Extreme value limit theorems in discrete mathematics

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von M.Sc. Philip Dörr,

geboren am 21. Juni 1996 in Karlsruhe,

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. Thomas Kahle
Prof. Dr. Christoph Thäle

eingereicht am: 27.08.2024

Verteidigung am: 21.01.2025

Summary

The present thesis deals with limit theorems for the numbers of inversions and descents, as well as for related combinatorial objects. This subject combines two different areas of mathematics: extreme value theory and statistical algebra. Permutation statistics have been studied and investigated for a long time, while also generating recent interest. Many of them can be generalized from symmetric groups to finite Coxeter groups, as symmetric groups are a subclass of these. When we treat the underlying Coxeter group as a probability space, commonly equipped with the uniform distribution, the numbers of inversions and descents become random variables, which we call X_{inv} and X_{des}. The study of their probabilistic and asymptotic properties, such as the central limit theorem (CLT), is part of statistical algebra.

The validity of the CLT of X_{inv} and X_{des} was recently classified on all finite Coxeter groups, which serves as a motivation to subsequently investigate the extreme value behaviors of these statistics. Hence, in this thesis, we initiate the study of *extreme values of permutation statistics*. By this, we understand the maximum of a permutation statistic within a collection of independent samples. To get meaningful asymptotic statements, we need to construct a triangular array that encompasses an infinite sequence $(W_n)_{n\in\mathbb{N}}$ of finite Coxeter groups of increasing ranks and contains k_n i.i.d. samples $X_{n1}, \dots, X_{nk_n}$ drawn from W_n. So far, there is little understanding and classification for extreme values of triangular arrays, which is why we aim to find new ways to tackle these extreme values.

Due to the asymptotic normality of inversions and descents, the Gumbel distribution is a natural candidate for the extreme value limit distribution in a triangular array built from X_{inv} or X_{des}. However, the choice of the number of samples k_n plays a crucial role. On one hand, k_n must diverge to infinity, but on the other hand, it must not be too large to avoid degeneracy of extreme values. One novel contribution is a universal bound on k_n which is yielded by the Berry-Esseen bound. However, we are clearly interested in methods and results that make use of the underlying permutation statistics and its specific properties. We find such methods that can yield significantly stronger bounds on k_n, which are ideally subexponential, and this turns out to be successful for both the number of inversions and the number of descents. This is highlighted as the first main result of this thesis.

In addition, we provide asymptotic theory for the joint distribution $(X_{\text{inv}}, X_{\text{des}})^\top$, which is less understood than the individual statistics X_{inv} and X_{des}. Up to date, $(X_{\text{inv}}, X_{\text{des}})^\top$ is only known to be asymptotically normal on symmetric groups. The primary challenge here is the dependency structure between X_{inv} and X_{des}, so the joint distribution needs to be handled with more elaborate methods. To tackle the dependency, we use the Hájek projection $\hat{X}_{\text{inv}}$ of X_{inv}, giving an m-dependent approximation $(\hat{X}_{\text{inv}}, X_{\text{des}})^\top$ for which we can obtain a suitable quantitative Gaussian approximation. This leads to several novel contributions, namely, the extension of the CLT for $(X_{\text{inv}}, X_{\text{des}})^\top$ beyond symmetric groups, and the max-attraction of $(X_{\text{inv}}, X_{\text{des}})^\top$ to a bivariate Gumbel distribution.

The latter is highlighted as the second main result of this thesis. We further argue that this result can be obtained for other permutation statistics with certain properties as well.

Finally, we extensively analyze the applicability of these methods to *generalized* inversions and descents, which are an interesting combinatorial extension of common inversions and descents. These generalized statistics are indexed by a parameter d whose magnitude is significant for asymptotic considerations, so we determine its impact on the CLT and the extreme value limit theorems.

This thesis is organized as follows. Chapter 1 introduces the basics and different facets of extreme value theory in both univariate and multivariate scenarios, and outlines the research history of extremes of triangular arrays.

Chapter 2 gives preliminaries on inversions and descents first on symmetric groups, then in the more general framework of finite Coxeter groups. Here, we also outline the research history of the CLT for the random number of inversions and descents.

In Chapter 3, we prove the max-attraction to the Gumbel distribution of inversions and descents for a large class of finite Coxeter groups, based on generating functions and their decomposition into independent but not identically distributed parts. For this scenario, we employ large deviations theory to obtain tail equivalence to the standard normal distribution, which leads to extreme value limit theorems. We also discuss the asymptotic upper bound on the number of samples k_n in each row of the triangular array. For X_{des}, we find an even better bound on k_n than for X_{inv}. Moreover, we show that a weak but universal extreme value limit theorem applies to a very large class of random variables, including other permutation statistics.

Chapter 4 is extensively devoted to the asymptotics of the joint distribution $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$. We first introduce the Hájek projection $\hat{X}_{\mathrm{inv}}$ and a high-dimensional Gaussian approximation for m-dependent random vectors. With these tools, we first extend the CLT for $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$ from symmetric groups to the signed and even-signed permutation groups, on which we also propose a biased random choice of signs. More importantly though, we show that $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$ is in the maximum domain of attraction of the two-dimensional Gumbel distribution with independent marginals.

In the concluding Chapter 5, we determine the univariate and bivariate extreme value asymptotics of generalized inversions and descents with the previously developed methods.

Zusammenfassung

Die vorliegende Dissertation beschäftigt sich mit Grenzwertsätzen für die Anzahlen von Inversions und Descents sowie für ähnliche kombinatorische Objekte. Dieses Thema verbindet zwei verschiedene Gebiete der Mathematik: Extremwerttheorie und statistische Algebra. Permutationsstatistiken werden seit Langem erforscht und sind auch in jüngster Zeit von Interesse. Viele können von symmetrischen Gruppen auf allgemeine endliche Coxeter-Gruppen übertragen werden, da symmetrische Gruppen eine Teilfamilie der letzteren bilden. Wir können die zugrundeliegende Coxetergruppe als Wahrscheinlichkeitsraum auffassen, üblicherweise mit der Gleichverteilung. Die Anzahl der Inversions bzw. Descents ist dann eine Zufallsvariable, die wir mit X_{inv} bzw. X_{des} benennen. Die Erforschung der stochastischen und asymptotischen Eigenschaften solcher Zufallsvariablen, etwa des zentralen Grenzwertsatzes (ZGWS), gehört zum Gebiet der statistischen Algebra.

Die Gültigkeit des ZGWS für X_{inv} und X_{des} wurde erst kürzlich für alle endlichen Coxetergruppen charakterisiert. Dies ist für uns eine Motivation, auch die Extremwerte dieser Statistiken zu untersuchen. Diese Dissertationsarbeit initiiert somit das Studium der *Extremwerte von Permutationsstatistiken.* Darunter verstehen wir den größten Wert einer Permutationsstatistik innerhalb von unabhängigen Stichproben. Um sinnvolle asymptotische Aussagen erhalten zu können, müssen wir ein Dreiecksschema konstruieren, das eine unendliche Folge $(W_n)_{n\in\mathbb{N}}$ von immer größer werdenden endlichen Coxeter-Gruppen umfasst. Auf der n-ten Gruppe W_n ziehen wir dann k_n unabhängige Stichproben der interessierenden Permutationsstatistik. Allerdings sind Extremwerte von Dreiecksschemata nur unzureichend verstanden bzw. klassifiziert, weshalb wir nach neuen Wegen suchen, um mit ihnen umzugehen.

Angesichts des ZGWS ist die Gumbel-Verteilung ein natürlicher Kandidat für die Extremwertverteilung eines Dreiecksschemas aus X_{inv} oder X_{des}. Allerdings spielen die Zeilenlängen k_n des Dreiecksschemas eine entscheidende Rolle. Einerseits muss k_n bestimmt divergieren, andererseits darf es nicht zu groß sein, damit das Extremwertverhalten nicht entartet. Eine neue Erkenntnis ist die Einführung einer universellen Schranke für k_n, die aus der Berry-Esseen-Fehlerabschätzung resultiert. Wir interessieren uns jedoch besonders für Methoden und Erkenntnisse, welche die zugrundeliegende Permutationsstatistik berücksichtigen und ihre Eigenschaften ausnutzen. Wir entwickeln solche Methoden, die zu einer viel besseren, idealerweise subexponentiellen k_n-Schranke führen, sowohl für die Anzahl der Inversions als auch für die Anzahl der Descents. Dies ist das erste Hauptresultat dieser Arbeit.

Es ist auch interessant, die gemeinsame Verteilung $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$, die bislang weniger verstanden ist als ihre individuellen Komponenten X_{inv} und X_{des}, auf ihre asymptotischen Eigenschaften zu untersuchen. Nach heutigem Stand ist für sie nur der ZGWS auf symmetrischen Gruppen bekannt. Die wesentliche Herausforderung ist hier die Abhängigkeit zwischen X_{inv} und X_{des}, weshalb hier aufwändigere Methoden erforderlich sind. Um das Problem der Abhängigkeit zu lösen, benutzen wir die Hájek-Projektion $\hat{X}_{\mathrm{inv}}$ von X_{inv}

zwecks einer m-abhängigen Approximation $(\hat{X}_{\text{inv}}, X_{\text{des}})^\top$, für die wiederum eine geeignete Gauß-Approximation genutzt werden kann. Dies führt zu mehreren neuen Ergebnissen, nämlich die Verallgemeinerung des ZGWS auf andere endliche Coxeter-Gruppen sowie die Max-Anziehung von $(X_{\text{inv}}, X_{\text{des}})^\top$ zu einer zweidimensionalen Gumbel-Verteilung. Letzteres ist das zweite Hauptresultat dieser Arbeit. Es zeigt sich außerdem, dass dieses Resultat auch für andere Permutationsstatistiken mit bestimmten Eigenschaften gültig ist.

Abschließend analysieren wir in aller Ausführlichkeit die Anwendbarkeit dieser Methoden auf die sog. *verallgemeinerten* Inversions und Descents, die eine interessante kombinatorische Erweiterung der herkömmlichen Inversions und Descents darstellen. Diese erweiterten Statistiken werden durch einen Parameter d indiziert, dessen asymptotische Größenordnung entscheidend ist. Somit bestimmen wir seinen Einfluss auf die Gültigkeit des ZGWS und auf das Extremwertverhalten.

Diese Arbeit ist wie folgt strukturiert: Kapitel 1 führt in die Grundlagen und verschiedenen Richtungen der Extremwerttheorie im Ein- wie im Mehrdimensionalen ein und rekapituliert die Forschungsgeschichte von Extremwerten auf Dreiecksschemata.

Kapitel 2 liefert die Grundlagen für Inversions und Descents zunächst auf symmetrischen, dann allgemeiner auf endlichen Coxeter-Gruppen. Dabei besprechen wir auch die bislang bekannten zentralen Grenzwertsätze für die zufällige Anzahl von Inversions oder Descents.

In Kapitel 3 beweisen wir die Max-Anziehung von Inversions und Descents zur Gumbel-Verteilung für eine große Klasse endlicher Coxeter-Gruppen. Dabei nutzen wir Zerlegungen ihrer erzeugenden Funktionen in unabhängige aber nicht identisch verteilte Summanden und verwenden die sog. *Large Deviations Theory*, um Tail-Äquivalenz zur Standardnormalverteilung zu erhalten. Wir diskutieren dabei auch die asymptotische Schranke für die Längen k_n der Zeilen im Dreiecksschema. Für X_{des} ergibt sich dabei sogar eine bessere Schranke als für X_{inv}. Zudem zeigen wir, dass ein schwächerer Extremwertsatz mit einer strengen Schranke für k_n für eine sehr allgemeine Klasse von Zufallsvariablen gilt, die auch andere Permutationsstatistiken beinhaltet.

Kapitel 4 ist ausführlich der Asymptotik der gemeinsamen Verteilung $(X_{\text{inv}}, X_{\text{des}})^\top$ gewidmet. Wir führen zunächst die Hájek-Projektion $\hat{X}_{\text{inv}}$ und die hochdimensionale Gauß-Approximation m-abhängiger Zufallsvektoren ein. Mit diesen Hilfsmitteln erhalten wir den ZGWS für $(X_{\text{inv}}, X_{\text{des}})^\top$ auch auf den sog. *(even-)signed permutation groups*, für die wir auch eine verzerrte Auswahl der Vorzeichen diskutieren. Noch wichtiger ist jedoch der Beweis, dass $(X_{\text{inv}}, X_{\text{des}})^\top$ im Max-Anziehungsbereich der zweidimensionalen Gumbel-Verteilung mit unabhängigen Rändern liegt.

Im abschließenden Kapitel 5 betrachten wir die verallgemeinerten Inversions und Descents, und bestimmen mit den zuvor entwickelten Techniken auch für diese Klasse von Permutationsstatistiken die uni- und bivariate Extremwert-Asymptotik.

Declaration of Authorship

The results and scientific contributions of this thesis have been published in the following papers:

- Philip Dörr and Thomas Kahle. *Extreme Values of Permutation Statistics.* The Electronic Journal of Combinatorics **31** (2024), no. 3, Paper No. 3.10, 1–18. [40]
- Philip Dörr and Johannes Heiny. *Joint extremes of inversions and descents of random permutations.* Journal of Theoretical Probability **38** (2025), no. 2, 1–37. [39]
- Philip Dörr. *Extremes of generalized inversions on symmetric groups.* arXiv preprint arXiv:2404.06598 (2024). Submitted to: Séminaire Lotharingien de Combinatoire. [38]

Chapter 3 is based on the joint publication [40] where I have taken the leading role in developing the theorems and proofs. In the numbering of Chapter 3, the sections 3.2–3.5 stick closely to their counterparts in the accepted version of [40]. There are slight corrections in Remark 3.3.1 and Theorems 3.3.2, 3.4.1. Corollary 3.3.4, Remark 3.5.3, and Section 3.6 are additions that are not found in [40].

Chapter 4 is based on [39] which is a collaboration with Johannes Heiny originating from a one-week research visit at the Ruhr University Bochum. The main ideas resulted from joint conversations during this week. In the later process, Johannes Heiny contributed the idea of introducing p-biased signed permutations. Both authors contributed to the proofs of Theorems 4.3.4 and 4.4.1. I have taken the leading role in developing the other theorems and technical lemmas as well as in performing the structuring and writing. In the numbering of this thesis, Section 4.1 corresponds with [39, Section 2]. However, in Section 4.2, we subsequently cover the Hájek projections on the groups B_n, D_n and the p-bias, after which Theorems 4.3.4 and 4.4.1 are stated for all classical Weyl groups. In contrast, [39, Section 3] focuses on proving Theorem 4.4.1 only for symmetric groups, and [39, Section 4] introduces the other classical Weyl groups and states the CLT in [39, Theorem 4.7], which corresponds to Theorem 4.3.1.

The work on generalized inversions and descents in Chapter 5 has been published separately in [38], of which I am the only author. Sections 5.1 and 5.3 stick closely to their counterparts [38, Section 2] and [38, Section 3], which both use additional preliminaries from [39]. Section 5.2 is a more detailed and illustrated introduction towards the Hájek projection of $X_{\text{inv}}^{(d)}$, whose essentials are contained in [38, Section 4] along with the technical details in Section 5.4.

Contents

Notations and abbreviations **1**

1 Introduction to Extreme Value Theory **3**
1.1 Sequences of independent and identically distributed random variables . . . 4
1.2 Dependent sequences and triangular arrays . . . 10
1.3 Multivariate extreme value theory . . . 14

2 Finite Coxeter groups and permutation statistics **19**
2.1 Finite Coxeter groups, inversions and descents . . . 19
2.2 Stochastic properties . . . 25
2.3 Central Limit Theorems . . . 28

3 Extremes of inversions and descents on finite Coxeter groups **35**
3.1 Introduction and framework . . . 35
3.2 Tail equivalence for non-identically distributed sums . . . 37
3.3 Sequences of classical Weyl groups . . . 40
3.4 Arbitrary finite Coxeter groups . . . 43
3.4.1 Sequences of products of classical Weyl groups . . . 44
3.4.2 Sequences of products involving dihedral groups . . . 44
3.5 Universal EVLT for other permutation statistics . . . 46
3.6 Open questions . . . 48

4 CLT and extremes of multivariate permutation statistics **51**
4.1 The Hájek projection of inversions and descents . . . 52
4.2 Signed and even-signed permutation groups . . . 56
4.2.1 Asymptotic vanishing of correlation . . . 57
4.2.2 Signs with random bias . . . 57
4.2.3 Products of classical Weyl groups . . . 60
4.3 High-dimensional Gaussian approximation and CLT . . . 61
4.4 The extreme value asymptotics of $(X_{\text{inv}}, X_{\text{des}})^\top$ . . . 64
4.5 Summary and open problems . . . 69
4.5.1 Extension of the upper bound on k_n . . . 69
4.5.2 Weakly dependent permutation statistics . . . 71
4.6 Technical proofs and simulation code . . . 74
4.6.1 Proof of Lemma 4.2.1 . . . 74
4.6.2 Proof of Lemma 4.2.3a) . . . 76
4.6.3 Proof of Lemma 4.2.3b) . . . 80
4.6.4 Proof of Lemma 4.2.8 . . . 82
4.6.5 Proof of Lemma 4.2.9 . . . 85
4.6.6 Proof of Lemma 4.2.11 . . . 87
4.6.7 Simulation code . . . 89

5 Generalized inversions and descents **93**
5.1 Basic definitions . . . 93
5.2 The Hájek approximation of generalized inversions . . . 97
5.3 Asymptotic results . . . 102
5.4 Technical proofs . . . 105
5.4.1 Proof of Lemma 5.2.1 . . . 105
5.4.2 Proof of Lemma 5.2.2 . . . 105
5.4.3 Proof of Lemma 5.2.4 . . . 106

List of Symbols **111**

List of Figures **115**

Bibliography **117**

Notations and abbreviations

We use the following basic notations:

- $\mathbb{N}$ is the set of positive integers (not including zero), $\mathbb{R}$ is the set of real numbers, and $\mathbb{R}^+$ is the set of positive real numbers. On the real numbers, $\lfloor\cdot\rfloor$ denotes the floor operator and $\lceil\cdot\rceil$ denotes the ceiling operator. For two real numbers x, y, we use the shorthand notations $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$.

