^2}{18}\right)^{\frac{1}{1+2\gamma}} \geq \frac{400}{18^{\frac{1}{1+2\gamma}}} \geq \frac{400}{18} \geq 22$ and

$$s \geq \left(\frac{\Delta^2}{18}\right)^{1/(1+2\gamma)} - 4 \geq 22 - 4 \geq 4$$

as well as

$$
\begin{aligned}
s &\geq \left(\frac{\Delta^2}{18}\right)^{1/(1+2\gamma)}\left(1 - 4\left(\frac{18}{\Delta^2}\right)^{1/(1+2\gamma)}\right) \\
&\geq \left(\frac{\Delta^2}{18}\right)^{1/(1+2\gamma)}\left(1 - \frac{4 \cdot 18}{400}\right) = 0.82\left(\frac{\sqrt{2}\Delta}{6}\right)^{2/(1+2\gamma)}. \qquad \square
\end{aligned}
$$

We note that Δ_s, which we introduced in order to use Theorem 2.2 and $\Delta_\gamma := \frac{(\Delta/\sqrt{18})^{1/(1+2\gamma)}}{6}$, which appears in Theorem 2.3, satisfy the following inequalities:

$$\Delta_s \geq \frac{1}{6}\Delta\left(\frac{\Delta^2}{18}\right)^{-\gamma/(1+2\gamma)} = \sqrt{18} \cdot \frac{1}{6}\left(\frac{\Delta}{\sqrt{18}}\right)^{1/(1+2\gamma)}\right) = \sqrt{18}\Delta_\gamma \geq 4\Delta_\gamma. \quad (8.4)$$

We also note that in (8.4) we have $\Delta_s = \sqrt{18}\Delta_\gamma$ if in the definition of s in (8.2) we neglect the floor function. Further, we have

$$
\begin{aligned}
\frac{\Delta_s}{\Delta_\gamma} &= \frac{\Delta}{6(s+2)^\gamma} \frac{1}{\frac{1}{6}(\Delta/\sqrt{18})^{\frac{1}{1+2\gamma}}} \leq \Delta^{1-\frac{1}{1+2\gamma}}\sqrt{18}^{\frac{1}{1+2\gamma}} \frac{1}{2^\gamma\left(\frac{1}{2}\left(\Delta^2/18\right)^{\frac{1}{1+2\gamma}} - 1\right)^\gamma} \\
&= \sqrt{18}\frac{1}{\left(1 - 2\left(\frac{18}{\Delta^2}\right)^{\frac{1}{1+2\gamma}}\right)^\gamma} \leq \sqrt{18}\left(\frac{1}{1 - \frac{9}{100}}\right)^\gamma \leq \sqrt{18}(1.1)^\gamma. \quad (8.5)
\end{aligned}
$$

Together with Lemma 8.2, which reads $\sqrt{s} \geq 6\sqrt{0.82}\Delta_\gamma$, (8.5) implies that for $\Delta \geq 20^{1+2\gamma}$, the quantity Δ_s^2/s^2 is bounded by some constant that does not depend on Δ. We further note that from the definition of s we have

$$s \leq 2(1/2(\Delta^2/18)^{1/(1+2\gamma)}) = 36\Delta_\gamma^2. \quad (8.6)$$

Next, let $\tilde{\lambda}_k$, $k \geq 3$, be the coefficients defined with the help of the truncated functions below Definition A.3, with s given by (8.2). Set

$$\tilde{L}(x) := \sum_{j=3}^{\infty} \tilde{\lambda}_j x^j. \quad (8.7)$$

The function appears naturally when applying Theorem 2.2 in the proof of Theorem 2.3. A first rough estimate for $\tilde{L}(x)$ is obtained as follows. Because of Lemma 8.1 we can apply the bound for $\tilde{\lambda}_j$ from Proposition A.5 (remember $\tilde{\lambda}_k = -\tilde{b}_{k-1}/k$) with Δ_s instead of Δ, which gives

$$|\tilde{\lambda}_j| \leq \frac{1}{0.6\,j(j-1)}\frac{1}{(0.3\,\Delta_s)^{j-2}} \quad \text{for all } j \geq 3. \quad (8.8)$$

Consequently for $|x| \leq 0.3 \, \Delta_s$ we have

$$|\tilde{L}(x)| \leq \frac{x^3}{0.3 \, \Delta_s} \sum_{j=3}^{\infty} \frac{1}{0.6 \, j(j-1)} = \frac{x^3}{0.36 \, \Delta_s} \leq \frac{x^3}{1.54 \, \Delta_\gamma}. \tag{8.9}$$

This bound should be compared to $|L(x)| \leq C'x^3/\Delta$ in Theorem 2.1 and $|\tilde{L}(x)| \leq 1.2 \, x^2/\Delta$ in Theorem 2.2. Notice, however, that x^3/Δ_s can be fairly large. Therefore in some circumstances it can be useful to write $\tilde{L}(x)$ as a polynomial in x, obtained by truncating the Cramér-Petrov series, plus a correction term that is small when x is small compared to Δ_s.

Lemma 8.3. *Suppose that condition* $(\mathcal{S}_\gamma)$ *holds true with* $\Delta > 20^{1+2\gamma}$. *Let* s *be the integer from* (8.2) *and*

$$m := \min\left(\left\lceil \frac{1}{\gamma} \right\rceil + 1, s\right). \tag{8.10}$$

Then

$$\tilde{L}(x) = \sum_{j=3}^{m} \lambda_j x^j + \theta C_\gamma \left(\frac{x}{\Delta_\gamma}\right)^{m+1} \tag{8.11}$$

for some constant $C_\gamma > 0$ *and all* $x \in (0, \Delta_\gamma)$.

Proof. We use $\tilde{\lambda}_k = \lambda_k$ for $k \leq s$ (see (A.8)) and split

$$\sum_{j=m+1}^{\infty} \tilde{\lambda}_j x^j = \sum_{j=m+1}^{s} \lambda_j x^j + \sum_{j=s+1}^{\infty} \tilde{\lambda}_j x^j. \tag{8.12}$$

The first sum on the right-hand side is set to zero if $m = s$. For the second sum we use (8.8) and obtain

$$\sum_{j=s+1}^{\infty} |\tilde{\lambda}_j| x^j \leq \sum_{j=s+1}^{\infty} \frac{1}{j(j-1)} \frac{1}{0.6} \frac{x^j}{(0.3\Delta_s)^{j-2}} = \frac{5x^2}{3} \sum_{j=s+1}^{\infty} \frac{1}{j(j-1)} \left(\frac{x}{0.3 \, \Delta_s}\right)^{j-2}$$

$$\leq \frac{5x^2}{3s^2} \frac{(x/(0.3\Delta_s))^{s-1}}{1 - x/(0.3 \, \Delta_s)} \leq \frac{5(0.3)^2 \Delta_s^2}{3s^2} \frac{(x/(0.3 \, \Delta_s))^{s+1}}{1 - x/(0.3 \, \Delta_s)}$$

$$\leq 0.15 \frac{\Delta_s^2}{s^2} \frac{1}{1 - x/(0.3 \, \Delta_s)} (x/\Delta_\gamma))^{s+1}.$$

Here, we used $\Delta_s \geq 4\Delta_\gamma$ (see (8.4)) for the last inequality. The factor $\frac{1}{1 - x/(0.3 \, \Delta_s)}$ is bounded for $x \in (0, \Delta_\gamma)$ and also Δ_s^2/s^2 is bounded by some constant that does not depend on Δ (see (8.5)). Hence, we get

$$\sum_{j=s+1}^{\infty} |\tilde{\lambda}_j| x^j \leq C'_\gamma \left(\frac{x}{\Delta_\gamma}\right)^{s+1} \tag{8.13}$$

for some constant $C'_\gamma > 0$ and all $x \in (0, \Delta_\gamma)$.

If $m = s$, then (8.11) follows, with $C_\gamma = C'_\gamma$.

If $m < s$, we use the bound from Proposition A.6, which yields

$$
\sum_{j=m+1}^{s} |\lambda_j| x^j \leq x^2 \sum_{j=m+1}^{s} (j+1)!^\gamma \Big(\frac{15x}{\Delta}\Big)^{j-2}
$$

$$
\leq (m+3)!^\gamma\, x^2 (s+2)^{-(m-1)\gamma} \sum_{j=m+1}^{s} (s+2)^{(j-2)\gamma} \Big(\frac{15x}{\Delta}\Big)^{j-2}
$$

$$
\leq (m+3)!^\gamma\, x^2 (s+2)^{-(m-1)\gamma} \frac{(2.5\, x/\Delta_s)^{m-1}}{1 - 2.5\, x/\Delta_s}
$$

$$
\leq \frac{(m+3)!^\gamma}{6.25} \frac{\Delta_s^2}{(s+2)^{(m-1)\gamma}} \frac{(2.5\, x/\Delta_s)^{m+1}}{1 - 2.5\, x/\Delta_s}. \tag{8.14}
$$

Notice that $\Delta_\gamma \leq 0.25\, \Delta_s \leq \Delta_s/2.5$, hence the geometric series in the previous inequality are indeed absolutely convergent for $x \in (0, \Delta_\gamma)$.

By Lemma 8.2, for large s (i.e. large Δ, see (8.2)), s is of the order of Δ_s^2 and if in addition x is of the order of Δ_s (and strictly smaller than $\Delta_s/2.5$) then (8.14) is of the order of

$$
\frac{\Delta_s^2}{\Delta_s^{2\gamma(m-1)}},
$$

which stays bounded because $m \geq 1 + 1/\gamma$. It follows that for some constant $C''_\gamma > 0$ and all $x \in (0, \Delta_\gamma)$,

$$
\sum_{j=m+1}^{s} |\lambda_j| x^j \leq C''_\gamma \Big(\frac{x}{\Delta_\gamma}\Big)^{m+1}.
$$

We combine this bound with (8.13) and obtain (8.11) for $m < s$. The case $m = s$ was already addressed above, the proof is complete. $\qquad\square$

Proof of Theorem 2.3. We start with the proof for $\Delta \geq 20^{1+2\gamma}$. Set $\Delta_\gamma := \frac{1}{6}(\Delta\sqrt{2}/6)^{1/(1+2\gamma)}$ and let Δ_s and s be as in Eqs. (8.1) and (8.2). By Lemma 8.2, for $\Delta \geq 20^{1+2\gamma}$, we have

$$
\Delta_\gamma \leq \frac{\sqrt{s}}{6\sqrt{0.82}} \leq \frac{\sqrt{s}}{3\sqrt{e}}.
$$

For $\delta := x/\Delta_\gamma$ we have $x = \delta\Delta_\gamma \leq \delta\sqrt{s}/(3\sqrt{e})$. Theorem 2.2 is stated for equality $x = \delta\sqrt{s}/(3\sqrt{e})$ but it extends to $x \leq \delta\sqrt{s}/(3\sqrt{e})$ because $f(s, \delta)$ is monotone increasing in δ. Therefore we obtain

$$
\mathbb{P}(X \geq x) = e^{\tilde{L}(x)} \mathbb{P}(Z \geq x)\Big(1 + \theta f(\delta, s)\frac{x+1}{\sqrt{s}}\Big)
$$

$\tilde{L}$ given in (8.7). Then

$$
f(\delta, s) = \frac{1}{1 - \delta}\Big(127 + 113\, s\, e^{-(1-\delta)s^{1/4}/2}\Big)
$$

$$\leq \frac{1}{1-\delta}\Big(127 + 113 \cdot 36\Delta_\gamma^2 \exp\big(-\frac{1}{2}(1-\delta)(3\sqrt{e}\,\Delta_\gamma)^{1/2}\big)\Big)$$

$$\leq \frac{1}{1-\delta}\Big(127 + 113 \cdot 36\Delta_\gamma^2 \exp\big(-(1-\delta)\sqrt{\Delta_\gamma}\big)\Big),$$

where we used (8.6) for the first inequality. Hence we have

$$\frac{1}{\sqrt{s}}f(\delta, s) \leq \frac{1}{\Delta_\gamma}\frac{1}{6\sqrt{0.82}}\frac{1}{1-\delta}\Big(127 + 113 \cdot 36\Delta_\gamma^2 \exp\big(-(1-\delta)\sqrt{\Delta_\gamma}\big)\Big)$$

$$\leq \frac{1}{\Delta_\gamma}\frac{1}{1-\delta}\Big(24 + 749\Delta_\gamma^2 \exp\big(-(1-\delta)\sqrt{\Delta_\gamma}\big)\Big) = \frac{1}{\Delta_\gamma}g(\delta, \Delta_\gamma)$$

Then we obtain for $\Delta \geq 20^{1+2\gamma}$

$$\mathbb{P}(X \geq x) = e^{\tilde{L}(x)}\mathbb{P}(Z \geq x)\Big(1 + \theta g(\delta, \Delta_\gamma)\frac{x+1}{\Delta_\gamma}\Big), \quad x \in [0, \Delta_\gamma). \qquad (8.15)$$

We check that (8.15) is trivial for $\Delta \leq 20^{1+2\gamma}$ as follows. For $0 < \Delta < 20^{1+2\gamma}$ we have

$$0 \leq x < \Delta_\gamma \leq \frac{1}{6}\left(\frac{20^{1+2\gamma}\sqrt{2}}{6}\right)^{\frac{1}{1+2\gamma}} \leq \frac{5\sqrt{2}}{9} \leq 0.8.$$

By (8.9) we have for $x \in [0, \Delta_\gamma)$

$$|L_\gamma(x)| \leq \frac{x^3}{1.54\Delta_\gamma} \leq \frac{x^2}{1.54} \leq 0.42$$

We supplement this with the inequality

$$g(\delta, \Delta_\gamma)\frac{x+1}{\Delta_\gamma} = \frac{1}{1-\delta}\Big(24 + 749\Delta_\gamma^2 \exp\big(-(1-\delta)\sqrt{\Delta_\gamma}\big)\Big)\frac{x+1}{\Delta_\gamma} \geq 24.$$

This implies

$$\big(1 - \Phi(x)\big)e^{L_\gamma(x)}\left(1 + g(\delta, \Delta_\gamma)\frac{x+1}{\Delta_\gamma}\right) \geq 0.79 \cdot 0.65 \cdot 25 \geq 12 \geq 1 \geq 1 - F_X(x)$$

as well as

$$\frac{1 - F_X(x)}{e^{L_\gamma(x)}\big(1 - \Phi(x)\big)} - 1 \geq -1 \geq -24 \geq -g(\delta, \Delta_\gamma)\frac{x+1}{\Delta_\gamma}.$$

This proves that (8.15) is also true for $\Delta \leq 20^{1+2\gamma}$. The statement on $\tilde{L}_\gamma = \tilde{L}$ in the theorem follows from Eq. (8.9) and Lemma 8.3. $\qquad \square$

9. Berry-Esseen bound. Proof of Theorem 2.4

The proof of Theorem 2.4 is similar to the proof of the normal approximation of the tilted measure for the proof of Theorem 2.2, see Section 7.6. The primary

ingredients are a smoothing inequality, e.g. Lemma 4.1, and bounds on the characteristic function of X. The bounds on characteristic functions are similar to the bounds from Sections 7.4 and 7.5, the estimates are slightly easier because we do not need to take tilt parameters into account.