- If x is a vector, then $x^\top$ denotes its transpose. If the components of a vector are listed within continuous text, we write, e.g., $x = (x_1, \dots, x_n)^\top$ to indicate that x is a column vector. We also often use the bold notation $\mathbf{x}$ for fixed vectors in a multidimensional Euclidean space $\mathbb{R}^d$, to distinguish vectors from real numbers.

- For any collection of sets $A_1, \dots, A_n$, their joint Cartesian product is denoted by $\prod_{i=1}^n A_i$. The complement of a set A in a larger set is denoted by A^c.

- If A is a set, then $\mathbf{1}_A$ denotes its indicator function. Typically, A is expressed by a condition such as, e.g., $A = \{x > y\}$, for which we use the notation $\mathbf{1}\{x > y\}$ or $\mathbf{1}_{x>y}$.

- The quantile function of a distribution function F is denoted by $F^\leftarrow$.

- $\mathbb{E}(X)$ denotes the expected value of a random variable or vector.

- $\mathrm{Var}(X)$ denotes the variance and $\sigma(X) = \sqrt{\mathrm{Var}(X)}$ denotes the standard deviation of a random variable. If X is a random vector, then $\mathrm{Var}(X)$ denotes its covariance matrix.

- $\mathrm{Cov}(X, Y)$ is the covariance of two random variables X, Y.

- $(X_n)_{n\in\mathbb{N}}$ denotes a sequence, and $(X_{nj})_{j=1,\dots,k_n}$ denotes a triangular array of random variables, with $(k_n)_{n\in\mathbb{N}}$ denoting the length of its rows. In the case of $k_n = n \ \forall n \in \mathbb{N}$, we refer to $(X_{n1}, \dots, X_{nn})$ as a *uniform* triangular array. For any doubly indexed objects, we omit the comma separation of the two indices if both are uniliteral placeholders. If the first index is a fixed number, we separate it with a comma for the sake of clarification.

- $\longrightarrow$ denotes a limit that refers to $n \to \infty$, unless specified otherwise. When we speak of divergence, we always mean divergence to infinity.

- The symbols $\sim$ and $\overset{\mathcal{D}}{=}$ mean equality in distribution. The symbol $\overset{\mathcal{D}}{\longrightarrow}$ means convergence in distribution, and $\overset{\mathbb{P}}{\longrightarrow}$ means convergence in probability.

In addition, we use the following symbols for important distributions, and also for correspondingly distributed variables if the meaning is clear from the context:

- $U(a, b)$ denotes the uniform distribution on the interval $[a, b]$ for $a, b \in \mathbb{R}$, and $U(\{0, 1, \ldots, n\})$ denotes the discrete uniform distribution on the set $\{0, 1, \ldots, n\}$.
- $\mathrm{Bin}(n, p)$ denotes the binomial distribution with n trials and success probability p. In particular, $\mathrm{Bin}(1, p)$ is the Bernoulli distribution with success probability p.
- $\mathrm{Po}(\lambda)$ denotes the Poisson distribution, and $\mathrm{Exp}(\lambda)$ denotes the exponential distribution with parameter λ.
- $\mathrm{N}(0, 1)$ denotes the standard normal distribution on the real numbers. $\mathrm{N}_d(0, \Sigma)$ denotes the d-variate centered normal distribution with covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$. In particular, I_d denotes the d-dimensional identity matrix.

All Coxeter groups considered in this thesis are finite. When speaking about *products of Coxeter groups*, we always indicate direct products.

Following the common O-notation, we express magnitude relations for positive sequences $(a_n)_{n \in \mathbb{N}}, (b_n)_{n \in \mathbb{N}}$ as follows:

- $a_n = O(b_n)$ means that a_n grows *at most as fast* as b_n, i.e., $\limsup_{n \to \infty} \dfrac{a_n}{b_n} < \infty$. This is also written as $a_n \lesssim b_n$.
- $a_n = o(b_n)$ or $b_n = \omega(a_n)$ means that a_n grows *slower* than b_n, or is *negligible* compared to b_n, i.e., $\lim_{n \to \infty} \dfrac{a_n}{b_n} = 0$. This is also written as $a_n \ll b_n$ or $b_n \gg a_n$.
- $a_n = \Theta(b_n)$ means that a_n and b_n have *the same order of magnitude*, i.e., both $a_n = O(b_n)$ and $b_n = O(a_n)$ hold. Even stronger, $a_n \sim b_n$ means that a_n and b_n are *asymptotically equivalent,* i.e., $\lim_{n \to \infty} \dfrac{a_n}{b_n} = 1$.
- $a_n = b_n + o_{\mathbb{P}}(1)$ means that a_n, b_n are random variables or vectors with $a_n - b_n \xrightarrow{\mathbb{P}} 0$.

Throughout this thesis, we use the following abbreviations:

- CDA = copula domain of attraction
- CDF = cumulative distribution function
- CLT = central limit theorem
- EVD = extreme value distribution
- EVLT = extreme value limit theorem
- GRF = Gaussian random field
- MDA = max-domain of attraction
- MEVD = multivariate extreme value distribution

1 Introduction to Extreme Value Theory

Extreme value theory, or extreme value analysis, deals with extremely large values of random variables or vectors. These extreme values (or *extremes*) are commonly the maxima or high threshold exceedances of sequences, triangular arrays, stochastic processes, or other families of random objects. Therefore, a typical issue in extreme value theory is to investigate the asymptotic properties of extreme values. This theory is an important field of probability theory with scientific and practical applications in all areas concerned with extreme events and their modeling. See, e.g., [52, 69] for general overviews of applications, and see [62] for optimization, [16] for engineering, [61] for meteorology, [103] for public health, and [72, 93] for finance and risk management. One of the most seminal works in extreme value theory is the book *Statistics of Extremes* by Gumbel [55]. Another frequently cited reference is the book by Leadbetter *et al.* [71].

The simplest setting is a sequence $X_1, X_2, \ldots$ of independent and identically distributed (i.i.d.) random variables. The *central limit theorem* (CLT) states that

$$\frac{\mathcal{S}_n - \mathbb{E}(\mathcal{S}_n)}{\sigma(\mathcal{S}_n)} \xrightarrow{\mathcal{D}} \mathrm{N}(0,1) \tag{1.1}$$

for the sequence of partial sums $\mathcal{S}_n := X_1 + X_2 + \ldots + X_n$. The standardization of $\mathcal{S}_n$ constitutes an affine-linear rescaling. In some scenarios, each $\mathcal{S}_n$ is constructed as $\mathcal{S}_n = X_{n,1} + \ldots + X_{n,k_n}$ from an *individual* block of variables of divergent sizes $k_n \in \mathbb{N}$, $k_n \longrightarrow \infty$. This gives a *triangular array* which is denoted by $(X_{nj})_{j=1,\ldots,k_n}$. Oftentimes, $(k_n)_{n\in\mathbb{N}}$ matches the sequence of natural numbers, giving a *uniform* triangular array, but this is not mandatory. Moreover, the i.i.d. case can be generalized in various ways, such as allowing non-identical distributions or weak dependency structures. For these reasons, there are several stronger or weaker versions of the CLT. The property (1.1) is called *asymptotic normality,* and the question of its validity is also interesting if $\mathcal{S}_n$ is not represented as a partial sum of other random variables.

The CLT serves as a motivation to obtain statements analogous to (1.1) for extreme values. For a sequence $X_1, X_2, \ldots$ of random variables, we consider the sequence of maxima

$$M_n := \max\{X_1, X_2, \ldots, X_n\}\,.$$

For a triangular array, we accordingly consider the *row-wise maximum*

$$M_n := \max\{X_{n1}, \ldots, X_{nk_n}\}\,.$$

In either case, we want to find deterministic sequences $(a_n) \subseteq \mathbb{R}^+, (b_n) \subseteq \mathbb{R}$ and a *non-degenerate* limit distribution G so that

$$\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{D}} G\,. \tag{1.2}$$

The affine-linear rescaling of M_n is necessary to give a meaningful limit distribution rather than a plain constant, which is not useful for inference purposes. Similar to the CLT, the existence of such G, a_n, b_n depends on the properties of the underlying sequence or triangular array. We call statements and results in the way of (1.2) *extreme value limit theorems,* or in short, *EVLTs.* These can be viewed as the max-analogues of CLTs, so whenever the question of a CLT is raised, a natural follow-up question is that of an EVLT.

Section 1.1 gives the basic understanding of EVLTs and the classification of G, a_n, b_n for i.i.d. sequences. Section 1.2 discusses dependent stationary sequences and important examples for extremes of triangular arrays. In particular, we review the use of tail equivalence to the standard normal distribution. Section 1.3 introduces multivariate extreme value theory for random vectors.

1.1 Sequences of independent and identically distributed random variables

The foundation of extreme value theory is to understand the extremes of i.i.d. random variables $X_1, X_2, \dots$ with a joint distribution function F. In this scenario, (1.2) translates to

$$\forall x \in \mathcal{C}(G)\colon\ \mathbb{P}\left(\frac{M_n - b_n}{a_n} \le x\right) = \mathbb{P}(M_n \le a_n x + b_n) = F(a_n x + b_n)^n \longrightarrow G(x)\,,$$

where $\mathcal{C}(G)$ denotes the continuity region of G. In fact, it turns out that $\mathcal{C}(G) = \mathbb{R}$ in all relevant cases.

Definition 1.1.1. Let G be a non-degenerate distribution function. Let $F, X_1, X_2, \dots$ be as above. Then, F is in the *max-domain of attraction* (MDA) of G if there exist sequences $a_n > 0$, $b_n \in \mathbb{R}$ such that (1.2) is satisfied. For this situation, we write $F \in \mathrm{MDA}(G)$. Any distribution function G with a non-empty MDA is called an *extreme value distribution* (EVD). Important extreme value distributions are, for $t \in \mathbb{R}$:

- the *Gumbel distribution* $\Lambda(t) := \exp(-\exp(-t))$,
- the *Fréchet distributions* $\Phi_\alpha(t) := \exp(-t^{-\alpha})\mathbf{1}_{t>0}$, where $\alpha > 0$ is a fixed parameter,
- the *Weibull distributions*

$$\Psi_\alpha(t) := \begin{cases} \exp(-|t|^\alpha), & t \le 0 \\ 1, & t \ge 0 \end{cases}, \qquad \alpha > 0\,.$$

Example 1.1.2. The fact that the aforementioned distributions $\Lambda, \Phi_\alpha, \Psi_\alpha$ are extreme value distributions can be verified by the following examples:

- Let $X_1, X_2, \dots \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(1)$. Then, $M_n - \log(n) \xrightarrow{\mathcal{D}} \Lambda$ (see [71, Example 1.7.2]).
- Let $X_1, X_2, \dots$ be i.i.d. samples of the Pareto distribution function

$$F(t) = \begin{cases} 1 - t^{-\alpha}, & t \ge 1 \\ 0, & t < 1 \end{cases}.$$

Then, $n^{-1/\alpha} M_n \xrightarrow{\mathcal{D}} \Phi_\alpha$ (see [71, Example 1.7.6]).

- Let $X_1, X_2, \dots$ be i.i.d. samples of the distribution function

$$F(t) = \begin{cases} 0, & t \le -1 \\ 1 - (-t)^\alpha, & -1 \le t \le 0 \\ 1, & t \ge 0 \end{cases}.$$

  Then, $n^{1/\alpha} M_n \xrightarrow{\mathcal{D}} \Psi_\alpha$ (see [71, Example 1.7.10]).

Illustrations of the Gumbel distribution and the families of the Fréchet and Weibull distributions are given in Figures 1.1, 1.2, and 1.3.

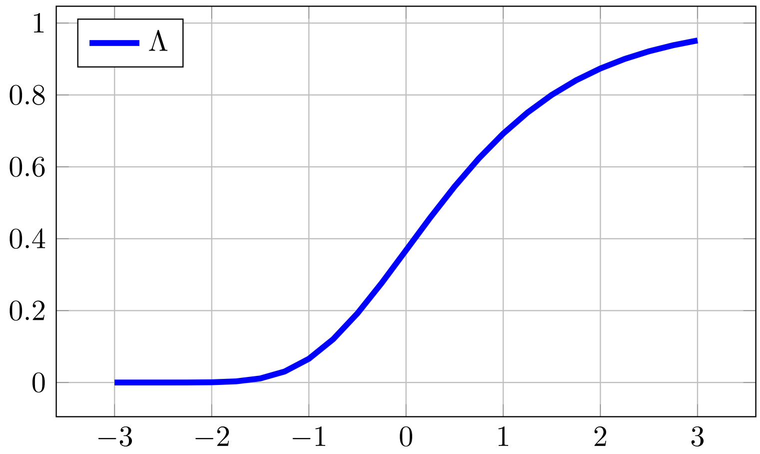


Figure 1.1: A plot of the (standard) Gumbel distribution.

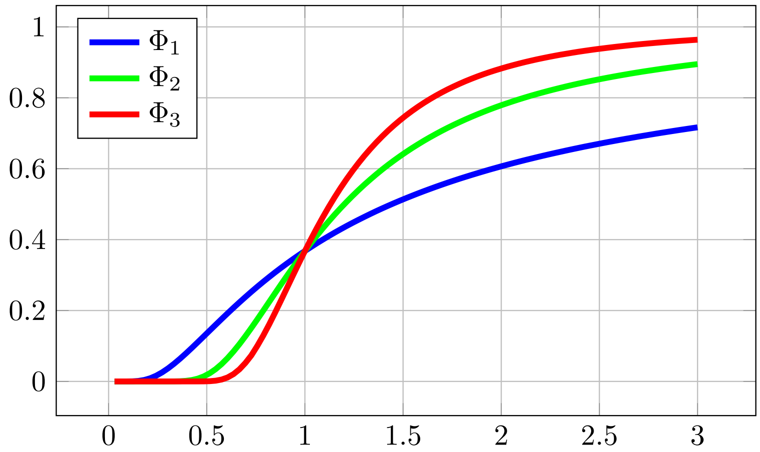


Figure 1.2: A plot of three exemplary Fréchet distributions. Large choices of the shape parameter α produce more curved distributions.

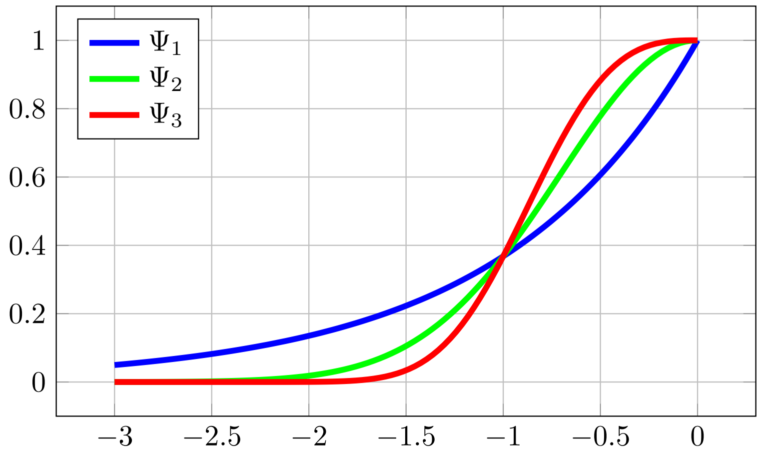


Figure 1.3: A plot of three exemplary Weibull distributions, each cut off at the right endpoint $x^* = 0$. Large choices of the shape parameter α produce more curved distributions.