Proof of Theorem 2.4. Replacing the exponentials by their second-order Taylor approximations, we immediately get

$$\left| \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big] - \mathrm{e}^{-t^2/2} \right| \leq t^2$$

for all $t \in \mathbb{R}$, compare Section 7.5. But now suppose first that $\Delta > 20^{1+2\gamma}$. Define Δ_γ, $s := s_\gamma$, and m_γ as in (2.9) and $\Delta_s = \Delta/(6(s+2)^\gamma)$ as in (8.1). In Section 8 we showed that $s \leq 2\Delta_s^2$ and $s \geq 4$, moreover X satisfies condition $(\mathcal{S}^*)$ with Δ_s instead of Δ, i.e., $|\kappa_j| \leq (j-2)!/\Delta_s^{j-2}$ for $j = 3, \dots, s+2$. Let $\tilde{\varphi}(t) = \frac{t^2}{2} + \sum_{j=1}^s \kappa_j t^j/j!$. We split

$$\left| \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big] - \mathrm{e}^{-t^2/2} \right| \leq \left| \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big] - \mathrm{e}^{\tilde{\varphi}(\mathrm{i}t)} \right| + \left| \mathrm{e}^{\tilde{\varphi}(\mathrm{i}t)} - \mathrm{e}^{-t^2/2} \right|. \tag{9.1}$$

For $|t| \leq \Delta_s$, we have

$$\left| \tilde{\varphi}(\mathrm{i}t) + \frac{t^2}{2} \right| \leq t^2 \sum_{j=3}^s \frac{(t/\Delta_s)^{j-2}}{j(j-1)} \leq \frac{t^3}{2\Delta_s}.$$

In the last inequality we have used $\sum_{j=3}^\infty 1/[j(j-1)] = \sum_{j=2}^\infty (1/j - 1/(j+1)) = 1/2$. For the first term on the right-hand side of (9.1), we use a couple of relations from Section 7.1, notably (7.5), involving the truncated moments $\tilde{m}_j$ and the functions $g_t(x) = \exp_s(tx) + x^2\tilde{r}(t)$, which give

$$\left| \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big] - \mathrm{e}^{\tilde{\varphi}(\mathrm{i}t)} \right| = \left| \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX} - g_{\mathrm{i}t}(X)\big] \right| \leq \mathbb{E}\big| \mathrm{e}^{\mathrm{i}tX} - \exp_s(\mathrm{i}tX) \big| + |\tilde{r}(\mathrm{i}t)|,$$

which is similar to (7.15) with $h = 0$ (note $g_h(x) = 1$ when $h = 0$) and added expected values. Let us assume that $s \geq 30$, which is the case when Δ is large enough, so that the bounds from Section 7.4 and 7.5 are applicable. Let $a := \sqrt{s/(4\mathrm{e})}$ and $\delta_2 \in (0,1)$ with $\delta_2^{-2} \leq a$. Then, proceeding as in Lemma 7.8, we obtain the upper bound

$$\left| \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big] - \mathrm{e}^{\tilde{\varphi}(\mathrm{i}t)} \right| \leq 4\sqrt{2}\, \frac{\delta_2^{\lfloor 4a \rfloor}}{1 - \delta_2} \quad (|t| \leq \delta_2 a).$$

We deduce

$$\left| \mathbb{E}\big[\mathrm{e}^{\mathrm{i}tX}\big] - \mathrm{e}^{-t^2/2} \right| \leq 4\sqrt{2}\, \frac{\delta_2^{\lfloor 4a \rfloor}}{1 - \delta_2} + \mathrm{e}^{-t^2/2} \big| \mathrm{e}^{t^3/(2\Delta_s)} - 1 \big| \quad (|t| < \delta_2 a).$$

A reasoning similar to Lemma 7.13 shows that the Kolmogorov distance between the normal law and the law of X is bounded by some constant times $1/\sqrt{s}$ hence also by some constant times $1/\Delta_\gamma$,

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(X \geq x) - \mathbb{P}(Z \geq x) \right| \leq \frac{C}{\Delta_\gamma}.$$

This holds true if Δ is large enough so that $s = s_\gamma$ is larger than 30, say $\Delta \geq \Delta^*$. For smaller Δ, the bound is trivially true if we choose $C \geq 2\sup_{\Delta \leq \Delta^*} \Delta_\gamma$. $\qquad\square$

Appendix A: Cramér-Petrov series

Here we recall some facts on the Cramér-Petrov series, also called Cramér series, for the reader's convenience. The series was introduced by Cramér [19] and appeared in a limit theorem subsequently improved by Petrov [91], see [92, Theorem 5.23] or [64, Chapter 8]. The recurrence relation (A.2) below can be found in [101].

A.1. When Cramér's condition is satisfied

Even though we are primarily interested in heavy-tailed variables, the definition of the Cramér-Petrov series is best understood by looking first at a random variable that satisfies Cramér's condition. Thus let X be a real-valued random variable X such that

$$\mathbb{E}[\exp(tX)] < \infty$$

for all $t \in (-\Delta, \Delta)$, for some $\Delta > 0$. Further assume that X is not almost surely constant so that the variance σ^2 is non-zero. Then the cumulant generating function $\varphi(t) = \log \mathbb{E}[\exp(tX)]$ is analytic in some neighborhood of the origin and the Taylor expansion

$$\varphi(t) = \mu t + \frac{1}{2}\sigma^2 t^2 + \sum_{j=3}^{\infty} \frac{\kappa_j}{j!} t^j$$

has a strictly positive radius of convergence. Recall $I(x) = \sup_{t \in \mathbb{R}}(tx - \varphi(t))$.

Proposition A.1. *Let X be a real-valued random variable that is not almost surely constant. Assume $\mathbb{E}[\exp(tX)] < \infty$ for all $t \in (-\Delta, \Delta)$ for some $\Delta > 0$. Let $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathbb{V}(X)$. Then the Taylor expansion of I at $\mu = \mathbb{E}[X]$ is of the form*

$$I(\mu + \tau) = \frac{\tau^2}{2\sigma^2} - \sum_{j=3}^{\infty} \lambda_j \tau^j \tag{A.1}$$

and has non-zero radius of convergence. The coefficients $(\lambda_j)_{j \geq 3}$ are given by $\lambda_k = -b_{k-1}/k$ with coefficients $(b_k)_{k \geq 2}$ computed recursively as follows: $b_1 = 1/\sigma^2$ and for all $k \geq 2$,

$$b_k = -\frac{1}{\sigma^2} \sum_{r=2}^{k} \frac{\kappa_{r+1}}{r!} \sum_{\substack{1 \leq j_1, \ldots, j_r \leq k-1: \\ j_1 + \cdots + j_r = k}} b_{j_1} \cdots b_{j_r}. \tag{A.2}$$

The coefficients b_k have a significance of their own: for small t,

$$\varphi'(t) = \mu + \tau \iff t = \frac{\tau}{\sigma^2} + \sum_{k=2}^{\infty} b_k \tau^k.$$

Proof. The restriction of the cumulant generating function to $(-\Delta, \Delta)$ is strictly convex and in $C^\infty((-\Delta, \Delta))$, its derivative φ' is a strictly increasing smooth bijection from $(-\Delta, \Delta)$ onto some open interval $(a, b) \subset \mathbb{R}$, and $\varphi''(t) > 0$ on $(-\Delta, \Delta)$. Therefore the inverse map $(\varphi')^{-1} : (a, b) \to (-\Delta, \Delta)$ is smooth as well, i.e., $(\varphi')^{-1} \in C^\infty((a; b))$. By standard facts on Legendre transforms, setting $t_x := (\varphi')^{-1}(x)$, we have for all $x \in (a, b)$

$$I(x) = t_x x - \varphi(t_x), \quad \varphi'(t_x) = x, \quad I'(x) = t_x, \quad I''(x) = \frac{1}{\varphi''(t_x)}.$$

In view of $\varphi'(0) = \mu$ and $\varphi''(0) = \sigma^2$, we have $t_\mu = 0$ and

$$I(\mu) = 0, \quad I'(\mu) = 0, \quad I''(\mu) = \frac{1}{\sigma^2},$$

a well-known property of the Cramér rate function $I(x)$. Consequently the Taylor series of $I(x)$ at $x = \mu$ is of the form (A.1) with

$$\lambda_j = -\frac{1}{j!} \frac{\mathrm{d}^j}{\mathrm{d}x^j} I(x) \Big|_{x=\mu}.$$

From the analyticity of the cumulant generating function $\varphi(t)$ at $t = 0$, the fact $\varphi''(0) = 1 \neq 0$, and the holomorphic inverse function theorem, we know that there exist *complex* open neighborhoods $U, V \subset \mathbb{C}$ of $t = 0$ and μ respectively such that φ' is a bijection from U onto V with holomorphic inverse $(\varphi')^{-1} : V \to U$. In particular, its Taylor series around μ has non-zero radius convergence. Thus

$$t(\tau) := (\varphi')^{-1}(\mu + \tau) = \sum_{k=0}^\infty b_k \tau^k$$

for suitable coefficients $b_k \in \mathbb{R}$, and the series is absolutely convergent for sufficiently small t. The first two terms are easily seen to be $b_0 = 0$ and $b_1 = 1/\sigma^2$, thus

$$t(\tau) = \frac{\tau}{\sigma^2} + \sum_{k=2}^\infty b_k \tau^k. \tag{A.3}$$

Therefore

$$I(\mu + \tau) = \int_0^\tau I'(\mu + u)\mathrm{d}u = \int_0^\tau t(u)\mathrm{d}u = \frac{\tau^2}{2\sigma^2} + \sum_{k=2}^\infty b_k \frac{\tau^{k+1}}{k+1}$$

hence the Taylor series of I and $(\varphi')^{-1}$ around μ have the same radius of convergence, moreover

$$\lambda_k = -\frac{b_{k-1}}{k} \qquad (k \geq 3).$$

The coefficients b_k from (A.3) are computed as follows: we must have

$$\mu + \tau = \varphi'\big(t(\tau)\big) = \mu + \sigma^2 \Big(\sum_{k\geq 1} b_k \tau^k\Big) + \sum_{r=2}^\infty \frac{\kappa_{r+1}}{r!} \Big(\sum_{j=1}^\infty b_j \tau^j\Big)^r$$

for all sufficiently small τ. The left- and right-hand sides are power series in τ and therefore must have all their coefficients equal. The coefficients of order zero and one are equal because $\mu = \mu$ and $1 = b_1 \sigma^2$. For orders $k \geq 2$, we obtain the equation

$$0 = \sigma^2 b_k + \sum_{r=2}^{\infty} \frac{\kappa_{r+1}}{r!} \sum_{\substack{j_1,\ldots,j_r \geq 1: \\ j_1 + \cdots + j_r = k}} b_{j_1} \cdots b_{j_r}.$$

In the second summand, because of $j_\ell \geq 1$, the only relevant contributions come from $r \leq k$ and $1 \leq j_1, \ldots, j_r \leq k-1$, so the recurrence relation for the b_k's follows. $\qquad\square$

An immediate consequence is the following: If $\sigma^2 = 1$, then each coefficient λ_j is a polynomial of the cumulants $\kappa_3, \ldots, \kappa_j$. Explicit formulas for the first few coefficients $b_1, \ldots, b_4$ are given in [101, p. 19], see also [64, Eq. (7.2.20)].

It is instructive to work out an explicit bound on the radius of convergence.

Proposition A.2. *Assume $\mathbb{E}[X] = 0$, $\mathbb{V}[X] = 1$ and $|\kappa_j| \leq (j-2)!/\Delta^{j-2}$ for all $j \geq 3$. Then the radius of convergence of the Cramér-Petrov series is at least $\frac{3}{10}\Delta$.*

Notice $\frac{\sqrt{2}}{3\sqrt{e}} \approx 0.2859 < 0.3$ so the radius of convergence from Proposition A.2 is slightly better than the lower bound $2/(3\sqrt{e})$ to the radius of convergence to the Cramér-Petrov series proven in [101].

Proof. We bound the radius of convergence of the series $\sum_k b_k \tau^k$ with the help of Lagrange inversion, a trick used for the virial expansion in classical statistical mechanics [74]. The radius of convergence of $\sum_j \frac{\kappa_j}{j!} t^j$ is at least Δ (obvious), and for $|t| < \Delta$ we have

$$|\varphi''(t)| \leq \frac{1}{1 - |t|/\Delta}.$$

and

$$\left| \frac{\varphi'(t)}{t} - 1 \right| \leq \sum_{j=3}^{\infty} \frac{|\kappa_j|}{(j-1)!} |t|^{j-2} \leq \sum_{j=3}^{\infty} \frac{1}{j-1} \frac{|t|^{j-2}}{\Delta^{j-2}}$$

$$= \frac{1}{|t|/\Delta} \left[-\log\left(1 - \frac{|t|}{\Delta}\right) - \frac{|t|}{\Delta} \right] =: \varepsilon\left(\frac{|t|}{\Delta}\right). \qquad (A.4)$$

By the Lagrange inversion formula [43, Appendix A.6], the coefficient of τ^k in the expansion of $t(\tau)$ is equal to $1/k$ times the coefficient of t^{k-1} in the expansion of $(\varphi'(t)/t)^{-k}$, which we write as

$$b_k = [\tau^k] t(\tau) = \frac{1}{k} [t^{k-1}] \left(\frac{\varphi'(t)}{t} \right)^{-k}. \qquad (A.5)$$

Since coefficients in convergent integrals can be extracted by complex contour integrals, it follows that

$$b_k = \frac{1}{k} \frac{1}{2\pi i} \oint \frac{1}{(\varphi'(t)/t)^k} \frac{dt}{t^k}$$

with the contour of integration a circle $|t| = r$ with $r \in (0, \Delta)$. It follows that

$$|b_k| \leq \frac{1}{kr^{k-1}} \sup_{|t|=r} \frac{1}{|\varphi'(t)/t|^k} \qquad (r \in (0, \Delta))$$

which yields, together with (A.4),

$$|b_k| \leq \frac{1}{kr^{k-1}(1 - \varepsilon(r/\Delta))^k} \qquad (r \in (0, \Delta)).$$

Let us choose $r = \Delta/2$. Then

$$\varepsilon(r/\Delta) = \varepsilon(\frac{1}{2}) = 2(\log 2 - \frac{1}{2}) = \log 4 - 1 \simeq 0.386 \leq 0.4$$

hence $r(1 - \varepsilon(r/\Delta)) \geq \frac{1}{2}\Delta(1 - 0.4) = 0.3\,\Delta$ and

$$|b_k| \leq \frac{r}{k} \frac{1}{(r[1 - \varepsilon(\frac{1}{2})])^k} \leq \frac{\Delta}{2k} \frac{1}{(0.3\,\Delta)^k}. \tag{A.6}$$

Therefore the radius of convergence of $\sum_k b_k t^k$ and $\sum_j \lambda_j t^j$ is at least $0.3\,\Delta$. $\qquad \square$

A.2. When X has moments up to order $s \geq 3$

More generally, we adopt the recurrence relation from Proposition A.1 as a definition of coefficients b_k and λ_k.

Definition A.3. *Fix $s \geq 3$. Let X be a real-valued random variable with mean $\mathbb{E}[X] = \mu$ and variance $\sigma^2 = \mathbb{V}(X) > 0$. Assume $\mathbb{E}[|X|^s] < \infty$. Then we define coefficients $b_1, \ldots, b_{s-1}$ and $\lambda_2, \ldots, \lambda_s$ as follows:*

- *$b_1 := 1/\sigma^2$.*
- *$b_2, \ldots, b_{s-1}$ are defined recursively by (A.2).*
- *$\lambda_k := -b_{k-1}/k$ for $k = 1, \ldots, s$.*

If $s = \infty$, the series $\sum_{j=0}^{\infty} \lambda_{j+3} \tau^j$ is called Cramér-Petrov series.