Remark 1.1.3. An affine transformation of a distribution function G is given by

$$G'(x) = G\left(\frac{x-\mu}{\sigma}\right)$$

for some $\sigma > 0$, $\mu \in \mathbb{R}$. If G is an extreme value distribution, then so is G' and $\mathrm{MDA}(G) = \mathrm{MDA}(G')$. In other words, MDAs are invariant under affine transformations. Two distribution functions G, G' are *of the same type* if they differ only by an affine transformation. If two EVDs G, G' are not of the same type, then their MDAs are disjoint. This is substantiated by *Khintchine's Convergence of Types Theorem* (see, e.g., [71, Theorem 1.2.3]). △

The following theorem is one of the fundamental theorems of extreme value theory. It substantiates the outstanding importance of the EVDs introduced in Definition 1.1.1.

Theorem 1.1.4. (Fisher–Tippett–Gnedenko, cf. [71], Theorem 1.4.2) *The Gumbel, Fréchet, and Weibull distributions are the only types of extreme value distributions for sequences of i.i.d. random variables. In conclusion, if a distribution function F belongs to some MDA, then it is in the MDA of either the Gumbel, a Fréchet or a Weibull distribution.*

Remark 1.1.5. According to [71, Theorem 1.3.1], the extreme value distributions are exactly the *max-stable* distributions G which satisfy $G(c_n x + d_n)^n = G(x)\ \forall x \in \mathbb{R}$ for some constants $c_n > 0$, $d_n \in \mathbb{R}$. In that case, by [71, Corollary 1.3.2], there even exist real-valued functions $c\colon \mathbb{R}^+ \to \mathbb{R}^+$, $d\colon \mathbb{R}^+ \to \mathbb{R}$ so that

$$G\big(c(s)t + d(s)\big)^s = G(t) \quad \forall s, t \in \mathbb{R}\,.$$

The proof of the Fisher–Tippett–Gnedenko theorem narrows down to solving this functional equation. This approach was developed by de Haan [31] and simplified the longer original proof by Gnedenko [54].

Convention: Let F be any distribution function. For simplicity, we say that F *has Gumbel behavior* whenever $F \in \mathrm{MDA}(\Lambda)$ is true. Accordingly, we say that F has Fréchet behavior if $F \in \mathrm{MDA}(\Phi_\alpha)$, and that F has Weibull behavior if $F \in \mathrm{MDA}(\Psi_\alpha)$, for some $\alpha > 0$.

Definition 1.1.6. Let F be a distribution function. The *tail function* (or simply, the *tail*) of F is $\overline{F}(x) := 1 - F(x)$. The *right endpoint* of F is $x^* := \sup\{x \in \mathbb{R} \colon F(x) < 1\}$. See Figure 1.4 for an illustration. △

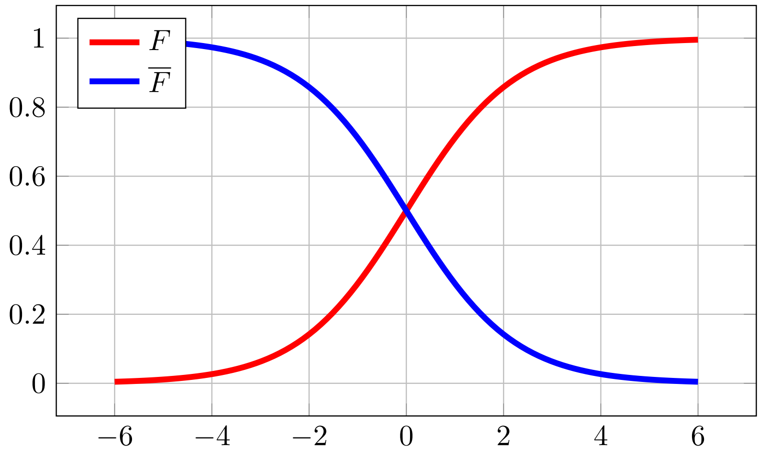


Figure 1.4: Display of a distribution function F together with its tail function $\overline{F}$.

There is a complete classification showing which distribution functions F belong to which MDA, based on the tail and the right endpoint. To describe the asymptotics of the tail, we need the following definition.

Definition 1.1.7. A measurable function $L \colon \mathbb{R}^+ \to \mathbb{R}^+$ is *slowly varying* if

$$\lim_{x\to\infty} \frac{L(\lambda x)}{L(x)} = 1 \quad \forall \lambda > 0\,.$$

The class of slowly varying functions includes those converging to a positive limit, as well as any other function that grows slower than all x^ε and faster than any $x^{-\varepsilon}$, $\varepsilon > 0$ (e.g., logarithmic functions). More generally, a measurable function $R \colon \mathbb{R}^+ \to \mathbb{R}^+$ is *regularly varying* with index $\alpha \in \mathbb{R}$ if

$$\lim_{x\to\infty} \frac{R(\lambda x)}{R(x)} = \lambda^\alpha \quad \forall \lambda > 0\,.$$

This property is denoted by $f \in \mathrm{RV}_\alpha$. It is satisfied if and only if there exists a slowly varying function L with $R(x) = x^\alpha L(x)$. △

The following theorem gives a classification of MDAs and additionally lists feasible choices of the normalization sequences a_n, b_n in (1.2). It shows that the asymptotic extreme value behavior is closely related to the behavior of the corresponding tail function. For proofs, we again refer to [31, 54].

Theorem 1.1.8. *Let F be a distribution function and let $F^{\leftarrow}$ be its quantile function. Let $\gamma_n := F^{\leftarrow}(1-1/n)$.*

(a) *$F \in \mathrm{MDA}(\Phi_\alpha) \iff \overline{F} \in \mathrm{RV}_{-\alpha}$ and $x^* = \infty$. In this case, we can choose $a_n = \gamma_n$ and $b_n = 0$.*
(b) *Assuming $x^* < \infty$, let $F^*(x) := F(x^* - 1/x)$. Then, $F \in \mathrm{MDA}(\Psi_\alpha) \iff x^* < \infty$ and $\overline{F^*} \in \mathrm{RV}_{-\alpha}$. In this case, we can choose $a_n = x^* - \gamma_n$ and $b_n = x^*$.*
(c) *$F \in \mathrm{MDA}(\Lambda)$ if and only if there exists a measurable function $g\colon \mathbb{R} \to \mathbb{R}^+$ such that*

$$\lim_{t\to x^*} \frac{\overline{F}(t + xg(t))}{\overline{F}(t)} = e^{-x} \quad \forall x \in \mathbb{R}\,.$$

In this case, one can choose any b_n such that $n\overline{F}(b_n) \longrightarrow 1$, and $a_n = g(b_n)$.

Remark 1.1.9. The statements of Theorem 1.1.8 can be described as follows: The Fréchet MDAs are for heavy-tailed distributions. The Weibull MDAs are designed for any distribution with a finite right endpoint and polynomial behavior near the right endpoint. For example, this applies to the continuous uniform distributions. The Gumbel MDA is an intermediate case comprising distributions with both finite and infinite right endpoint. Moreover, the suggested choices of a_n and b_n are not unique. For instance, any $b_n + o_{\mathbb{P}}(a_n)$ fulfills the same purpose as b_n due to Slutsky's lemma. △

The following lemma embodies the fundamental connection between tails and extremes. It is the basis of the proof of Theorem 1.1.8 and serves many other purposes as well. In the basic scenario of i.i.d. sequences, it can be proved by simple means, but in more general scenarios, its validity cannot be assumed without further investigation.

Lemma 1.1.10. (see [71], Theorem 1.5.1)
Let F be a distribution function and let $(u_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. Let $\tau \in [0,\infty]$ and let M_n be as in (1.2). The following are equivalent:

(a) *$n(1 - F(u_n)) \longrightarrow \tau$.*
(b) *$\mathbb{P}(M_n \le u_n) = F(u_n)^n \longrightarrow e^{-\tau}$.*

A very important extreme value behavior is that of the standard normal distribution. It is often possible to derive similarities between the extreme values of interest and the extremes of the standard normal distribution. The following fundamental theorem states that the standard normal distribution $\mathrm{N}(0,1)$ has Gumbel behavior.

Theorem 1.1.11. (cf. [71], Theorem 1.5.3)
Let M_n be the maximum of n i.i.d. standard normal variables, and let

$$\alpha_n := \left(\sqrt{2\log(n)}\right)^{-1},$$
$$\beta_n := \sqrt{2\log(n)} - \frac{\log(4\pi\log(n))}{2\sqrt{2\log(n)}}\,.$$

Then, it holds that

$$\frac{M_n - \beta_n}{\alpha_n} \xrightarrow{\mathcal{D}} \Lambda\,.$$

This can be proved in several ways, e.g., by *defining* u_n so that $1 - \Phi(u_n) = n^{-1}e^{-x}$ in Lemma 1.1.10, and then proving that $u_n = \alpha_n x + \beta_n + o(1)$. See [71, Theorem 1.5.3] for details. Figure 1.5 displays the normalization constants. It is interesting that the scaling constant α_n decays as $n \to \infty$, while the translation constant β_n diverges.

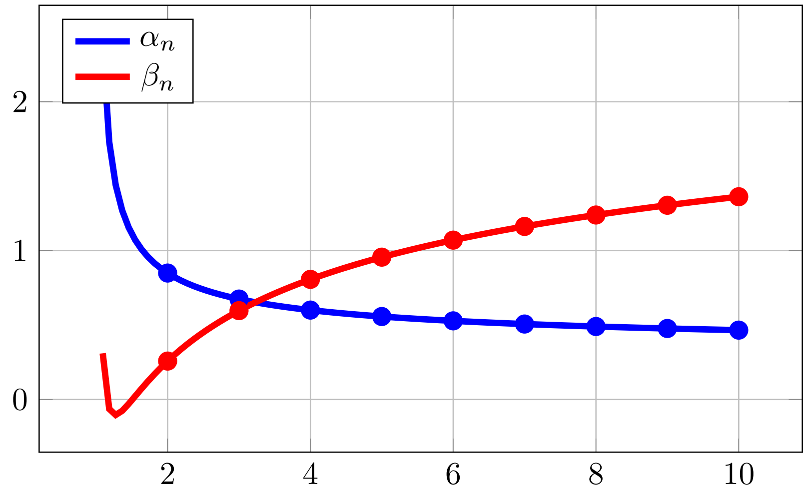


Figure 1.5: Plot of the normalization constants α_n, β_n for $n = 2, \ldots, 10$.

Note that in the literature, it is often preferred to take $\alpha_n := \sqrt{2\log(n)}$, and to write $\alpha_n(M_n - \beta_n) \xrightarrow{\mathcal{D}} \Lambda$. Throughout this thesis, we consistently use the notation in Theorem 1.1.11 since it is analogous to that in (1.2).

Remark 1.1.12. The MDAs of the extreme value distributions do not encompass all probability distributions of real-valued random variables. This especially concerns discrete and discontinuous distributions. Consider any distribution function F with $x^* < \infty$ and $F(x^*-) = \lim_{x \to x^*-} F(x) < 1$, implying $\mathbb{P}(X = x^*) > 0$ for $X \sim F$. This means that for an i.i.d. sequence $X_1, X_2, \ldots \sim F$, there will be some $N \in \mathbb{N}$ with $X_N = x^*$ almost surely, and $M_N = M_{N+1} = \ldots = x^*$. Therefore, the limit distribution of $(M_n)_{n\in\mathbb{N}}$ is a degenerate Dirac measure in x^*, and any affine-linear rescaling still gives a Dirac measure. Such probability distributions have no meaningful extreme value behavior in the context of i.i.d. sequences. Regarding discrete distributions supported on infinitely many numbers, the extreme value asymptotics were investigated and classified by Anderson [1]. According to [1, Eq. (1.3)], a necessary condition for the existence of a non-degenerate extreme value limit distribution is

$$\lim_{x \to x^*} \frac{1 - F(x)}{1 - F(x-)} = 1\,.$$

For example, the Poisson distributions and the geometric distributions do not satisfy this condition. However, any heavy-tailed discrete distribution is in the MDA of a Fréchet distribution according to [98].

An example of a continuous distribution outside of all MDAs is $F(t) = 1 - \log(t)^{-1}$. Its tail is too heavy to satisfy any of the conditions of Theorem 1.1.8.

1.2 Dependent sequences and triangular arrays

For sequences of i.i.d. variables, the classification of extreme value asymptotics is well understood. Now, we focus on extensions of this theory to more general scenarios, such as dependence and non-identical distributions. It turns out that if the random variables show a sufficiently mild degree of dependence, then their extreme values often behave similarly to those of corresponding i.i.d. variables.

Definition 1.2.1. A stationary sequence $(X_n)_{n\in\mathbb{N}}$ satisfies the *strong mixing condition* $\mathfrak{D}$ if there is a decreasing sequence $(g_l)_{l\in\mathbb{N}} \searrow 0$ such that for all $u \in \mathbb{R}$ and all $p, q \in \mathbb{N}$, $i_1 < \ldots < i_p$ and $j_1 < \ldots < j_q$ with $j_1 - i_p > l$:

$$\left|\mathbb{P}(X_{i_1}, \ldots, X_{i_p}, X_{j_1}, \ldots, X_{j_q} \le u) - \mathbb{P}(X_{i_1}, \ldots, X_{i_p} \le u)\mathbb{P}(X_{j_1}, \ldots, X_{j_q} \le u)\right| \le g_l\,.$$

This means that if the distance between the blocks $X_{i_1}, \ldots, X_{i_p}$ and $X_{j_1}, \ldots, X_{j_q}$ increases, then their dependence decreases, regardless of the block sizes. This condition can be weakened by addressing only a sequence of thresholds $(u_n)_{n\in\mathbb{N}}$ rather than all real numbers u. So, we say that $(X_n)_{n\in\mathbb{N}}$ satisfies the *mixing condition* $\mathfrak{D}(u_n)$ if there is a double-indexed sequence $(\alpha_{n,l})$ such that $\alpha_{n,l_n} \searrow 0$ for some $l_n = o(n)$ and such that for all $i_1 < \ldots < i_p$, $j_1 < \ldots < j_q$ as above:

$$\left|\mathbb{P}(X_{i_1} \ldots, X_{i_p}, X_{j_1}, \ldots, X_{j_q} \le u_n) - \mathbb{P}(X_{i_1} \ldots, X_{i_p} \le u_n)\mathbb{P}(X_{j_1}, \ldots, X_{j_q} \le u_n)\right| \le \alpha_{n,l}\,.$$

Theorem 1.2.2. (cf. [71], Theorem 3.3.3)
Let $(X_n)_{n\in\mathbb{N}}$ be a stationary sequence of random variables, and assume the existence of sequences $a_n > 0$, $b_n \in \mathbb{R}$ such that:

- $a_n^{-1}(M_n - b_n) \xrightarrow{\mathcal{D}} G$ *for a non-degenerate distribution function G.*
- *For all $x \in \mathbb{R}$, the condition $\mathfrak{D}(a_n x + b_n)$ is satisfied.*

Then, G must be of a Gumbel, Fréchet, or Weibull type.

However, in order to transfer the classification of Theorem 1.1.8, and to preserve the fundamental equivalence in Lemma 1.1.10, an additional condition is required.

Definition 1.2.3. Let $(X_n)_{n\in\mathbb{N}}$ be a stationary sequence of random variables and let $(u_n)_{n\in\mathbb{N}}$ be a sequence of thresholds. Then, (X_n) satisfies the *anticlustering condition* $\mathfrak{D}'(u_n)$ if

$$\lim_{k\to\infty}\left[\limsup_{n\to\infty}\, n \sum_{j=1}^{\lfloor n/k \rfloor} \mathbb{P}(X_1 > u_n, X_j > u_n)\right] = 0\,.$$

Now, Lemma 1.1.10 carries over as follows:

Lemma 1.2.4. *Let $(X_n)_{n\in\mathbb{N}}$ be a stationary sequence of random variables, and let $(u_n)_{n\in\mathbb{N}}$ be a sequence such that both $\mathfrak{D}(u_n)$ and $\mathfrak{D}'(u_n)$ are satisfied. Then, for any $\tau \in [0, \infty)$:*

$$\lim_{n\to\infty} \mathbb{P}(M_n \le u_n) = e^{-\tau} \iff \lim_{n\to\infty} n\mathbb{P}(X_1 > u_n) = \tau\,.$$

In conclusion, if both mixing conditions can be verified, then the extreme value behavior of a stationary sequence $(X_n)_{n\in\mathbb{N}}$ can be derived from that of an i.i.d. sequence $(X_n^*)_{n\in\mathbb{N}}$ with $X_1^* \stackrel{\mathcal{D}}{=} X_1$. In a more general framework, it is possible that $(M_n - b_n)/a_n \xrightarrow{\mathcal{D}} G^\vartheta$ for some index $\vartheta \in [0,1]$, while $(\max\{X_1^*, \dots, X_n^*\} - b_n)/a_n \xrightarrow{\mathcal{D}} G$. The index ϑ is called the *extremal index* and expresses the degree of clustering of extreme values.

Very recently, the theory of extremes of identically distributed but dependent sequences has been significantly expanded by Herrmann *et al.* [59]. For such a sequence $(X_n)_{n\in\mathbb{N}}$, the dependency structure of the first n members can be encoded and investigated with help of copulas.

Definition 1.2.5. An n-dimensional *copula* is any distribution function $C\colon [0,1]^n \to [0,1]$ for which all marginal distributions are $U(0,1)$, i.e.,

$$\forall j = 1, \dots, n \ \forall u \in [0,1]\colon \quad C(\underbrace{1, \dots, 1}_{j-1 \text{ times}}, u, \underbrace{1, \dots, 1}_{n-j \text{ times}}) = u\,.$$

If F is an n-dimensional distribution function with continuous marginals $F_1, \dots, F_n$, then there is a unique copula C with $F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))\ \forall x_1, \dots, x_n \in \mathbb{R}$ by Sklar's theorem (see, e.g., [42]). Therefore, we say that C is *the copula* of F, and we see that copulas store information on the dependency structure of $F_1, \dots, F_n$.

For a sequence $(X_n)_{n\in\mathbb{N}}$ of random variables with common distribution function F, Herrmann *et al.* [59] consider the copulas C_n of the random vectors $(X_1, \dots, X_n)$ and make use of the fact that $\forall x \in \mathbb{R}\colon\ \mathbb{P}(M_n \le x) = \delta_n(F(x))$, where $\delta_n\colon [0,1] \to [0,1]$, $\delta_n(u) = C(u, \dots, u)$ denotes the *diagonal* of C. Their main result reads as follows:

Theorem 1.2.6. (see [59], Theorem 2.2)
Let $M_n := \max\{X_1, \dots, X_n\}$ for a sequence $(X_n)_{n\in\mathbb{N}}$ of identically distributed random variables with $X_1 \sim F$, and let $M_n^* := \max\{X_1^*, \dots, X_n^*\}$ for corresponding i.i.d. variables $X_1^*, X_2^*, \dots \sim F$. If $F \in \mathrm{MDA}(G)$ for an extreme value distribution G by means of $\mathbb{P}(M_n^* \le a_n^* x + b_n^*) \longrightarrow G(x)\ \forall x \in \mathbb{R}$, and if there exist a rate sequence *$(r_n)_{n\in\mathbb{N}}$ and a continuous function D on $[0,1]$ such that $\delta_n(u^{1/r_n}) \longrightarrow D(u)\ \forall u \in [0,1]$, then*

$$\forall x \in \mathbb{R}\colon \quad \mathbb{P}\left(M_n \le a^*_{\lceil r_n \rceil} x + b_{\lceil r_n \rceil}\right) \longrightarrow D(G(x))\,.$$

We now turn our attention to triangular arrays. As explained in Remark 1.1.12, almost all discrete distributions do not have a meaningful extreme value behavior for i.i.d. sequences. For these families of distributions, we instead aim to find such meaningful behavior within triangular arrays. This approach allows to consider an infinite family of probability distributions, which is not possible in the i.i.d. scenario. Families of discrete distributions are often parametrized by some real number, and limit theorems on extremes of these families often impose restrictions on the choice of the parameters. Formally, we consider a triangular array $(X_{nj})_{j=1,\dots,k_n}$ with identically distributed rows, where F_n denotes the joint distribution of the n-th row. Moreover, $M_n := \max\{X_{n1}, \dots, X_{nk_n}\}$ denotes the maxima of the rows. If there exist a non-degenerate distribution G and constants a_n, b_n such that $(M_n - b_n)/a_n \xrightarrow{\mathcal{D}} G$, then we adopt the terminology in Definition 1.1.1 and say that the family of $F_1, F_2, \dots$ is in the max-domain of attraction of G.

The question of classifying the non-degenerate limits of $(M_n - b_n)/a_n$ for triangular arrays has already been raised by Anderson *et al.* [2, p. 960]. However, this does not seem possible, particularly since there exist triangular arrays whose maxima are attracted to distributions other than the classical EVDs. For example, Panov and Morozova [80] studied the extremes of triangular arrays stemming from mixture models with heavy-tailed impurity. Within a certain regime of the impurity, the limit distribution turns out to be discontinuous, while all classical EVDs are continuous. The recent copula approach for dependent sequences in [59] may also help to better understand the extremes of certain triangular arrays.

Several efforts have been made to solve the issues explained in Remark 1.1.12 for common families of discrete distributions. For example, when considering uniform triangular arrays of the geometric distribution or the Poisson distribution, choosing the underlying parameter within a suitable regime leads to a Gumbel limit.

Theorem 1.2.7. (see [78], Theorem 1)
Let $X^*_{n1}, \dots, X^*_{nn} \sim \mathrm{Geo}(1/n)$ be i.i.d. geometric variables for all $n \in \mathbb{N}$ and let $M^*_n := \max\{X^*_{n1}, \dots, X^*_{nn}\}$. Then, $\mathbb{P}(M^*_n \leq n(x + \log n)) \longrightarrow \Lambda(x)$. More precisely,

$$\mathbb{P}(M^*_n \leq n(x + \log n)) - \Lambda(x) \sim \frac{e^{-x}\Lambda(x)\log(n)}{2n}.$$

The proof is based on the rate of convergence in Lemma 1.1.10 and a Taylor series of $\log(x)$ for $|x-1| < 1$. In [78], the statement of Theorem 1.2.7 was extended to a triangular array of *dependent* geometric variables that describe waiting times for Laplacian random variables. For i.i.d. $Z_{n1}, Z_{n2}, \dots \sim U(\{1, \dots, n\})$ and $j \in \{1, \dots, n\}$, let $X_{nj} := \min\{k \in \mathbb{N} : Z_{nk} = j\}$ be the waiting times of the events $\{Z_{nk} = j\}$.

Theorem 1.2.8. (see [78], Theorem 2)
The statement of Theorem 1.2.7 holds true for the row-wise dependent triangular array $(X_{nj})_{j=1,\dots,n}$, i.e., for $M_n = \max\{X_{n1}, \dots, X_{nn}\}$, we have

$$\mathbb{P}(M_n \leq n(x + \log n)) - \Lambda(x) = O\left(\frac{\log(n)}{n}\right).$$

The proof focuses on verifying the mixing conditions $\mathfrak{D}(u_n), \mathfrak{D}'(u_n)$ for $u_n := n(x + \log n)$ through extensive but straightforward calculations.

For some families of distributions specified by a real parameter, it is sufficient to state a minimum or maximum growth rate of this parameter. Anderson *et al.* [2] proved a Gumbel limit for a triangular array of Poisson variables $(R_{n,i})_{i=1,\dots,n} \sim \mathrm{Po}(\lambda_n)$, where the sequence $(\lambda_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$ satisfies a minimum growth rate.

Theorem 1.2.9. (see [2], Proposition 1)
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence with $\log(n) = o\left(\lambda_n^{(r+1)/(r+3)}\right)$ for some $r \in \mathbb{N}_0$. Let $R_{n1}, \dots, R_{nn} \sim \mathrm{Po}(\lambda_n)$ be a uniform row-wise i.i.d. triangular array and $M_n := \max\{R_{n1}, \dots, R_{nn}\}$. Then, there is a linear normalization $u_n(x) = \lambda_n + \sqrt{\lambda_n}\,(\beta^r_n + \alpha_n x)$ so that

$$\lim_{n\to\infty} \mathbb{P}\left(M_n \leq u_n(x)\right) = \exp(-\exp(-x)).$$

The sequences α_n, β_n are specified in [2, Section 3]. In the case of $r = 0$, the constants α_n, β_n are exactly those appearing in Theorem 1.1.11, and we have $\lambda_n = \omega(\log(n)^3)$. Moreover, Anderson *et al.* note in [2, Section 6] that if $\lambda_n = o(\log n)$, there exists no linear normalization to attract the triangular array to the Gumbel distribution or any other EVD. In this case, the triangular array is asymptotically too similar to i.i.d. sequences $X_1, X_2, \dots \sim \mathrm{Po}(\lambda)$ with fixed λ, which have no non-degenerate extreme value behavior.

Remark 1.2.10. As a consequence of Theorem 1.2.9, we note that for a triangular array $(X_{nj})_{j=1,\ldots,k_n}$ with $X_{n1} \sim \mathrm{Po}(n)$, the Gumbel attraction $\mathbb{P}(M_n \leq u_n(x)) \longrightarrow \exp(-e^{-x})$ with $u_n(x) = n + \sqrt{n}(\beta_n + \alpha_n x)$ is true for all $k_n = \exp(o(n^{1/3}))$. On the contrary, if k_n is too large, then this triangular array has no non-degenerate extreme value behavior for the aforementioned reasons. These considerations are not exclusive to the Poisson distribution. In [2, Section 4], it is argued that Theorem 1.2.9 can be generalized to other distributions represented as sums of i.i.d. random variables. However, the applicability of these observations is limited by the assumption of identical distribution. △

For the more general case of a row-wise stationary triangular array $(X_{nj})_{j=1,\ldots,k_n}$, there is a Gumbel EVLT due to Dkenge *et al.* [37] based on the mixing conditions introduced in Definitions 1.2.1 and 1.2.3. However, it additionally requires that all variables involved have an infinite right endpoint. This excludes discrete distributions on finitely many numbers.

Theorem 1.2.11. (see [37], Theorem 2.1)
Let $(X_{nj})_{j=1,\ldots,k_n}$ be a row-wise stationary triangular array, where μ_n and σ_n^2 denote the mean and variance of the n-th distribution. Moreover, let $u_n(x) := \alpha_n x + \beta_n$, where α_n, β_n are defined as in Theorem 1.1.11. We need the following conditions:

- *All X_{nj} have an infinite right endpoint.*
- *All X_{nj} have a moment generating function that exists in an open neighborhood of the origin.*
- $k_n = \exp\left(o(\sigma_n^{2/3})\right)$.
- *For $v_n = v_n(x) := \sigma_n u_{k_n}(x) + \mu_n$, both $\mathfrak{D}(v_n)$ and $\mathfrak{D}'(v_n)$ are satisfied.*

Then, for all $x \in \mathbb{R}$:

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_{n,k_n} - \mu_n}{\sigma_n} \leq u_{k_n}(x)\right) = \exp(-\exp(-x))\,.$$

For binomial distributions, a result similar to Theorem 1.2.9 is given by Nadarajah & Mitov [81]. The connection of the binomial distributions to the standard normal distribution is drawn by means of tail equivalence. We review [81, Theorem 3] and its proof in detail since it can be applied once we achieve tail equivalence for other distributions of interest.

In what follows, let $q := 1 - p$ for any $p \in (0,1)$ and let $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mathrm{d}x$ be the cumulative distribution function (CDF) of the standard normal distribution. At first, the tail equivalence of binomial distributions to the standard normal distribution is given as follows:

Theorem 1.2.12. (see Feller [47], p. 193)
Let $X_n \sim \mathrm{Bin}(n,p)$ for a fixed parameter $p \in (0,1)$ and let $Y_n := \frac{X_n}{\sqrt{npq}}$. Then, if $x_n = o(n^{1/6})$, we have $\mathbb{P}(Y_n > x_n) \sim 1 - \Phi(x_n)$.

Next, Nadarajah & Mitov show how to use tail equivalence combined with the extreme value behavior of the standard normal distribution. Recall the constants $\alpha_n = (2\log n)^{-1/2}$, $\beta_n = \alpha_n^{-1} - \alpha_n(\log\log n + \log(4\pi))/2$ from Theorem 1.1.11.

Theorem 1.2.13. **(see Nadarajah & Mitov [81], Theorem 3)**
Let $(X_{n1},\dots,X_{nn}) \overset{\text{i.i.d.}}{\sim} \mathrm{Bin}(N_n,p)$, *where* $p\in(0,1)$ *is fixed and* $N_n \gg \log(n)^3$. *Moreover, let* $a_n := \sqrt{pqN_n}\alpha_n$ *and* $b_n := pN_n + \sqrt{pqN_n}\beta_n$. *Then, we have* $\forall x\in\mathbb{R}$:

$$\mathbb{P}(M_n \le a_n x + b_n) \longrightarrow \exp(-\exp(-x))\,.$$

Proof. From Theorem 1.2.12, we know that if $N_n\to\infty$ and $x = x_n = o(N_n^{1/6})$, then

$$1 - F_n\left(pN_n + \sqrt{pqN_n}x\right) \sim 1 - \Phi(x)\,, \tag{1.3}$$

where F_n is the CDF of $\mathrm{Bin}(N_n,p)$. On the other hand, from Lemma 1.1.10 we know that

$$n(1-\Phi(\alpha_n x + \beta_n)) \longrightarrow e^{-x}.$$

In this limit process, x is treated as a constant. To plug the sequence $(\alpha_n x+\beta_n)_{n\in\mathbb{N}}$ in (1.3) with fixed x, we must ensure that $\alpha_n x + \beta_n = o(N_n^{1/6})$. Due to $\beta_n = \sqrt{2\log(n)} + o(1)$, we need $\log(n)^3 = o(N_n)$. Hence, combining both limit processes yields

$$n\left(1 - F_n\left(pN_n + \sqrt{pqN_n}x\right)\right) \sim n(1-\Phi(\alpha_n x+\beta_n)) \longrightarrow e^{-x}. \qquad \square$$

Remark 1.2.14. Interestingly, Isaev *et al.* [63, Lemma 5.1] note that Theorem 1.2.13 holds true not only for fixed p, but also for $p \longrightarrow 0$ or $p\longrightarrow 1$ under the condition $N_n pq \gg \log(n)^3$. Moreover, in some situations, Lemma 1.1.10 can be applied through direct analytical calculations, without making use of tail equivalence. See [12, 77] for examples.

1.3 Multivariate extreme value theory

The previous two sections introduced theory and examples for the extremes of *univariate* random variables. However, it is likewise interesting and useful to develop extreme value theory for *multivariate random vectors* in $d\ge 2$ dimensions. In many applications, it is important to know the dependence between extremes of different random quantities (see [41, 43] for examples). To develop such theory, it is necessary to specify what is meant by extremely large values in multiple dimensions, since there is no clear concept of ordering. The most useful and intuitive concept is the component-wise partial ordering $\mathbf{x}\le\mathbf{y} \iff \forall i = 1,\dots,d\colon x^{(i)} \le y^{(i)}$ for $\mathbf{x} = (x^{(1)},\dots,x^{(d)})^\top, \mathbf{y} = (y^{(1)},\dots,y^{(d)})^\top \in\mathbb{R}^d$. This gives a comprehensive theory, despite the fact that many vectors are not mutually comparable.

For any d-variate random vector $X = (X^{(1)},\dots,X^{(d)})^\top$, its CDF is $F\colon \mathbb{R}^d\to[0,1]$, $\mathbf{x}\mapsto \mathbb{P}(X\le\mathbf{x}) := \mathbb{P}(X^{(i)}\le x^{(i)})$. Its *marginal distributions* $F_1,\dots,F_d$ (or *marginals*, for short) are defined as

$$F_i\colon \mathbb{R}^d\to[0,1], \quad \mathbf{x}\mapsto F_i(\mathbf{x}) = \mathbb{P}(X^{(i)}\le x^{(i)})\,.$$

Definition 1.3.1. Let $X_n = (X_n^{(1)},\dots,X_n^{(d)})^\top \sim F$ be i.i.d. random vectors. For any $i=1,\dots,d$, let $M_n^{(i)} := \max_{k=1,\dots,n} X_k^{(i)}$. If there are constants $a_n^{(i)}>0, b_n^{(i)}\in\mathbb{R}$ such that

$$\mathbb{P}\left(\frac{M_n^{(i)} - b_n^{(i)}}{a_n^{(i)}} \le x^{(i)}\ \forall i\right) = F\left(a_n^{(1)}x^{(1)} + b_n^{(1)},\dots,a_n^{(d)}x^{(d)}+b_n^{(d)}\right)^n \longrightarrow G(\mathbf{x}) \tag{1.4}$$

for all $\mathbf{x} \in \mathbb{R}^d$, and if each marginal distribution of G is non-degenerate, then we call G a *multivariate extreme value distribution* (MEVD) and write $F \in \mathrm{MDA}(G)$, in analogy to the univariate setting. We can write (1.4) more concisely as

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq \mathbf{x}\right) \longrightarrow G(\mathbf{x})\,,$$

where $M_n := (M_n^{(1)}, \ldots, M_n^{(d)})^\top, a_n := (a_n^{(1)}, \ldots, a_n^{(d)})^\top, b_n := (b_n^{(1)}, \ldots, b_n^{(d)})^\top \in \mathbb{R}^d$, and all operations are taken component-wise.