A substitute for Proposition A.1 is the following. Define

$$\tilde{\varphi}(t) := \sum_{j=1}^{s} \frac{\kappa_j}{j!} t^j, \quad \tilde{\kappa}_j := \begin{cases} \kappa_j, & j \leq s, \\ 0, & j > s. \end{cases}$$

Let $(\tilde{b}_k)_{k \geq 1}$ and $(\tilde{\lambda}_k)_{k \geq 2}$ be the coefficients defined by $\tilde{b}_1 = 1/\sigma^2$, $\tilde{\lambda}_k = -\tilde{b}_{k-1}/k$, and the recurrence relation (A.2) with $\tilde{\kappa}_{r+1}$ instead of κ_{r+1} or equivalently,

$$\tilde{b}_k = -\frac{1}{\sigma^2} \sum_{r=2}^{s-1} \frac{\kappa_{r+1}}{r!} \sum_{\substack{1 \leq j_1, \ldots, j_r \leq k-1: \\ j_1 + \cdots + j_r = k}} \tilde{b}_{j_1} \cdots \tilde{b}_{j_r}. \tag{A.7}$$

A finite induction over k shows that

$$\forall k \in \{2, \ldots, s-1\} : \tilde{b}_k = b_k, \quad \forall k \in \{2, \ldots, s\} : \tilde{\lambda}_k = \lambda_k. \tag{A.8}$$

Proposition A.4. *Under the assumptions of Definition A.3, there exist open intervals $(-\varepsilon, \varepsilon') \ni 0$, $(\mu - \delta, \mu + \delta)$ such that the following holds true:*

(a) *$\check{\varphi}'$ is a bijection from $(-\varepsilon, \varepsilon')$ onto $(\mu - \delta, \mu + \delta)$.*
(b) *The series $\sum_{k=1}^{\infty} \tilde{b}_k \tau^k$ and $\sum_{k=3}^{\infty} \tilde{\lambda}_k \tau^k$ have radius of convergence larger or equal to δ.*
(c) *If $t \in (-\varepsilon, \varepsilon')$ and $\tau \in (-\delta, \delta)$, then*

$$\check{\varphi}'(t) = \mu + \tau \iff t = \sum_{k=1}^{\infty} \tilde{b}_k \tau^k.$$

(d) *For $x \in (\mu - \delta, \mu + \delta)$ and $t \in (-\varepsilon, \varepsilon')$ the solution of $\check{\varphi}'(t) = x$, we have*

$$\tilde{I}(x) := tx - \check{\varphi}(t) = -\frac{x^2}{2\sigma^2} + \sum_{j=3}^{\infty} \tilde{\lambda}_j x^j.$$

Proof. Parts (a) and (b) follow from the holomorphic inverse function theorem. The function $\check{\varphi}' : \mathbb{C} \to \mathbb{C}$ is a polynomial, so in particular holomorphic. Its derivative at 0 is $\check{\varphi}''(0) = \frac{1}{\sigma^2} \neq 0$, and at 0 it takes the value $\check{\varphi}'(0) = \mu$. Therefore there exist open neighborhoods $U, V \subset \mathbb{C}$ of 0 and μ, respectively, such that $\check{\varphi}'$, restricted to U, is a bijection from U onto V with holomorphic inverse[2] $(\varphi')^{-1}$.

Let $\delta > 0$ be such that the open disk $B(\mu, \delta) := \{ z \in \mathbb{C} \mid |z - \mu| < \delta \}$ is contained in V and $(-\varepsilon, \varepsilon') := (\check{\varphi}')^{-1}((\mu - \delta, \mu + \delta))$. Then part (a) of the lemma is clearly satisfied. The Taylor series of $(\check{\varphi}')^{-1}$ at μ has radius of convergence at least δ. A reasoning completely analogous to the proof of Proposition A.1 shows that the coefficients of the Taylor series are equal to $\tilde{b}_k$. Part (b) and (c) of the lemma follow.

For (d), we note

$$\frac{\mathrm{d}}{\mathrm{d}x} \tilde{I}(x) = (\check{\varphi}')^{-1}(x) + x \frac{1}{\check{\varphi}''((\check{\varphi}')^{-1}(x))} - \check{\varphi}'\left((\check{\varphi}')^{-1}(x)\right) \frac{1}{\check{\varphi}''((\check{\varphi}')^{-1}(x))} = (\check{\varphi}')^{-1}(x)$$

and conclude by an argument similar to the proof of Proposition A.1. $\qquad\square$

Eq. (A.6) in the proof of Proposition A.2 has a counterpart as well.

Proposition A.5. *Assume $\mathbb{E}[X] = 0$, $\mathbb{V}(X) = 1$, and $|\kappa_j| \leq (j-2)!/\Delta^{j-2}$ for all $j = 3, \ldots, s$. Then*

$$|\tilde{b}_k| \leq \frac{\Delta}{2k} \frac{1}{(0.3\,\Delta)^k}$$

for all $k \geq 2$.

The proof is similar to the proof of Eq. (A.6) and is omitted.

---

[2]Strictly speaking, we should write $(\check{\varphi}'|_U)^{-1}$, since $\check{\varphi}'$ with domain $\mathbb{C}$ is *not* injective.

We conclude with a representation of the coefficients λ_k needed in the proof of Proposition A.6 below. We assume $\mathbb{E}[X] = 0$ and $\mathbb{V}(X) = 1$. Eq. (A.8) and the analogue of (A.5) for $\tilde{b}_k$ instead of b_k yields

$$
\lambda_k = -\frac{\tilde{b}_{k-1}}{k} = -\frac{1}{k(k-1)}[t^{k-2}]\Big(\frac{\tilde{\varphi}'(t)}{t}\Big)^{-(k-1)}
$$

$$
= -\frac{1}{k(k-1)}[t^{k-2}]\Big(1 + \sum_{j=3}^{s}\frac{\kappa_j}{(j-1)!}t^{j-2}\Big)^{-(k-1)} \tag{A.9}
$$

for all $k \leq s$.

A.3. Bounds under the Statulevičius condition

Proposition A.6. *Under condition $\mathcal{S}_\gamma$, the coefficients λ_k of the Cramér-Petrov series $\sum_{k\geq3}\lambda_k x^k$ satisfy*

$$
|\lambda_k| \leq \frac{(k+2)!^\gamma}{(\Delta/15)^{k-2}} \quad (k \geq 3).
$$

The proposition is a slightly improved version of [101, Eq. (2.67)].

Proof. By Eq. (A.9) (applied to $s = k$), we have

$$
\lambda_k = -\frac{1}{k(k-1)}[t^{k-2}]\Big(1 + \sum_{j=3}^{k}\frac{\kappa_j}{(j-1)!}t^{j-2}\Big)^{-(k-1)}.
$$

Let $g_k > 0$ be such that $|\kappa_j| \leq j!/g_k^{j-2}$ for all $j = 3, \ldots, k$; a bound for g_k is given shortly. Then Cauchy's inequality yields

$$
|\lambda_k| \leq \frac{1}{k(k-1)}\inf_r r^{-(k-2)}\Big(1 - \sum_{j=3}^{k}j\Big(\frac{r}{g_k}\Big)^{j-2}\Big)^{-(k-1)}.
$$

The infimum is over intervals $r \in [0, \alpha]$ on which the denominator is non-zero. We write $r = \rho\,g_k$, bound the sum by a series, and obtain

$$
|\lambda_k| \leq \frac{g_k^{-(k-2)}}{k(k-1)}\inf_\rho \rho^{-(k-2)}\Big(1 - \sum_{j=3}^{\infty}j\rho^{j-2}\Big)^{-(k-1)}.
$$

A numerical evaluation yields

$$
\inf_\rho \rho^{-1}\Big(1 - \sum_{j=3}^{\infty}j\rho^{j-2}\Big)^{-1} \simeq 14.5 \leq 15
$$

and the minimum is attained at $\rho \simeq 0.126 \leq 0.13$, therefore

$$
|\lambda_k| \leq \frac{g_k^{-(k-2)}}{k(k-1)}0.13 \cdot 15^{k-1} \leq 2\frac{(15/g_k)^{(k-2)}}{k(k-1)}. \tag{A.10}
$$