Remark 1.3.2. Trivially, if (1.4) is true, then the marginal distributions F_i, G_i, $i = 1, \ldots, d$, must satisfy

$$\forall 1 \leq i \leq d\colon F_i^n\left(a_n^{(i)}\mathbf{x} + b_n^{(i)}\right) \longrightarrow G_i(\mathbf{x})\,. \tag{1.5}$$

This means that $F \in \mathrm{MDA}(G)$ implies $F_i \in \mathrm{MDA}(G_i)$ for all marginals. Due to Theorem 1.1.4, we can assume that all marginals G_i are either Gumbel, Fréchet, or Weibull distributions.

Interestingly, the marginals of G can be specified to be the Fréchet-1 distribution Φ_1 without restriction, by an according transformation of any $F \in \mathrm{MDA}(G)$.

Theorem 1.3.3. (see [91], Proposition 5.10)
Let G be a d-variate distribution function with continuous marginals. For $i = 1, \ldots, d$ and $\mathbf{x} \in \mathbb{R}^d$, let

$$\psi_i(\mathbf{x}) = \left(\frac{1}{-\log(G_i)}\right)^{\leftarrow}(x)\,, \qquad G_*(\mathbf{x}) := G(\psi_1(x_1), \ldots, \psi_d(x_d))\,.$$

Then, G_* has Φ_1 marginals and is an MEVD if G is an MEVD. For a distribution function F with marginals $F_1, \ldots, F_d$, let $U_i := 1/(1 - F_i)$ and let F_* be the CDF of $\left(U_1(X_1^{(1)}), \ldots, U_d(X_1^{(d)})\right)^\top$, such that

$$F_*(\mathbf{x}) = F\left(U_1^{\leftarrow}(x^{(1)}), \ldots, U_d^{\leftarrow}(x^{(d)})\right).$$

If (1.4) is satisfied, then $F_* \in \mathrm{MDA}(G_*)$ and

$$\mathbb{P}\left(\max_{j=1,\ldots,n} \frac{U_i(X_j^{(i)})}{n} \leq x^{(i)}, 1 \leq i \leq d\right) = F_*^n(n\mathbf{x}) \longrightarrow G_*(\mathbf{x})\,. \tag{1.6}$$

Conversely, if (1.5) and (1.6) hold, and G_* has non-degenerate marginals, then (1.4) holds as well.

Using this standardization, it is possible to obtain a description of MDAs in the multivariate case. For the classification theorem 1.1.8 in the univariate case, the concept of slow variation and regular variation was essential (see Definition 1.1.7). It is possible to specify a multivariate analogue.

Definition 1.3.4. Let $C \subseteq \mathbb{R}^d$ be a cone, i.e., $\mathbf{x} \in C \iff t\mathbf{x} \in C\ \forall t > 0$, with $\mathbf{1} \in C$. A measurable function $h\colon C \to (0, \infty)$ is said to be *regularly varying* with *limit function* $\lambda\colon C \to (0, \infty)$ if $\lambda(\mathbf{1}) = 1$ and

$$\forall \mathbf{x} \in C\colon \lim_{t\to\infty} \frac{h(t\mathbf{x})}{h(t\mathbf{1})} = \lambda(\mathbf{x})\,.$$

Equivalently, h is λ-regularly varying if and only if there exists a regularly varying function $V\colon (0,\infty) \to (0,\infty)$ such that

$$\forall \mathbf{x} \in C\colon \lim_{t\to\infty} \frac{h(t\mathbf{x})}{V(t)} = \lambda(\mathbf{x})\,.$$

Theorem 1.3.5. (see [91], Proposition 5.15)
For a distribution function F and an MEVD G, let G_*, U_i, F_* be as in Theorem 1.3.3. Then, the following hold:

(a) $F_* \in \mathrm{MDA}(G_*)$ *if and only if* $1 - F_*$ *is regularly varying on* $(0,\infty)^d$ *with limit* $\lambda(\mathbf{x}) = -\log G_*(\mathbf{x})/-\log G_*(\mathbf{1})$. *Precisely,*

$$\lim_{t\to\infty} \frac{1 - F_*(t\mathbf{x})}{1 - F_*(t\mathbf{1})} = \frac{-\log G_*(\mathbf{x})}{-\log G_*(\mathbf{1})}\,,$$

where $\mathbf{1} := (1,\dots,1)^\top$.

(b) $F \in \mathrm{MDA}(G)$ *if and only if* $F_* \in \mathrm{MDA}(G_*)$ *and* (1.5) *holds.*

While this is a straightforward way to derive the asymptotic behavior of multivariate extreme values, there are further methods that sometimes prove useful. As highlighted above, it is necessary that the marginals converge to a univariate EVD, for which we have a complete classification. The difficulty of the multivariate scenario arises from the dependencies between different marginals. We have already seen in Definition 1.2.5 that these dependencies are elegantly represented by copulas, and that every continuous distribution function has a unique associated copula by Sklar's theorem. Therefore, Theorem 1.3.3 expresses a transformation of only the marginals, but not of the associated copula. In fact, the underlying copula is invariant under strictly increasing transformations.

Definition 1.3.6. If G is a d-variate MEVD, then its copula C is called an *extreme value copula*. Some terminology from univariate extreme value theory can be transferred to copulas. Equivalently, we say that C is an extreme value copula if there exists a copula C_0 for which

$$\lim_{n\to\infty} C_0(u^{1/n})^n = C(u) \qquad \forall u \in [0,1]^d.$$

In that case, C_0 is in the *copula domain of attraction* of C, and we write $C_0 \in \mathrm{CDA}(C)$. Furthermore, extreme value copulas can be characterized by max-stability, i.e., $C(u) = C(u^{1/m})^m$ for all $u \in [0,1]^d$ and $m \in \mathbb{N}$. △

We now obtain a multivariate analogue of the Fisher–Tippett–Gnedenko theorem. However, a convenient classification of MEVDs remains impossible due to their diversity.

Theorem 1.3.7. (see [33], Theorem 3.1)
Let F be a d-variate distribution with copula C_0 and let G be a d-variate extreme value distribution with copula C. Then, $F \in \mathrm{MDA}(G)$ if and only if the marginals satisfy (1.6) and $C_0 \in \mathrm{CDA}(C)$.

There are many different approaches to describe multivariate EVDs, most notably, spectral measures, tail dependence coefficients, stable tail dependence functions and Pickands dependence functions (the latter is a concept only in two dimensions). See, e.g., [43, 49, 92, 96] for introductions of these concepts. We will not go into the details, but in light of the univariate normal distribution being attracted to the Gumbel distribution by Theorem 1.1.11, we now explain the extreme value behavior of multivariate normal distributions. The two-dimensional case is particularly important.

Definition 1.3.8. Let X_1, X_2 be two random variables. The *upper tail dependence coefficient* of X_1 and X_2 is

$$\lambda_u(X_1, X_2) := \lim_{q \to 1-} \mathbb{P}(X_2 > F_2^{\leftarrow}(q) \mid X_1 > F_1^{\leftarrow}(q)),$$

provided this limit exists. It indicates the asymptotic probability of exceeding a high quantile in one component, given an exceedance in the other component. However, this quantity does not depend on the marginal distributions, but only on the dependency structure embodied by the copula.

Two random variables X_1, X_2 are called *asymptotically independent if* $\lambda_u(X_1, X_2) = 0$. On the contrary, if $\lambda_u(X_1, X_2) > 0$, then X_1 and X_2 are called *asymptotically dependent*. A remarkable fact is that the marginals of bivariate normal distributions are always asymptotically independent if they are not perfectly correlated, regardless of the degree of deviation from perfect correlation. This is known due to [99, Theorem 3].

Remark 1.3.9. Let $X = (X_1, X_2)$ be bivariate normal with $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $i = 1, 2$ and let $\rho := \operatorname{corr}(X_1, X_2)$. The copula of X, which depends on ρ, is denoted by $C_\rho \sim (U_1, U_2)$. Trivially, if $\rho = 1$, then $\lambda_u = 1$, and if $\rho = -1$, then $\lambda_u = 0$. Now, we assume $\rho \in (-1, 1)$ and write

$$\lambda_u = \lim_{q \to 1-} \mathbb{P}(U_2 > q \mid U_1 = q) + \mathbb{P}(U_1 > q \mid U_2 = q). \tag{1.7}$$

For the first summand, we have

$$\begin{aligned} \mathbb{P}(U_2 > q \mid U_1 = q) &= \mathbb{P}\left(F_2^{-1}(U_2) > F_2^{-1}(q) \;\middle|\; F_1^{-1}(U_1) = F_1^{-1}(q)\right) \\ &= \mathbb{P}\left(X_2 > \mu_2 + \sigma_2 \Phi^{-1}(q) \;\middle|\; X_1 = \mu_1 + \sigma_1 \Phi^{-1}(q)\right). \end{aligned}$$

Under the condition $X_1 = \mu_1 + \sigma_1 x$, we find that $(X_2 - \mu_2)/\sigma_2$ is normally distributed with mean ρx and variance $1 - \rho^2$. It follows that

$$\begin{aligned} \lim_{q \to 1-} \mathbb{P}(U_2 > q \mid U_1 = q) &= \lim_{x \to \infty} \mathbb{P}\left(\frac{X_2 - \mu_2}{\sigma_2} > x \;\middle|\; X_1 = \mu_1 + \sigma_1 x\right) \\ &= \lim_{x \to \infty} \Phi\left(\frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} x\right) = 0. \end{aligned}$$

For reasons of symmetry, the second summand in (1.7) behaves identically. This implies that the Gaussian copula C_ρ is in the CDA of the independence copula $C_0(x_1, x_2) = x_1 x_2$. Theorem 1.3.7 therefore gives:

Theorem 1.3.10. *The bivariate normal distribution* $\mathcal{N}_\rho \sim \mathrm{N}_2\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ *with* $\rho \in [-1, 1)$ *is in the MDA of the bivariate Gumbel distribution with independent marginals, namely,*

$$\Lambda_2(\mathbf{x}) := \exp\left(-e^{-x_1} - e^{-x_2}\right) = \Lambda(x_1)\Lambda(x_2), \qquad \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2.$$

According to (1.5)*, by taking the constant vectors* $\boldsymbol{\alpha}_n := (\alpha_n, \alpha_n)$ *and* $\boldsymbol{\beta}_n := (\beta_n, \beta_n)$ *with* α_n, β_n *as in Theorem* 1.1.11*, we have*

$$\frac{M_n - \boldsymbol{\beta}_n}{\boldsymbol{\alpha}_n} \xrightarrow{\mathcal{D}} \Lambda_2 \,,$$

where M_n *is the component-wise maximum of i.i.d. samples of* $\mathcal{N}_\rho$*.*

Figure 1.6 provides a visualization of Λ_2.

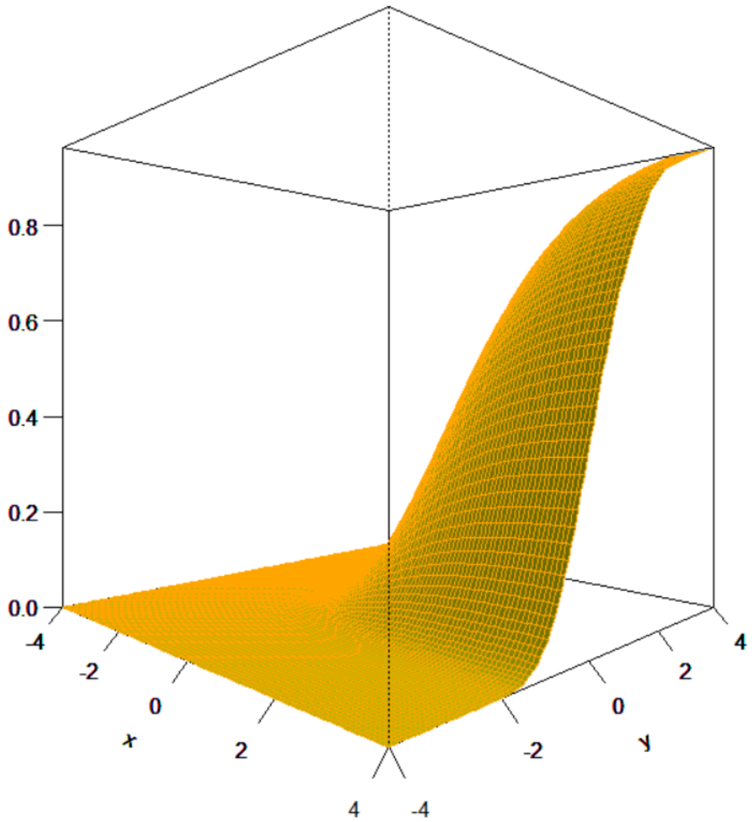


Figure 1.6: Plot of the two-dimensional Gumbel distribution $\Lambda_2(x, y)$ with independent marginals, for $-4 \leq x, y \leq 4$.

2 Finite Coxeter groups and permutation statistics

The first chapter gave an introduction to EVLTs for i.i.d. sequences, triangular arrays, and multivariate settings. Now, we present the objects for which we aim to find new EVLTs, namely, the numbers of inversions and descents on symmetric groups and, more generally, on finite Coxeter groups. These numbers are typical examples of *permutation statistics,* i.e., families of maps from permutation groups to the non-negative integers. They can be considered as random variables once the underlying group is equipped with a probability measure. Section 2.1 is devoted to the classification of finite Coxeter groups and the numbers of inversions and descents on them. Section 2.2 gives basic stochastic properties of these random variables. Finally, Section 2.3 outlines the history of the CLT for random inversions and descents on finite Coxeter groups.

2.1 Finite Coxeter groups, inversions and descents

Definition 2.1.1. Let S_n denote the *symmetric group* on n elements, i.e., the group of permutations on the set $\{1, \ldots, n\}$ with composition as group operation. We denote a permutation $\pi \in S_n$ using *in-line notation*:

$$\pi = (\pi(1), \pi(2), \ldots, \pi(n)) .$$

It is well known that each permutation decomposes into disjoint cycles, which in turn decompose into transpositions, which finally decompose into transpositions of neighboring numbers (*neighboring transpositions,* for short). For $i = 1, \ldots, n-1$, write τ_i for the neighboring transposition that permutes i and $i+1$. Then, the system $\{\tau_1, \ldots, \tau_{n-1}\}$ is a minimal generator of S_n.

Definition 2.1.2. Let $\pi \in S_n$ be a permutation. An *inversion* of π is any pair (i, j) with $i < j$ and $\pi(i) > \pi(j)$. The name suggests that π inverts the order of these two numbers, which is visualized in Figure 2.1. The set of inversions is denoted by $\mathrm{Inv}(\pi)$, and its cardinality is the *number of inversions* $\mathrm{inv}(\pi)$.

Definition 2.1.3. A *descent* of $\pi \in S_n$ is an inversion of two adjacent numbers, i.e., any number $i \in \{1, \ldots, n-1\}$ with $\pi(i) > \pi(i+1)$. Figure 2.2 visualizes descents as the starting points of descending segments between the points $(1, \pi(1)), (2, \pi(2)), \ldots, (n, \pi(n))$. The set of descents is denoted by $\mathrm{Des}(\pi)$, and its cardinality is the *number of descents* $\mathrm{des}(\pi)$. The *Eulerian numbers* are the numbers of permutations with a fixed number $k \in \{1, \ldots, n-1\}$ of descents:

$$\left\langle {n \atop k} \right\rangle := |\{\pi \in S_n \colon \mathrm{des}(\pi) = k\}| \, .$$

An extensive review of Eulerian numbers and related concepts is given by Petersen [84].

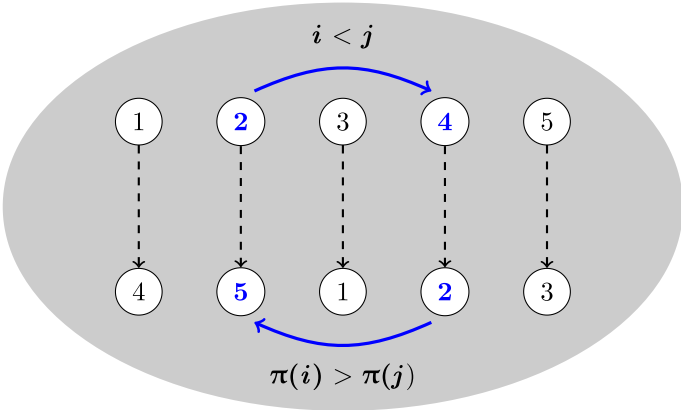


Figure 2.1: Example of a permutation $\pi = (4, 5, 1, 2, 3)$ on five numbers, highlighting an inverted pair: $2 < 4$, but $\pi(i) = 5 > 2 = \pi(j)$.