In order to get a bound on g_k, we check that

$$j!^{1/(j-2)} \leq (k+2)!^{1/(k-2)} \quad (j = 3, \ldots, k) \tag{A.11}$$

or equivalently, $j!^{k-2} \leq (k+2)!^{j-2}$. The proof is by induction over $k \geq j$ at fixed $j \geq 3$. For $k = j$, the claim is obviously true. For the induction step, we note that for all $k \geq j \geq 3$, we have $j! \leq 3! \, j^{j-3}$ hence

$$\frac{j!}{(k+3)^{j-2}} \leq \frac{3! \, j^{j-3}}{(j+3)^{j-2}} \leq \frac{3!}{j+3} \leq 1.$$

Therefore, under the induction hypothesis $(k+2)!^{j-2} \geq j!^{k-2}$, we have

$$(k+3)!^{j-2} = (k+3)^{j-2}(k+2)!^{j-2} \geq (k+3)^{j-2}j!^{k-2} \geq j!^{k-1}.$$

This completes the induction. Condition $(\mathcal{S}_\gamma)$ and Eq. (A.11) yield $|\kappa_j| \leq j!/g_k^{j-2}$ for all $j \leq k$ by choosing

$$\frac{1}{g_k} = \frac{1}{\Delta}(k+2)!^{\gamma/(k-2)}.$$

We deduce from (A.10) that

$$|\lambda_k| \leq \frac{2}{k(k-1)} \, (15/\Delta)^{k-2}(k+2)!^\gamma \leq (15/\Delta)^{k-2}(k+2)!^\gamma. \qquad \square$$

Acknowledgments

References

[1] AMOSOVA, N. N. (1999). Necessity of the Cramér, Linnik and Statulevičius conditions for the probabilities of large deviations. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **260** 9–16, 317. Translated in J. Math. Sci. (New York) 109 (2002), no. 6, 2031–2036. MR1759151

[2] AMOSOVA, N. N. (1999). On the necessity of the Statulevičius condition in limit theorems for probabilities of large deviations. *Liet. Mat. Rink.* **39** 293–303. Translation in Lithuanian Math. J. 39 (1999), no. 3, 231–239. MR1802030

[3] ARIZMENDI, O., HASEBE, T., LEHNER, F. and VARGAS, C. (2015). Relations between cumulants in noncommutative probability. *Adv. Math.* **282** 56–92. MR3374523

[4] Arratia, R., Barbour, A. D. and Tavaré, S. (2003). *Logarithmic combinatorial structures: a probabilistic approach. EMS Monographs in Mathematics.* European Mathematical Society (EMS), Zürich. MR2032426

[5] Bahadur, R. R. and Ranga Rao, R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027. MR0117775

[6] Barbour, A. D., Kowalski, E. and Nikeghbali, A. (2014). Mod-discrete expansions. *Probab. Theory Related Fields* **158** 859–893. MR3176367

[7] Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic techniques for use in statistics. Monographs on Statistics and Applied Probability.* Chapman & Hall, London. MR1010226

[8] Baryshnikov, Y. and Yukich, J. E. (2005). Gaussian limits for random measures in geometric probability. *Ann. Appl. Probab.* **15** 213–253. MR2115042

[9] Bentkus, R. and Rudzkis, R. (1980). On exponential estimates of the distribution of random variables. *Lithuanian Math. J.* **20** 15–30. MR0575427

[10] Berry, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49** 122–136. MR0003498

[11] Błaszczyszyn, B., Yogeshwaran, D. and Yukich, J. E. (2019). Limit theory for geometric statistics of point processes having fast decay of correlations. *Ann. Probab.* **47** 835–895. MR3916936

[12] Bobkov, S. G. (2016). Closeness of probability distributions in terms of Fourier-Stieltjes transforms. *Uspekhi Mat. Nauk* **71** 37–98. translation in Russian Math. Surveys 71 (2016), no. 6, 1021–1079. MR3588939

[13] Bodineau, T., Gallagher, I., Saint-Raymond, L. and Simonella, S. (2020). Statistical dynamics of a hard sphere gas: fluctuating Boltzmann equation and large deviations. Online preprint arXiv:2008.10403.

[14] Borot, G. and Guionnet, A. (2013). Asymptotic expansion of β matrix models in the one-cut regime. *Comm. Math. Phys.* **317** 447–483. MR3010191 MR3010191

[15] Brillinger, D. R. (1975). Statistical inference for stationary point processes. In *Stochastic processes and related topics (Proc. Summer Res. Inst. Statist. Inference for Stochastic Processes, Indiana Univ., Bloomington, Ind., 1974, Vol. 1; dedicated to Jerzy Neyman)* 55–99. MR0381201

[16] Bryc, W. (1993). A remark on the connection between the large deviation principle and the central limit theorem. *Statist. Probab. Lett.* **18** 253–256. MR1245694

[17] Chhaibi, R., Najnudel, J. and Nikeghbali, A. (2017). The circular unitary ensemble and the Riemann zeta function: the microscopic landscape and a new approach to ratios. *Invent. Math.* **207** 23–113. MR3592756

[18] Costin, O. and Lebowitz, J. L. (1995). Gaussian Fluctuation in Random Matrices. *Phys. Rev. Lett.* **75** 69–72. MR3155254

[19] CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actual. Sci. Industr.* **736** 5–23.

[20] CRAMÉR, H. and TOUCHETTE (TRANSLATOR), H. (2018). On a new limit theorem in probability theory. (Translation of: Sur un nouveau théorème-limite de la théorie des probabilités). Electronic preprint arXiv:1802.05988v3 [math.HO].

[21] DELBAEN, F., KOWALSKI, E. and NIKEGHBALI, A. (2014). Mod-φ convergence. *Int. Math. Res. Not. IMRN* **2015** 3445–3485. MR3373056

[22] DEMBO, A. and ZEITOUNI, O. (1998). *Large deviations techniques and applications*, second ed. *Applications of Mathematics (New York)* **38**. Springer-Verlag, New York. MR1619036

[23] DENISOV, D., DIEKER, A. B. and SHNEER, V. (2008). Large deviations for random walks under subexponentiality: the big-jump domain. *Ann. Probab.* **36** 1946–1991. MR2440928

[24] DETTE, H. and TOMECKI, D. (2019). Determinants of block Hankel matrices for random matrix-valued measures. *Stochastic Process. Appl.* **129** 5200–5235. MR4025703

[25] DOBRUSHIN, R. L. and SHLOSMAN, S. B. (1987). Completely analytical interactions: constructive description. *J. Statist. Phys.* **46** 983–1014. MR0893129

[26] DÖRING, H. and EICHELSBACHER, P. (2013). Moderate deviations for the eigenvalue counting function of Wigner matrices. *ALEA Lat. Am. J. Probab. Math. Stat.* **10** 27–44. MR3083918

[27] DÖRING, H. and EICHELSBACHER, P. (2013). Edge fluctuations of eigenvalues of Wigner matrices. In *High dimensional probability VI. Progr. Probab.* **66** 261–275. Birkhäuser/Springer, Basel. MR3443505

[28] DÖRING, H. and EICHELSBACHER, P. (2013). Moderate deviations via cumulants. *J. Theoret. Probab.* **26** 360–385. MR3055808

[29] DOUSSE, J. and FÉRAY, V. (2019). Weighted dependency graphs and the Ising model. *Ann. Inst. Henri Poincaré D* **6** 533–571. MR4033680

[30] DUNEAU, M., IAGOLNITZER, D. and SOUILLARD, B. (1973). Decrease properties of truncated correlation functions and analyticity properties for classical lattices and continuous systems. *Comm. Math. Phys.* **31** 191–208. MR0337229

[31] EICHELSBACHER, P. and KNICHEL, L. (2019). Moment estimates of Rosenthal type via cumulants. arXiv:1901.04865.