Remark 2.1.4. Let e denote the identity map of S_n. The symmetric group S_n has the $n-1$ generators $\tau_1, \ldots, \tau_{n-1}$. All of these are self-inverse and it can be easily checked that $\tau_i\tau_{i+1}\tau_i = \tau_{i+1}\tau_i\tau_{i+1} \iff (\tau_i\tau_{i+1})^3 = e$ for all $i \in \{1, \ldots, n-2\}$, and $\tau_i\tau_j = \tau_j\tau_i \iff (\tau_i\tau_j)^2 = e$ for all non-adjacent i, j, i.e., $|i - j| \geq 2$. This information is sufficient to characterize the structure of S_n. The aforementioned relations between the generators $\tau_1, \ldots, \tau_{n-1}$ give a presentation of S_n. This shows that symmetric groups are part of a larger family of groups, namely, the so-called Coxeter groups.

Definition 2.1.5. Let $S = \{s_1, s_2, \ldots\}$ be an at most countable set. A symmetric matrix $M\colon S \times S \longrightarrow \{1, 2, \ldots, \infty\}$ is called a *Coxeter matrix* if $M(s_1, s_2) = 1 \iff s_1 = s_2$, meaning that ones are placed on the principal diagonal and nowhere else. A Coxeter matrix M gives rise to a *Coxeter group* W via the following presentation:

$$W = \langle S \mid (s_1 s_2)^{M(s_1,s_2)} = e \quad \text{for all } (s_1, s_2) \in S^2 \text{ with } M(s_1, s_2) < \infty \rangle \,. \tag{2.1}$$

This means that the Coxeter matrix contains all information about the structure of the corresponding Coxeter group, since there are no other relations by definition, and since there is no smaller generating set of W. The matrix entries $M(s_1, s_2)$ are exactly the orders of the products $s_1 s_2$, see [8, Proposition 1.1.1b)].

A pair (W, S) of a Coxeter group W and its generating system S satisfying (2.1) is called a *Coxeter system*. The underlying Coxeter matrix M is usually suppressed in the notation. The cardinality of S is called the *rank* of W and it is denoted by $\mathrm{rk}(W)$.

Definition 2.1.6. Let (W, S) be a Coxeter system based on a Coxeter matrix M. It is convenient to represent W as an undirected *Coxeter graph*, according to the following rules:

- Each generator gives a node.
- If $M(s_i, s_j) = 2$, there is no edge between s_i and s_j. In this case, $s_i s_j$ is self-inverse, so s_i and s_j commute.
- If $M(s_i, s_j) = 3$, there is an unlabeled edge between s_i and s_j.
- If $M(s_i, s_j) \geq 4$, there is an edge between s_i and s_j that is labeled with $M(s_i, s_j)$.

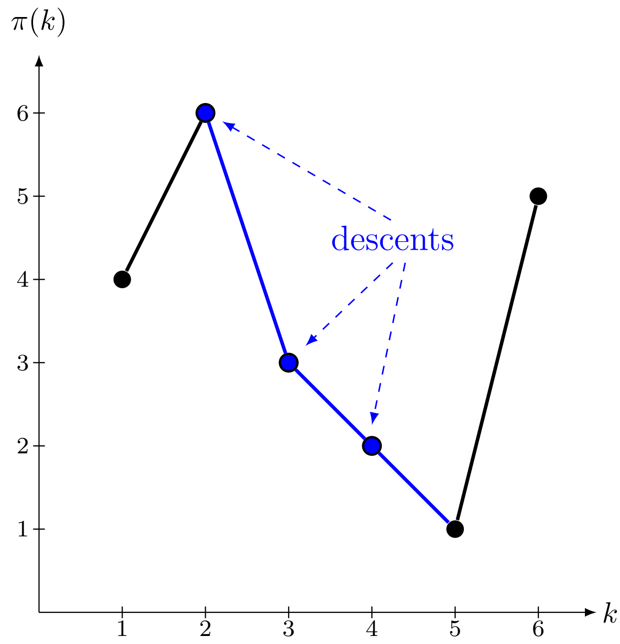


Figure 2.2: Graphical display of descents (highlighted in blue) of the permutation $\pi = (4, 6, 3, 2, 1, 5)$. The arrows emphasize the convention that the points at the beginning of descending segments are referred to as descents.

Remark 2.1.7. Products of Coxeter groups are again Coxeter groups. If (W_1, S_1) and (W_2, S_2) are two Coxeter systems with Coxeter matrices M_1, M_2, then $W := (W_1 \times W_2, S_1 \cup S_2)$ is a Coxeter system with the Coxeter matrix

$$M(s_i, s_j) = \begin{cases} M_1(s_i, s_j), & s_i, s_j \in S_1 \\ M_2(s_i, s_j), & s_i, s_j \in S_2 \\ 2, & \text{otherwise} \end{cases}, \qquad s_i, s_j \in S_1 \cup S_2 \,.$$

Definition 2.1.8. A Coxeter group or a Coxeter system is called *irreducible* if it is not a Cartesian product of smaller Coxeter groups. This is the case if and only if its Coxeter graph is connected.

Remark 2.1.9. There is a complete classification of finite irreducible Coxeter groups, which is given by [28]. It consists of:

- three families of groups called A_n, B_n, D_n, which will be explained in the following definitions. The index n corresponds to the rank of these groups. For simplicity, we refer to the entirety of these three families as *classical Weyl groups* since they are the Weyl groups of the classical groups (see, e.g., [106]).
- the family of *dihedral groups* $I_2(m) = \langle r, s \mid r^2, s^2, (rs)^m \rangle$, also known as the isometry groups of the planar regular m-gons. Here, r and s represent reflections of the m-gon by two symmetry axes that intersect at an angle of π/m.
- six exceptional groups, which are commonly known as $E_6, E_7, E_8, F_4, H_3, H_4$. They also appear as isometry groups of certain geometric structures. While we will not pay special attention to these groups, they will be implicitly covered in the results concerning the entirety of finite Coxeter groups.

Definition 2.1.10. The groups A_n can be introduced as the isometry groups of the n-simplex, which means each A_n is isomorphic to the symmetric group S_{n+1}. Note that A_n has rank n, which is one less than the cardinality of the underlying set for S_{n+1}. According to Remark 2.1.4, the Coxeter matrix of A_n is:

$$M(i,j) = \begin{cases} 1, & i = j \\ 3, & |i-j| = 1 \\ 2, & |i-j| \geq 2 \end{cases}.$$

According to Definition 2.1.6, the Coxeter graph of A_n is a straight path on the set $\{1, 2, \ldots, n\}$, as seen in Figure 2.3.

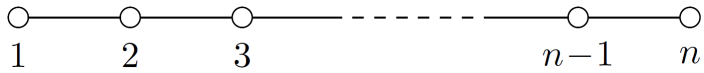


Figure 2.3: The Coxeter graph of $A_n = S_{n+1}$.

However, we will consistently keep referring to symmetric groups as S_n, since this is more suitable for the methods and concepts used in this thesis. When speaking of classical Weyl groups of rank n, we also refer to the groups S_n, B_n, D_n by convention, even though S_n has rank $n-1$.

Definition 2.1.11. The groups B_n can be introduced as the isometry groups of the n-hypercube, but we are interested in their combinatorial interpretation as an extension of the symmetric groups. For $\pi \in S_n$, the entries $\pi(1), \ldots, \pi(n)$ are each given a positive or negative sign, which yields a *signed permutation*. Therefore, we call B_n a *signed permutation group*. Figure 2.4 illustrates this with an example. △

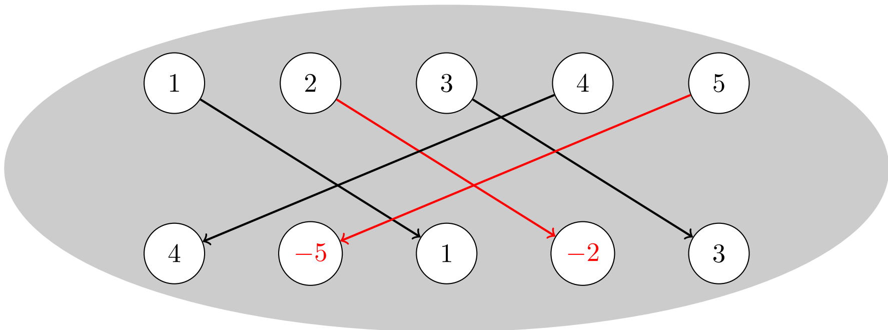


Figure 2.4: Example of a signed permutation derived from the permutation illustrated in Figure 2.1. This signed permutation can be written as $(4, -5, 1, -2, 3)$. The negative signs are marked in red.

The in-line notation used for symmetric groups can be adopted for signed permutation groups. We can also describe elements of B_n as permutations on $\{-n, \ldots, -1, 1, \ldots, n\}$ satisfying the antisymmetry constraint $\pi(i) = -\pi(i)$. For another combinatorial interpretation, see [8, Example 1.2.4]. The group B_n can be generated by taking the generators $\{\tau_1, \ldots, \tau_{n-1}\}$ of S_n and adding the element τ_0 which inverts the sign of the first entry. Due to $(\tau_0\tau_1)^2 = (\tau_1\tau_0)^2 = (-1, -2, 3, \ldots, n)$ being self-inverse, both $\tau_0\tau_1$ and $\tau_1\tau_0$ have order 4. The Coxeter graph of B_n is shown in Figure 2.5.

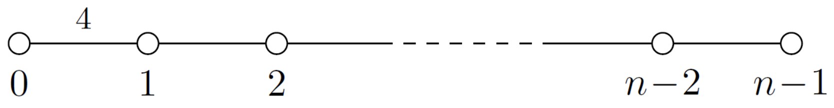


Figure 2.5: The Coxeter graph of the signed permutation group B_n. The leftmost node represents the additional generator τ_0.

Definition 2.1.12. The *even-signed permutation groups* D_n are the subgroups of B_n consisting of all signed permutations with an even number of negative signs. D_n is generated by $\{\tau_1, \ldots, \tau_{n-1}, \tilde{\tau}_0\}$, where $\tilde{\tau}_0 = (-2, -1, 3, \ldots, n)$. The Coxeter graph of D_n is shown in Figure 2.6. It is easy to verify the cubic relations between $\tilde{\tau}_0$ and τ_1, τ_2. △

Figure 2.6: The Coxeter graph of the even-signed permutation group D_n. The top node represents the additional generator $\tilde{\tau}_0$.

We now explain how to generalize the numbers of inversions and descents to Coxeter groups. Recall that symmetric groups are generated by neighboring transpositions. Multiplying such a neighboring transposition from the right either creates a new descent or cancels an existing one. Likewise, inversions are created or canceled by multiplying general transpositions from the right. Moreover, every conjugate of a neighboring transposition is a general transposition. For this reason, we introduce the concept of word length and reflections in order to generalize inversions and descents from symmetric groups to Coxeter groups.

Definition 2.1.13. Let (W, S) be a Coxeter system. Each $w \in W$ has a shortest representation $w = s_1 \ldots s_k$ with $s_1, s_2, \ldots \in S$. We call $k = l(w)$ the *word length* of w. Moreover, for $w \in W, s \in S$, the conjugate $t := wsw^{-1}$ is called a *reflection*. Let $T \subseteq W$ be the set of reflections. The (right) *inversions* of w are all $t \in T$ with $l(wt) < l(w)$, and the (right) *descents* of w are all $s \in S$ with $l(ws) < l(w)$. This also explains the quantities $\mathrm{inv}(w)$ and $\mathrm{des}(w)$.

Remark 2.1.14. On the symmetric group S_n, the number of inversions is equal to the word length. For $\pi = (\pi(1), \ldots, \pi(n)) \in S_n$, we see that $n - \pi^{-1}(n)$ indicates the number of inversions (i, j) with $i = n$, as well as the number of neighboring transpositions needed to shift n to its position. This argument is continued recursively. △

In the following, we explain the generalization of inversions and descents on the groups B_n and D_n, and again note that $\mathrm{inv}(\cdot)$ is equal to $l(\cdot)$.

Remark 2.1.15. On the groups B_n of signed permutations, using in-line notation, we count inversions as

$$\mathrm{inv}(\pi) = |\mathrm{Inv}^+(\pi)| + |\mathrm{Inv}^-(\pi)| + |\mathrm{Inv}^\circ(\pi)|\,,$$

where

$$\begin{aligned}
\mathrm{Inv}^+(\pi) &:= \{1 \le i < j \le n \mid \pi(i) > \pi(j)\}\,,\\
\mathrm{Inv}^-(\pi) &:= \{1 \le i < j \le n \mid -\pi(i) > \pi(j)\}\,,\\
\mathrm{Inv}^\circ\,(\pi) &:= \{1 \le i \le n \mid \pi(i) < 0\}\,.
\end{aligned}$$

The set $\mathrm{Inv}^+(\cdot)$ is analogous to inversions on symmetric groups. Note that on B_n and D_n, one has to pay attention to signs, i.e., a pair (i,j) with $\pi(i), \pi(j) < 0$ and $|\pi(i)| < |\pi(j)|$ also adds to $\mathrm{Inv}^+(\pi)$. Obviously, if $\pi \in S_n$, then the other two quantities $|\mathrm{Inv}^-(\cdot)|$ and $|\mathrm{Inv}^\circ(\cdot)|$ vanish. These two quantities are required so that $|\mathrm{inv}(\pi)|$ equals the word length on B_n with respect to the generating system $\{\tau_0, \tau_1, \ldots, \tau_{n-1}\}$ introduced in Definition 2.1.11. For details, see the proof of [8, Proposition 8.1.1].

On the groups D_n of even-signed permutations, inversions are counted similarly, except that we now have to omit the number of negative signs. That is,

$$\mathrm{inv}(\pi) = |\mathrm{Inv}^+(\pi)| + |\mathrm{Inv}^-(\pi)|\,.$$

The proof of [8, Proposition 8.2.1] shows that this matches the word length on D_n.

Remark 2.1.16. The number of descents on B_n can be written as follows. Expand the in-line notation by setting $\pi(0) := 0$. Then,

$$\mathrm{des}(\pi) = \sum_{i=0}^{n-1} \mathbf{1}\{\pi(i) > \pi(i+1)\}\,,$$

and on the even-signed permutations D_n, we set $\pi(0) := -\pi(2)$, giving

$$\mathrm{des}(\pi) = \sum_{i=1}^{n-1} \mathbf{1}\{\pi(i) > \pi(i+1)\} + \mathbf{1}\{-\pi(2) > \pi(1)\}\,.$$

It is easy to verify that these representations give the number of descents on B_n and D_n according to Definition 2.1.13, see [8, Propositions 8.1.2 and 8.2.2].

Definition 2.1.17. For inversions and descents on a Coxeter group W, we have the *generating functions*

$$\mathcal{G}_{\mathrm{inv}}(W;z) := \sum_{w \in W} z^{\mathrm{inv}(w)}\,, \qquad \mathcal{G}_{\mathrm{des}}(W;z) := \sum_{w \in W} z^{\mathrm{des}(w)}\,.$$

Obviously, these are polynomials with natural-numbered coefficients. The generating function of inversions is known as the *Mahonian polynomial*, while the generating function of descents is known as the *Eulerian polynomial*.

Remark 2.1.18. To decompose the Mahonian polynomial of a finite Coxeter group W, we need quantities known as the *degrees of fundamental invariants,* or simply, the *degrees* of W. Each finite Coxeter group is associated with a canonical action, e.g., S_n permutes the coordinates of $\mathbb{C}^n$, and B_n can additionally change their signs. Some polynomials are invariant under this action. Their degrees are the so-called *degrees of fundamental invariants* of W. For the main families of finite irreducible Coxeter groups, the degrees are as follows:

- S_n has degrees $2, 3, \ldots, n$.
- B_n has degrees $2, 4, \ldots, 2n$.
- D_n has degrees $2, 4, \ldots, 2n-2$ and n.
- $I_2(m)$ has degrees $2, m$.

In this thesis, we consistently denote the largest degree of a finite Coxeter group by $d_{\max}$, suppressing the rank n.

Theorem 2.1.19. (see [8], Theorem 7.1.5)
Let W be a finite Coxeter group with $\mathrm{rk}(W) = n$. Then,

$$\mathcal{G}_{\mathrm{inv}}(W; z) = \prod_{i=1}^{n} (1 + z + \ldots + z^{d_i - 1}),$$

where $d_1, \ldots, d_n$ are the degrees of W.

The generating function $\mathcal{G}_{\mathrm{des}}$ also has a decomposition, even into linear factors. This was proved by Brenti [14] for all irreducible finite Coxeter groups except for the groups D_n, which were handled by Savage & Visontai [97]. They proved that the Eulerian polynomial of these groups is real-rooted. From this, it is trivial to conclude that the roots are negative, since all coefficients of the Eulerian polynomial are positive.

Theorem 2.1.20. *Let W be a finite Coxeter group with $\mathrm{rk}(W) = n$. Then, $\mathcal{G}_{\mathrm{des}}(W; z)$ has only negative roots, i.e.,*

$$\mathcal{G}_{\mathrm{des}}(W; z) = \prod_{i=1}^{n} (z + q_i)$$

for some $q_1, \ldots, q_n > 0$.

2.2 Stochastic properties

Let (W, S) be a finite Coxeter system. Then, the discrete uniform distribution on the power set $\mathcal{P}(W)$ allows us to regard W as a probability space. Moreover, we now interpret $X_{\mathrm{inv}}(w) := \mathrm{inv}(w)$, $X_{\mathrm{des}}(w) := \mathrm{des}(w)$ as random variables on this probability space. This notation is used throughout to emphasize the stochastic background. On the symmetric groups S_n, the probability distribution of X_{inv} is known as the *Mahonian distribution* and that of X_{des} is known as the *Eulerian distribution*. Both have been studied in [5]. Moreover, on the classical Weyl groups, it is possible to give purely probabilistic representations of X_{inv} and X_{des} as follows:

Remark 2.2.1. Let $Z_1, Z_2, \dots, Z_n \sim U(0,1)$ be i.i.d. random variables with order statistics $Z_{(1)} < Z_{(2)} < \dots < Z_{(n)}$. Let $\pi = (\pi(1), \dots, \pi(n)) \in S_n$ be the unique permutation with $Z_i = Z_{(\pi(i))}$ $\forall i = 1, \dots, n$. This procedure yields a uniform random permutation. Therefore, X_{inv} and X_{des} can be expressed by

$$X_{\mathrm{inv}} = \sum_{1 \le i < j \le n} \mathbf{1}\{Z_i > Z_j\}, \qquad X_{\mathrm{des}} = \sum_{i=1}^{n-1} \mathbf{1}\{Z_i > Z_{i+1}\}. \tag{2.2}$$

On the signed and even-signed permutation groups B_n and D_n, we can analogously represent X_{inv} and X_{des} through i.i.d. variables $Z_1, \dots, Z_n \sim U(-1,1)$. Let X^B_{inv} and X^D_{inv} denote X_{inv} on B_n and D_n, respectively. From Remark 2.1.15, it follows that

$$\begin{aligned}
X^B_{\mathrm{inv}} &= \sum_{1 \le i < j \le n} \mathbf{1}\{Z_i > Z_j\} + \sum_{1 \le i < j \le n} \mathbf{1}\{-Z_i > Z_j\} + \sum_{i=1}^{n} \mathbf{1}\{Z_i < 0\}, \\
X^D_{\mathrm{inv}} &= \sum_{1 \le i < j \le n} \mathbf{1}\{Z_i > Z_j\} + \sum_{1 \le i < j \le n} \mathbf{1}\{-Z_i > Z_j\}.
\end{aligned}$$

Accordingly, let X^B_{des} and X^D_{des} denote X_{des} on B_n and D_n, respectively. From Remark 2.1.16, it follows that

$$\begin{aligned}
X^B_{\mathrm{des}} &= \sum_{k=1}^{n-1} \mathbf{1}\{Z_k > Z_{k+1}\} + \mathbf{1}\{Z_1 < 0\}, \\
X^D_{\mathrm{des}} &= \sum_{k=1}^{n-1} \mathbf{1}\{Z_k > Z_{k+1}\} + \mathbf{1}\{-Z_2 > Z_1\}.
\end{aligned}$$

Important stochastic quantities such as the mean and variance can be computed. For irreducible finite Coxeter groups, a summary of these is given in [66].

Theorem 2.2.2. **(see [66], Theorem 3.1 and Corollary 3.2)**
For any irreducible finite Coxeter group W of rank n, the random number of inversions X_{inv} has the mean and variance

$$\mathbb{E}(X_{\mathrm{inv}}) = \frac{1}{2}\sum_{k=1}^{n}(d_k - 1), \qquad \mathrm{Var}(X_{\mathrm{inv}}) = \frac{1}{12}\sum_{k=1}^{n}(d_k^2 - 1),$$

where $d_1, \dots, d_n$ are the degrees of W. In particular,

$$\mathbb{E}(X_{\mathrm{inv}}) = \begin{cases} n(n-1)/4, & W = S_n \\ n^2/2, & W = B_n \\ n(n-1)/2, & W = D_n \\ m/2, & W = I_2(m) \end{cases},$$

$$\mathrm{Var}(X_{\mathrm{inv}}) = \begin{cases} (2n^3 + 3n^2 - 5n)/72, & W = S_n \\ (4n^3 + 6n^2 - n)/36, & W = B_n \\ (4n^3 - 3n^2 - n)/36, & W = D_n \\ (m^2 + 2)/12, & W = I_2(m) \end{cases}.$$

Theorem 2.2.3. **(see [66], Theorem 4.1 and Corollary 4.2)**
For any irreducible finite Coxeter group W of rank n, the random number of descents X_{des} has the mean and variance

$$\mathbb{E}(X_{\text{des}}) = n/2\,, \qquad\qquad \operatorname{Var}(X_{\text{des}}) = \frac{n-2}{12} + \frac{1}{m}\,,$$

where m denotes half of the size of the largest dihedral subgroup of W. In particular,

$$\operatorname{Var}(X_{\text{des}}) = \begin{cases} (n+1)/12, & W = S_n \\ (n+1)/12, & W = B_n \\ (n+2)/12, & W = D_n \\ 1/m, & W = I_2(m) \end{cases}.$$

Remark 2.2.4. Let $W = W_1 \times W_2$ be a product of two Coxeter groups and let X_W, X_{W_1}, X_{W_2} denote the random number of either inversions or descents on these groups. Then, we have that $X_W = X_{W_1} + X_{W_2}$, and the two summands are independent. In particular:

(a) If $W = \prod_{i=1}^k I_2(m_i)$ is a product of dihedral groups, then $\operatorname{Var}(X_{\text{des}}) = \sum_{i=1}^k m_i^{-1}$.
(b) If $(W_n)_{n\in\mathbb{N}}$ are finite Coxeter groups without dihedral components, then we have $\operatorname{Var}(X_{\text{des}}) = \Theta(\operatorname{rk}(W_n))$. △

Any factorization of a generating function yields an independent sum decomposition of the corresponding statistic. In conclusion, due to the decompositions of $\mathcal{G}_{\text{inv}}$ and $\mathcal{G}_{\text{des}}$ in Theorems 2.1.19 and 2.1.20, X_{inv} and X_{des} can be written as sums of independent (but not identically distributed) variables.

Corollary 2.2.5. *Let W be a finite Coxeter group with $\operatorname{rk}(W) = n$. Then:*

(a) $X_{\text{inv}} = \sum_{i=1}^n X_{\text{inv}}^{(i)}$, *where* $X_{\text{inv}}^{(i)} \sim U\left\{0, 1, \ldots, d_i^{(n)} - 1\right\}$ *and $d_1, \ldots, d_n$ are the degrees of W.*

(b) $X_{\text{des}} = \sum_{i=1}^n X_{\text{des}}^{(i)}$, *where* $X_{\text{des}}^{(i)} \sim \operatorname{Bin}\left(1, (1 + q_i^{(n)})^{-1}\right)$ *and $q_1, \ldots, q_n$ are the negatives of the roots of $\mathcal{G}_{\text{des}}(W)$.*

Remark 2.2.6. For the symmetric groups, Corollary 2.2.5a) gives

$$X_{\text{inv}} = \sum_{i=1}^{n-1} U(\{0, \ldots, i\})\,, \tag{2.3}$$

which can also be explained by the argument sketched in Remark 2.1.14. First, we look at the last summand $U(\{0, 1, \ldots, n-1\})$. It encodes the number of inversions induced by the largest entry n. According to Remark 2.1.14, this number equals $n - \pi^{-1}(n)$, which is a uniformly random number in $\{0, \ldots, n-1\}$. Likewise, the choice of any of the remaining $n-1$ positions for $\pi^{-1}(n-1)$ determines the number of inversions induced by $\pi^{-1}(n-1)$, and this holds independently of $\pi^{-1}(n)$. Recursively, (2.3) follows.

2.3 Central Limit Theorems

The asymptotic normality of inversions and descents is an important and extensively investigated subject. For the special case of symmetric groups, there is a multifarious history of different proofs of the CLT. For illustration purposes, Figure 2.7 shows the probability mass function of X_{des} on the symmetric group S_{31}, which clearly displays similarity to a Gaussian density function. The following list of proofs of the CLT for X_{des} on symmetric groups can be found in [20, Section 3]:

- On symmetric groups, the representation (2.2) of X_{des} is a sum of m-dependent indicator variables. Thus, the CLT for X_{des} on S_n is an immediate consequence of the CLT for m-dependent random variables (see, e.g., [21] for a version with error bounds).
- Another proof is obtained from the fact that according to [100], $\mathbb{P}(X_{\mathrm{des}} = j) = \mathbb{P}(j \leq Z_1 + \ldots + Z_n \leq j+1)$ for $0 \leq j \leq n-1$ and i.i.d. $Z_1, \ldots, Z_n \sim U(0,1)$.
- For proofs based on the roots of the generating function of X_{des}, see [57, 88].
- For the use of other regularity properties of the generating function of X_{des}, see [5, Ex. 3.5 and 5.3].
- For applications of Stein's method of exchangeable pairs, see [26, 51].

Stein's method is a popular approach in other frameworks as well. See, e.g., [27] for permutations on multisets, [87] for generalized inversions (which will be extensively discussed in Chapter 5), and [4] for non-uniformly distributed permutations.

The first work giving a CLT for X_{inv} and X_{des} on symmetric groups is due to Bender [5, Ex. 5.3 and 5.5]. Inversions are also covered in the work of Fulman [51], which uses that inversions and descents are a special case of permutation statistics based on antisymmetric matrices. It is also possible to prove the CLT for X_{inv} by use of Janson's dependency criterion [64, Theorem 2], as shown in [10].

Moreover, since the real-rootedness of the Eulerian polynomial $\mathcal{G}_{\mathrm{des}}$ is known for the irreducible families S_n, B_n, D_n, the CLT for X_{des} on these families (and their products) is implied by [13, Theorem 2.1]. Most recently, Özdemir [82] gave a new proof of the CLT for X_{des} on symmetric groups by a martingale representation of X_{des} and a Lindeberg-Feller type CLT for martingale differences (see [56, Theorem 3.2]).

For general finite Coxeter groups, the statistics X_{inv} and X_{des} satisfy the CLT in most cases. The validity of the CLT was classified by Kahle & Stump [66]. In essence, the CLT is true if and only if the variances grow fast enough.

Theorem 2.3.1. (see [66], Theorem 6.1 and 6.2)
Let $W_1, W_2, \ldots$ be finite Coxeter groups with $\mathrm{rk}(W_n) = n$. Let $d_{\max}$ be the maximum degree of W_n. Let s_n^2 be the variance of $X_{\mathrm{inv}}^{(n)}$ or $X_{\mathrm{des}}^{(n)}$, respectively. Then:

(a) $X_{\mathrm{inv}}^{(n)}$ *satisfies the CLT if and only if* $d_{\max}/s_n \longrightarrow 0$.

(b) $X_{\mathrm{des}}^{(n)}$ *satisfies the CLT if and only if* $s_n \longrightarrow \infty$.

The proof is based on the representations in Corollary 2.2.5, which allow to represent the sequences $\left(X_{\mathrm{inv}}^{(n)}\right)_{n\in\mathbb{N}}$ and $\left(X_{\mathrm{des}}^{(n)}\right)_{n\in\mathbb{N}}$ as triangular arrays. The necessary and sufficient conditions guarantee that no summand in Corollary 2.2.5 dominates the others.

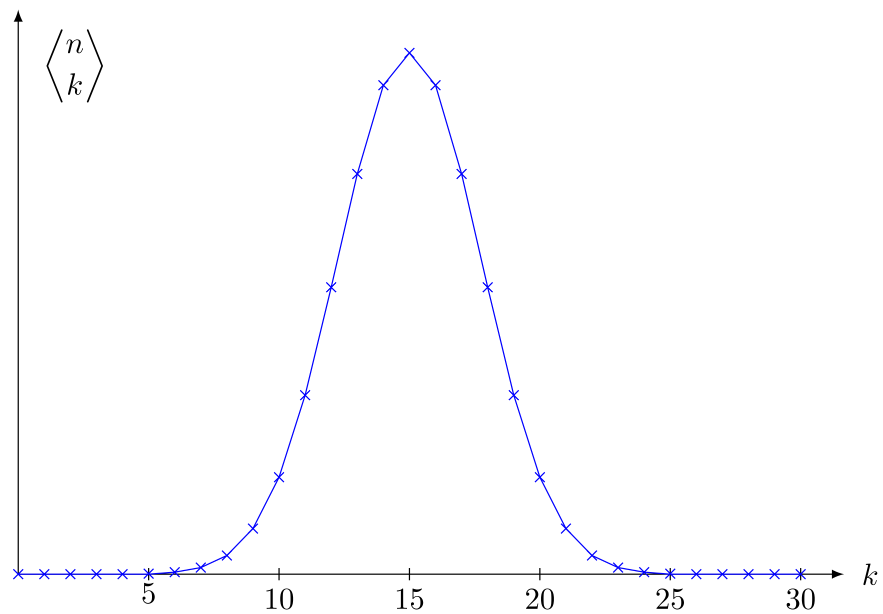


Figure 2.7: True to scale probability mass function of Eulerian numbers on S_n for $n = 31$.

Definition 2.3.2. Let $\left(X_i^{(n)}\right)_{i=1,\dots,k_n}$ be a triangular array of real-valued, row-wise independent random variables with $X^{(n)} := \sum_{i=1}^{k_n} X_i^{(n)}$ and $s_n^2 := \mathrm{Var}(X^{(n)})$. The triangular array satisfies the *Lindeberg condition* if

$$\forall \varepsilon > 0\colon \frac{1}{s_n^2} \sum_{i=1}^{k_n} \mathbb{E}\left(\left|X_i^{(n)}\right|^2 \mathbf{1}\left\{\left|X_i^{(n)}\right| \geq \varepsilon s_n\right\}\right) \longrightarrow 0\,.$$

Another way to formalize the non-dominance of a single summand is the *maximum condition*

$$\max_{i=1,\dots,k_n} \frac{\mathrm{Var}(X_i^{(n)})}{s_n^2} \longrightarrow 0\,.$$

Theorem 2.3.3. (see, e.g., [7], p. 361)
In the setting of Definition 2.3.2, $X^{(n)}$ satisfies the Lindeberg condition if and only if it satisfies the CLT and the maximum condition.

In the setting of Theorem 2.3.1, we have $k_n = n\ \forall n \in \mathbb{N}$. For the proof of Theorem 2.3.1, Kahle & Stump [66] use the following key steps:

- For $X_i^{(n)} \sim U(\{0, \dots, d_{ni} - 1\})$ with any choice of integers $2 \leq d_{n1} \leq \dots \leq d_{nn}$, the CLT is equivalent to the maximum condition (see [66, Proposition 6.12]).
- The maximum condition itself is equivalent to the condition $d_{nn} = o(s_n)$ which appears in Theorem 2.3.1a) (see [66, Lemma 6.13]).
- For any triangular array $\left(X_i^{(n)}\right)_{i=1,\dots,n}$ with globally bounded $X_i^{(n)}$ such that all $X^{(n)} - \mathbb{E}(X^{(n)})$ take values in a fixed lattice $\delta\mathbb{Z} \subseteq \mathbb{R}$ for some $\delta > 0$, the CLT is equivalent to the condition $\mathrm{Var}(X^{(n)}) \longrightarrow \infty$ which appears in Theorem 2.3.1b) (see [66, Propositions 6.14 and 6.15]).

Definition 2.3.4. Another interesting statistic on a Coxeter group W is

$$T(\pi) := \mathrm{des}(\pi) + \mathrm{des}(\pi^{-1}), \quad \pi \in W.$$

Accordingly, let $X_T := X_{\mathrm{des}}(\pi) + X_{\mathrm{ides}}(\pi)$ with $X_{\mathrm{ides}}(\pi) := X_{\mathrm{des}}(\pi^{-1})$ denote the corresponding random variable on W. We call X_T the *two-sided Eulerian statistic*, as it is a sum of two Eulerian variables. △

Obviously, X_{des} and X_{ides} are not independent. The mean and variance of X_T have the same magnitudes as those of X_{des}, see [66, Theorem 5.1 and Corollary 5.2]. A further observation is that the asymptotic normality of X_T is closely related to that of the joint distribution $(X_{\mathrm{des}}, X_{\mathrm{ides}})$ due to the Cramér-Wold device.