[32] EICHELSBACHER, P., RAIČ, M. and SCHREIBER, T. (2015). Moderate deviations for stabilizing functionals in geometric probability. *Ann. Inst. Henri Poincaré Probab. Stat.* **51** 89–128. MR3300965

[33] EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling extremal events. Applications of Mathematics (New York)* **33**. Springer-Verlag, Berlin For insurance and finance. MR1458613

[34] ERCOLANI, N. M., JANSEN, S. and UELTSCHI, D. (2019). Singularity analysis for heavy-tailed random variables. *J. Theoret. Probab.* **32** 1–46. MR3908904

[35] FELLER, W. (1971). *An introduction to probability theory and its applica-*

tions. Vol. II. Second edition. John Wiley & Sons, Inc., New York-London-Sydney. MR0270403

[36] Fenzl, M. and Lambert, G. (2020). Precise deviations for disk counting statistics of invariant determinantal processes.

[37] Féray, V. (2018). Weighted dependency graphs. *Electron. J. Probab.* **23** Paper No. 93, 65. MR3858921

[38] Féray, V. (2020). Central limit theorems for patterns in multiset permutations and set partitions. *Ann. Appl. Probab.* **30** 287–323. MR4068312

[39] Féray, V., Méliot, P. L. and Nikeghbali, A. (2016). *Mod-φ convergence. SpringerBriefs in Probability and Mathematical Statistics.* Springer, Cham Normality zones and precise deviations. MR3585777

[40] Féray, V., Méliot, P. L. and Nikeghbali, A. (2019). Mod-φ Convergence, II: Estimates on the Speed of Convergence. In *Séminaire de Probabilités L* (C. Donati-Martin, A. Lejay and A. Rouault, eds.) 405–477. Springer International Publishing, Cham. MR4291787

[41] Féray, V., Méliot, P. L. and Nikeghbali, A. (2020). Graphons, permutons and the Thoma simplex: three mod-Gaussian moduli spaces. *Proceedings of the London Mathematical Society* **121** 876–926. MR4105789

[42] Fisher, R. A. and Wishart, J. (1931). The derivation of the Pattern Formulae of Two-Way Partitions from those of Simpler Patterns. *Proc. London Math. Soc. (2)* **33** 195–208. MR1576825

[43] Flajolet, P. and Sedgewick, R. (2009). *Analytic combinatorics.* Cambridge University Press, Cambridge. MR2483235

[44] Gnedenko, B. V. and Kolmogorov, A. N. (1968). *Limit distributions for sums of independent random variables. Translated from the Russian, annotated, and revised by K. L. Chung. With appendices by J. L. Doob and P. L. Hsu. Revised edition.* Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills., Ont. MR0233400

[45] Gordon, R. D. (1941). Values of Mills' Ratio of Area to Bounding Ordinate and of the Normal Probability Integral for Large Values of the Argument. *Ann. Math. Statist.* **12** 364–366. MR0005558

[46] Götze, F., Heinrich, L. and Hipp, C. (1995). m-dependent random fields with analytic cumulant generating function. *Scand. J. Statist.* **22** 183–195. MR1339750

[47] Grote, J., Kabluchko, Z. and Thäle, C. (2019). Limit theorems for random simplices in high dimensions. *ALEA Lat. Am. J. Probab. Math. Stat.* **16** 141–177. MR3903027

[48] Grote, J. and Thäle, C. (2018). Concentration and moderate deviations for Poisson polytopes and polyhedra. *Bernoulli* **24** 2811–2841. MR3779703

[49] Grote, J. and Thäle, C. (2018). Gaussian polytopes: a cumulant-based approach. *J. Complexity* **47** 1–41. MR3804598

[50] Grübel, R. and Kabluchko, Z. (2017). Edgeworth expansions for profiles of lattice branching random walks. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 2103–2134. MR3729649

[51] Gusakova, A. and Thäle, C. (2021). The volume of simplices in high-

dimensional Poisson–Delaunay tessellations. *Annales Henri Lebesgue* **4** 121–153. MR4213157

[52] Hald, A. (2000). The Early History of the Cumulants and the Gram-Charlier Series. *International Statistical Review* **68** 137–153.

[53] Heinrich, L. (1985). Some estimates of the cumulant-generating function of a sum of m-dependent random vectors and their application to large deviations. *Math. Nachr.* **120** 91–101. MR0808334

[54] Heinrich, L. (1987). A method for the derivation of limit theorems for sums of weakly dependent random variables: a survey. *Optimization* **18** 715–735. MR0909654

[55] Heinrich, L. (1990). Some bounds of cumulants of m-dependent random fields. *Math. Nachr.* **149** 303–317. MR1124812

[56] Heinrich, L. (2005). Large deviations of the empirical volume fraction for stationary Poisson grain models. *Ann. Appl. Probab.* **15** 392–420. MR2115047

[57] Heinrich, L. (2007). An almost-Markov-type mixing condition and large deviations for Boolean models in the line. *Acta Appl. Math.* **96** 247–262. MR2327539

[58] Heinrich, L. and Richter, W. D. (1984). On moderate deviations of sums of m-dependent random vectors. *Math. Nachr.* **118** 253–263. MR0773624

[59] Heinrich, L. and Spiess, M. (2009). Berry-Esseen bounds and Cramér-type large deviations for the volume distribution of Poisson cylinder processes. *Lith. Math. J.* **49** 381–398. MR2591874

[60] Heinrich, L. and Spiess, M. (2013). Central limit theorems for volume and surface content of stationary Poisson cylinder processes in expanding domains. *Adv. in Appl. Probab.* **45** 312–331. MR3102453

[61] Hofer, L. (2017). A central limit theorem for vincular permutation patterns. *Discrete Math. Theor. Comput. Sci.* **19** Paper No. 9, 26. MR3783365

[62] Hwang, H. K. (1998). On convergence rates in the central limit theorems for combinatorial structures. *European J. Combin.* **19** 329–343. MR1621021

[63] Iagolnitzer, D. and Souillard, B. (1979). Lee-Yang theory and normal fluctuations. *Phys. Rev. B (3)* **19** 1515–1518. MR0526177

[64] Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, Groningen With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman.

[65] Ivanoff, G. (1982). Central limit theorems for point processes. *Stochastic Process. Appl.* **12** 171–186. MR0651902

[66] Jacod, J., Kowalski, E. and Nikeghbali, A. (2011). Mod-Gaussian convergence: new limit theorems in probability and number theory. *Forum Math.* **23** 835–873. MR2820392

[67] Janson, S. (1988). Normal convergence by higher semi-invariants with applications to sums of dependent random variables and random graphs.

Ann. Probab. **16** 305–312. MR0920273

[68] KALLABIS, R. S. and NEUMANN, M. H. (2006). An exponential inequality under weak dependence. *Bernoulli* **12** 333–350. MR2218558

[69] KEATING, J. P. and SNAITH, N. C. (2000). Random matrix theory and $\zeta(1/2 + it)$. *Comm. Math. Phys.* **214** 57–89. MR1794265

[70] KOWALSKI, E. and NIKEGHBALI, A. (2010). Mod-Poisson convergence in probability and number theory. *Int. Math. Res. Not. IMRN* **2010** 3549–3587. MR2725505

[71] KOWALSKI, E. and NIKEGHBALI, A. (2012). Mod-Gaussian convergence and the value distribution of $\zeta(\frac{1}{2} + it)$ and related quantities. *J. Lond. Math. Soc. (2)* **86** 291–319. MR2959306

[72] LAMBERT, G. (2018). Limit theorems for biorthogonal ensembles and related combinatorial identities. *Advances in Mathematics* **329** 590-648.

[73] LAMBERT, G. (2018). Mesoscopic fluctuations for unitary invariant ensembles. *Electron. J. Probab.* **23** 33 pp. MR3771744

[74] LEBOWITZ, J. L. and PENROSE, O. (1964). Convergence of virial expansions. *J. Mathematical Phys.* **5** 841–847. MR0163711

[75] LEBOWITZ, J. L., PITTEL, B., RUELLE, D. and SPEER, E. R. (2016). Central limit theorems, Lee-Yang zeros, and graph-counting polynomials. *J. Combin. Theory Ser. A* **141** 147–183. MR3479242

[76] LEONOV, V. P. and ŠIRJAEV, A. N. (1959). On a method of semi-invariants. *Theor. Probability Appl.* **4** 319–329. Translated from Teor. Verojatnost. i Primenen. **4** (1959), 342–355. MR0123345

[77] LUKKARINEN, J., MARCOZZI, M. and NOTA, A. (2018). Summability of connected correlation functions of coupled lattice fields. *J. Stat. Phys.* **171** 189–206. MR3779051

[78] MASON, D. M. and ZHOU, H. H. (2012). Quantile coupling inequalities and their applications. *Probab. Surv.* **9** 439–479. MR3007210

[79] McCULLAGH, P. (1987). *Tensor methods in statistics. Monographs on Statistics and Applied Probability.* Chapman & Hall, London. MR0907286

[80] MÉLIOT, P. L. and NIKEGHBALI, A. (2015). Mod-Gaussian convergence and its applications for models of statistical mechanics. In *In memoriam Marc Yor—Séminaire de Probabilités XLVII. Lecture Notes in Math.* **2137** 369–425. Springer, Cham. MR3444307

[81] MICHELEN, M. and SAHASRABUDHE, J. (2019). Central limit theorems from the roots of probability generating functions. *Adv. Math.* **358** 106840, 27. MR4021875

[82] MICHELEN, M. and SAHASRABUDHE, J. (2019). Central limit theorems and the geometry of polynomials. Online preprint arXiv:1908.09020 [math.PR].

[83] MIKOSCH, T. and NAGAEV, A. V. (1998). Large deviations of heavy-tailed sums with applications in insurance. *Extremes* **1** 81–110. MR1652936

[84] MÖHLE, M. and PITTERS, H. (2015). Absorption time and tree length of the Kingman coalescent and the Gumbel distribution. *Markov Process. Related Fields* **21** 317–338. MR3442826

[85] NAGAEV, A. V. (1968). Local limit theorems with regard to large deviations when Cramér's condition is not satisfied. *Litovsk. Mat. Sb.* **8** 553–579. Selected Transl. in Math. Stat. Probab. 11, 249–278 (1973). MR0243590

[86] NAGAEV, S. V. (1965). Some limit theorems for large deviations. *Teor. Verojatnost. i Primenen* **10** 231–254. English translation in Theor. Probability Appl. 10 (1965), 214–235. MR0185644

[87] NOURDIN, I. and PECCATI, G. (2010). Cumulants on the Wiener space. *J. Funct. Anal.* **258** 3775–3791. MR2606872

[88] PAN, G., WANG, S. and ZHOU, W. (2017). Limit theorems for linear spectrum statistics of orthogonal polynomial ensembles and their applications in random matrix theory. *J. Math. Phys.* **58** 103301. MR3709450

[89] PECCATI, G. and TAQQU, M. S. (2011). *Wiener chaos: moments, cumulants and diagrams. Bocconi & Springer Series* **1**. Springer, Milan; Bocconi University Press, Milan A survey with computer implementation, Supplementary material available online. MR2791919

[90] PEMANTLE, R. and WILSON, M. C. (2013). *Analytic combinatorics in several variables. Cambridge Studies in Advanced Mathematics* **140**. Cambridge University Press, Cambridge. MR3088495

[91] PETROV, V. V. (1954). Generalization of Cramér's limit theorem. *Uspehi Matem. Nauk (N.S.)* **9** 195–202. MR0065058

[92] PETROV, V. V. (1995). *Limit theorems of probability theory. Oxford Studies in Probability* **4**. The Clarendon Press, Oxford University Press, New York Sequences of independent random variables, Oxford Science Publications. MR1353441

[93] PITTERS, H. (2017). On the number of segregating sites. arXiv:1708.05634.

[94] PITTERS, H. (2019). The number of cycles in a random permutation and the number of segregating sites jointly converge to the Brownian sheet. arXiv:1903.04906.

[95] RIDER, B. and VIRÁG, B. (2007). The Noise in the Circular Law and the Gaussian Free Field. *International Mathematics Research Notices* **2007**. rnm006. MR2361453

[96] ROBBINS, H. (1955). A Remark on Stirling's Formula. *American Mathematical Monthly* **62** 402–405. MR0069328

[97] ROTA, G. C. and SHEN, J. (2000). On the combinatorics of cumulants. *J. Combin. Theory Ser. A* **91** 283–304. In memory of Gian-Carlo Rota. MR1779783

[98] RUDZKIS, R. and BAKSHAEV, A. (2017). General theorems on large deviations for random vectors. *Lith. Math. J.* **57** 367–390. MR3685166

[99] RUDZKIS, R., SAULIS, L. and STATULJAVIČUS, V. (1978). A general lemma on probabilities of large deviations. *Litovsk. Mat. Sb.* **18** 99–116, 217. Translated in Lithuanian Math. J. 18 (1978), no. 2, 226–238 (1979). MR0501287

[100] RUELLE, D. (1969). *Statistical mechanics: Rigorous results.* W. A. Benjamin, Inc., New York-Amsterdam. MR0289084

[101] Saulis, L. and Statulevičius, V. A. (1991). *Limit theorems for large deviations. Mathematics and its Applications (Soviet Series)* **73**. Kluwer Academic Publishers Group, Dordrecht Translated and revised from the 1989 Russian original. MR1171883

[102] Schulte, M. and Thäle, C. (2016). Cumulants on Wiener chaos: moderate deviations and the fourth moment theorem. *J. Funct. Anal.* **270** 2223–2248. MR3460239

[103] Schützenberger, P. M. (1954). Contribution aux applications statistiques de la théorie de l´information. Publ. Inst. Statist. Univ. Paris. Thèse d'État). MR0077816

[104] Scott, A. D. and Sokal, A. D. (2005). The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma. *J. Stat. Phys.* **118** 1151–1261. MR2130890

[105] Soshnikov, A. (1998). Level Spacings Distribution for Large Random Matrices: Gaussian Fluctuations. *Annals of Mathematics* **148** 573–617. MR1668559

[106] Soshnikov, A. (2000). The central limit theorem for local linear statistics in classical compact groups and related combinatorial identities. *Ann. Probab.* **28** 1353–1370. MR1797877

[107] Soshnikov, A. (2002). Gaussian limit for determinantal random point fields. *Ann. Probab.* **30** 171–187. MR1894104

[108] Speed, T. P. (1983). Cumulants and partition lattices. *Australian Journal of Statistics* **25** 378–388. MR0725217

[109] Sturmfels, B. and Zwiernik, P. (2013). Binary cumulant varieties. *Annals of combinatorics* **17** 229–250. MR3027579

[110] Thiele, T. N. (1889). Forelæsninger over almindelig Iagttagelseslære: Sandsynlighedsregning og mindste Kvadraters Methode. *Reitzel, Copenhagen.*

[111] Tsirelson, B. (2012). Bernstein inequality. Encyclopedia of Mathematics http://encyclopediaofmath.org/index.php?title=Bernstein_inequality &oldid=15217. Adapted from an original article by A.V. Prokhorov, N.P. Korneichuk, V.P. Motornyi (originator), which appeared in Encyclopedia of Mathematics - ISBN 1402006098.

[112] Zolotarev, V. M. (1965). On the closeness of the distributions of two sums of independent random variables. *Teor. Verojatnost. i Primenen.* **10** 519–526. MR0189109

[113] Zolotarev, V. M. (1967). A sharpening of the inequality of Berry-Esseen. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **8** 332–342. MR0221570

[114] Zwiernik, P. (2016). *Semialgebraic statistics and latent tree models. Monographs on Statistics and Applied Probability* **146**. Chapman & Hall/CRC, Boca Raton, FL. MR3379921