Chatterjee & Diaconis [20] proved a CLT for X_T on the symmetric groups S_n by constructing random permutations from order statistics of uniform variables on $[0,1]^2$, and using the so-called *method of interaction graphs*. This method was previously developed by Chatterjee in [19]. Here, the random quantity of interest is modeled as a multivariate function. An interaction graph displays the pairs of indices for which replacing both coordinates has a different impact on the quantity than replacing exactly one of the two coordinates. Brück & Röttger [15, 94] extended the result of [20] to B_n and D_n by suitably adding a random sign. Furthermore, they proved a CLT for X_T on a wide range of composed finite Coxeter groups. Their observations can be summarized as follows:

- If the groups W_n consist of only classical Weyl groups, then the CLT holds if their ranks are balanced in a way that the largest component rank does not grow as fast as the sum of all remaining ranks (see [15, Lemma 25]).
- For mixed groups, the sequence $(W_n)_{n\in\mathbb{N}}$ must satisfy a regularity condition called *well-behavedness* [15, Definition 28], however, no counterexample to this condition was found according to [15, Remark 29]. Then, for all well-behaved sequences $(W_n)_{n\in\mathbb{N}}$, the CLT holds if and only if $\mathrm{Var}(X_T) \longrightarrow \infty$ (see [15, Theorem 32]).

Moreover, Féray [48] gave a different proof of the CLT for X_T on *all* products of Coxeter groups with $\mathrm{Var}(X_T) \longrightarrow \infty$, without requiring the condition of well-behavedness. This proof relies on the CLT for X_T on classical Weyl groups. The components are split into classical Weyl groups of large rank (for which X_T is already close to normality) and other components, for which the CLT is derived by comparison of characteristic functions.

While the asymptotic normality of the *individual* statistics X_{inv} and X_{des} is well studied, comparatively little is known about their *joint distribution*. On the symmetric groups S_n, Fang & Röllin [45] proved a CLT for $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$. Furthermore, they proved an $O(n^{-1/2})$ rate of convergence using a multivariate generalization of [51]. So far, however, it is unclear whether their method can be generalized to other finite Coxeter groups.

Definition 2.3.5. Let $A \in \mathbb{R}^{n\times n}$ be an antisymmetric matrix, that is, $A_{uu} = 0$ and $A_{uv} = -A_{vu}$ for all $u, v = 1, \dots, n$, $u \neq v$. This matrix yields the following permutation statistic:

$$\mathcal{W}(\pi) = \sum_{1\le i<j\le n} A_{\pi(i)\pi(j)}, \tag{2.4}$$

for $\pi \in S_n$ uniformly at random. Given several antisymmetric matrices $A^{(1)}, \dots, A^{(d)}$, let

$$\mathcal{W}_1 := \sum_{1\le i<j\le n} A^{(1)}_{\pi(i)\pi(j)}, \quad \dots, \quad \mathcal{W}_d := \sum_{1\le i<j\le n} A^{(d)}_{\pi(i)\pi(j)}.$$

Example 2.3.6. Consider the following two matrices:

$$A^{(1)} := \begin{pmatrix} 0 & -1 & -1 & \cdots & -1 \\ 1 & 0 & -1 & \ddots & \vdots \\ 1 & \ddots & \ddots & \ddots & -1 \\ \vdots & \ddots & \ddots & 0 & -1 \\ 1 & \cdots & 1 & 1 & 0 \end{pmatrix}, \qquad A^{(2)} := \begin{pmatrix} 0 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & -1 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

Then, we see that

$$\begin{aligned}
\mathcal{W}_1 &= \sum_{1\le i<j\le n} \mathbf{1}\{\pi(i)=\pi(j)+1\} - \mathbf{1}\{\pi(j)=\pi(i)+1\} \\
&= \sum_{1\le i<j\le n} \mathbf{1}\{\pi(i)=\pi(j)+1\} - \sum_{i>j} \mathbf{1}\{\pi(i)=\pi(j)+1\} \\
&= 2\sum_{1\le i<j\le n} \mathbf{1}\{\pi(i)=\pi(j)+1\} - (n-1) \\
&= 2\sum_{i=1}^{n-1} \mathbf{1}\{\pi^{-1}(i+1)>\pi^{-1}(i)\} - (n-1) \\
&= 2\,\mathrm{Des}(\pi^{-1}) - (n-1)\,, \\
\mathcal{W}_2 &= \sum_{1\le i<j\le n} \mathbf{1}\{\pi(i)>\pi(j)\} - \mathbf{1}\{\pi(i)<\pi(j)\} \\
&= 2\,\mathrm{Inv}(\pi) - \binom{n}{2} = 2\,\mathrm{Inv}(\pi^{-1}) - \binom{n}{2}.
\end{aligned}$$

As π^{-1} also represents a uniformly random permutation, we can interpret the standardizations of X_{inv} and X_{des} as special instances of (2.4).

Theorem 2.3.7. (see [45], Corollary 3.8)
Let $\mathcal{N} \sim \mathrm{N}_2(0, \mathrm{I}_2)$ denote the two-dimensional standard normal distribution. For the joint distribution

$$\mathcal{W} = \left(\frac{X_{\mathrm{inv}} - n(n-1)/4}{\sqrt{n(n-1)(2n+5)/72}}, \frac{X_{\mathrm{des}} - (n-1)/2}{\sqrt{(n+1)/12}} \right),$$

there is a universal constant C so that

$$\sup_{\substack{A \subseteq \mathbb{R}^2 \\ A \text{ convex}}} |\mathbb{P}(\mathcal{W} \in A) - \mathbb{P}(\mathcal{N} \in A)| \le C n^{-1/2}.$$

In the following, while not listing all the technical details of the proof, we will comprehend the $O(n^{-1/2})$ rate of convergence.

Definition 2.3.8. (see [45], Definition 2.1)
A multivariate *Stein coupling* is a triple (X, X', G) of d-dimensional random vectors satisfying $\mathbb{E}(G^\top F(X') - G^\top F(X)) = \mathbb{E}(X^\top F(X))$ for all vector fields $F\colon \mathbb{R}^d \to \mathbb{R}^d$, provided both means exist. In this case, it immediately follows that $\mathbb{E}(X) = 0$ and $\mathbb{E}\big(G(X' - X)^\top\big) = \mathrm{Var}(X)$. Write $D := X' - X$. To shorten the notation, we write $E^X(\cdot) = \mathbb{E}(\cdot \mid X)$ for the conditional means. The main error bound from Stein's method is as follows:

Theorem 2.3.9. (see [45], Theorem 2.1)
Within the framework of Definition 2.3.8, assume that $\mathrm{Var}(X) = \mathrm{I}_d$, *and that* G *and* D *are bounded in the sense of*

$$\|G\| \le \alpha\,, \qquad \|D\| \le \beta\,,$$

for some constants $\alpha, \beta > 0$. *Let* $\mathcal{N} \sim \mathrm{N}_d(0, \mathrm{I}_d)$. *Then, there exists a universal constant* C *so that*

$$\begin{aligned} d(X, \mathcal{N}) &:= \sup_{\substack{A \subseteq \mathbb{R}^d \\ A \text{ convex}}} |\mathbb{P}(X \in A) - \mathbb{P}(\mathcal{N} \in A)| \\ &\le C\left(d^{7/4}\alpha \mathbb{E}\|D\|^2 + d^{1/4}\beta + d^{7/8}\sqrt{\alpha B_1} + d^{3/8}B_2 + d^{1/8}\sqrt{B_3}\right), \end{aligned}$$

where

$$\begin{aligned} B_1 &:= \sqrt{\mathrm{Var}\big(\mathbb{E}^X(\|D\|^2)\big)}\,, \\ B_2 &:= \sqrt{\sum_{i,j=1}^{d} \mathrm{Var}\big(\mathbb{E}^X(G_i D_j)\big)}\,, \\ B_3 &:= \sqrt{\sum_{i,j,k=1}^{d} \mathrm{Var}\big(\mathbb{E}^X(G_i D_j D_k)\big)}\,. \end{aligned}$$

According to Fulman [51], for any collection $\mathcal{W} = (\mathcal{W}_1, \ldots, \mathcal{W}_d)$ of doubly-indexed permutation statistics, a Stein coupling is constructed by taking $I \sim U(\{1, \ldots, n\})$ and $\mathcal{W}'(\pi) := \mathcal{W}(\pi')$, where $\pi' = \pi \circ (I \mapsto I+1 \mapsto \ldots \mapsto n \mapsto I)$. Fulman showed that $\mathbb{E}^\pi(\mathcal{W}' - \mathcal{W}) = -(2/n)\mathcal{W}$ for all $\pi \in S_n$. Taking $G := nD/4$ gives a Stein coupling. The entries of D are explicitly given by

$$D_r = -2\sum_{j>I} A^{(r)}_{\pi(I)\pi(j)}\,,$$

and with

$$\beta_1 := \max_{\substack{r=1,\ldots,d \\ u=1,\ldots,n}} \sum_{v=1}^{n} |A^{(r)}_{uv}|\,,$$

it is seen that there exists a constant C_d only depending on d giving the bounds

$$\|G\| \le C_d n \beta_1\,, \qquad \|D\| \le C_d \beta_1\,. \tag{2.5}$$

Extensive calculations in [45] show that this constant can also be used to give the bounds

$$\begin{aligned} \mathrm{Var}\big(\mathbb{E}^\pi(D_r D_s)\big) &\le \frac{C_d \beta^4}{n}\,, \\ \mathrm{Var}\big(\mathbb{E}^\pi(D_r D_s D_t)\big) &\le C_d \beta^4 \beta_2\,, \end{aligned}$$

where $1 \le r, s, t \le d$ and

$$\beta_2 := \max_{\substack{r=1,\ldots,d \\ u=1,\ldots,n}} \sum_{v=1}^{n} (A^{(r)}_{uv})^2.$$

We now apply Theorem 2.3.9 to verify the $O(n^{-1/2})$ bound for $\mathcal{W}$ as in Theorem 2.3.7. Since $d = 2$ remains constant, we can ignore the respective powers of d and write

$$d(\mathcal{W}, \mathcal{N}) \leq C\left(\alpha \mathbb{E}\|D\|^2 + \beta + \sqrt{\alpha B_1} + B_2 + \sqrt{B_3}\right), \tag{2.6}$$

where, according to (2.5), we have $\alpha = \Theta(n\beta_1)$, $\beta = \Theta(\beta_1)$. Due to the standardization, the matrices representing $\mathcal{W}$ are given by

$$A^{(1)}_{uv} = \sqrt{\frac{18}{n(n-1)(2n+5)}}\begin{cases}-1, & v > u\\ 1, & v < u\\ 0, & v = u\end{cases}, \qquad A^{(2)}_{uv} = \sqrt{\frac{3}{n+1}}\begin{cases}-1, & v = u+1\\ 1, & v = u-1\\ 0, & \text{otherwise}\end{cases}.$$

Therefore,

$$\beta_1 = \max_u \left(\sum_{v=1}^n |A^{(1)}_{uv}|\right) \vee \left(\sum_{v=1}^n |A^{(2)}_{uv}|\right).$$

In $A^{(1)}$, there are $n-1$ non-zero entries in each row, all giving a contribution of $\Theta(n^{-3/2})$, so the row-wise sums of absolute values are of order $\Theta(n^{-1/2})$. In $A^{(2)}$, there are only at most two non-zero entries per row, which are of the fitting order $\Theta(n^{-1/2})$. In conclusion, $\beta_1 = \Theta(n^{-1/2})$. Furthermore, $\alpha = \Theta(n^{1/2})$ and $\|D\| = O(n^{-1/2}) \implies \alpha\mathbb{E}(\|D\|^2) = O(n^{-1/2})$. In addition, $B_1 = \Theta(n^{-3/2}) \implies \alpha B_1 = \Theta(n^{-1}) \implies \sqrt{\alpha B_1} = \Theta(n^{-1/2})$. Next, as $G = nD/4$, we have

$$B_2 = \sqrt{\sum_{i,j=1}^2 \mathrm{Var}(\mathbb{E}^{\mathcal{W}}(G_i D_j))} = \sqrt{\sum_{i,j=1}^2 n^2 \mathrm{Var}(\mathbb{E}^{\mathcal{W}}(D_i D_j))} \leq \sqrt{\sum_{i,j=1}^2 C_d n \beta^4} = O(n^{-1/2}).$$

The same is obtained for B_3, as β_2 is bounded in the way of

$$\beta_2 = \max_u \underbrace{\left(\sum_{v=1}^n \left|A^{(1)}_{uv}\right|^2\right)}_{=\Theta(n^{-2})} \vee \underbrace{\left(\sum_{v=1}^n \left|A^{(2)}_{uv}\right|^2\right)}_{=\Theta(n^{-1})} = \Theta(n^{-1}).$$

In conclusion, every summand on the right hand side of (2.6) is $O(n^{-1/2})$, which explains the rate of convergence.

3 Extremes of inversions and descents on finite Coxeter groups

3.1 Introduction and framework

In Section 2.3, we discussed the CLT for the number of inversions and descents on finite Coxeter groups. Subsequently, we aim to understand the extreme value asymptotics of these statistics. Up to now, to the best of our knowledge, only two permutation statistics have been studied for the asymptotic behavior of their extremes, see [53, 79]. These two statistics are introduced as the maxima of a collection of substatistics $X_{n1}(\pi), \ldots, X_{nn}(\pi)$ for $\pi = (\pi(1), \ldots, \pi(n)) \in S_n$:

- The *largest gap statistic* is the maximum of $X_{nj}(\pi) := |\pi(j) - \pi(j+1)|$, $j = 1, \ldots, n$ (writing $\pi(n+1) := \pi(1)$).
- For fixed $k \in \mathbb{N}$ and $n > k$, the *largest consecutive k-sum statistic* is the maximum of $X_{nj}(\pi) := \sum_{i=0}^{k-1} \pi(i+j)$ (again using the ring notation $\pi(n+j) := \pi(j)$).

In both cases, the variables $(X_{n1}, \ldots, X_{nn})$ form a triangular array where in each row, the entries $X_{n1}, \ldots, X_{nn}$ are based on the same permutation $\pi \in S_n$.

Theorem 3.1.1. (see Mladenović [79], Theorem 1)
Let X_n be the largest gap statistic on S_n. Then, for all $x \in \mathbb{R}$:

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq x\sqrt{n} + n) = \Psi_2(x)\,,$$

where Ψ_2 denotes the CDF of the Weibull distribution with shape parameter 2.

The proof is divided in two steps. First, by means of Lemma 1.1.10 it is shown that the corresponding sequence of *independent* variables $X^*_{n1} \overset{\mathcal{D}}{=} X_{n1}, \ldots, X^*_{nn} \overset{\mathcal{D}}{=} X_{nn}$ satisfies the claim. Then, for the dependent but stationary segment $X_{n1}, \ldots, X_{nn}$, the mixing conditions introduced in Section 1.2 are shown by some combinatorial inclusion-exclusion arguments, so that the claim carries over from the independent variables. The following result is proved in the same way.

Theorem 3.1.2. (see Glavas et al. [53], Theorem 1.1)
Let X_n be the largest consecutive k-sum statistic on S_n for fixed $k \in \mathbb{N}$. Then, for all $x \in \mathbb{R}$:

$$\lim_{n\to\infty} \mathbb{P}\left(X_n \leq x(k!n^{k-1})^{1/k} + kn\right) = \Psi_k(x)\,,$$

where Ψ_k denotes the CDF of the Weibull distribution with shape parameter k.

For the numbers of inversions and descents, there is no feasible approach to represent them as a maximum of other random variables. Instead, we consider a triangular array consisting of *independent samples* drawn from a finite Coxeter group in each row.

It is not mandatory to draw exactly n samples from a finite Coxeter group of rank n. There are two equivalent frameworks (cf. Remark 1.2.10):

(a) We consider a sequence of finite Coxeter groups $(W_n)_{n\in\mathbb{N}}$ with $\mathrm{rk}(W_n) = n$ and a triangular array $(X_{nj})_{j=1,\dots,k_n}$ of arbitrary length.

(b) We use a uniform triangular array $(X_{nj})_{j=1,\dots,n}$, where $X_{n1}, \dots, X_{nn}$ are independent samples drawn from finite Coxeter groups of arbitrary ranks.

Figures 3.1 and 3.2 illustrate these two frameworks for comparison, showing exemplary triangular arrays with samples drawn from symmetric groups for simplicity.

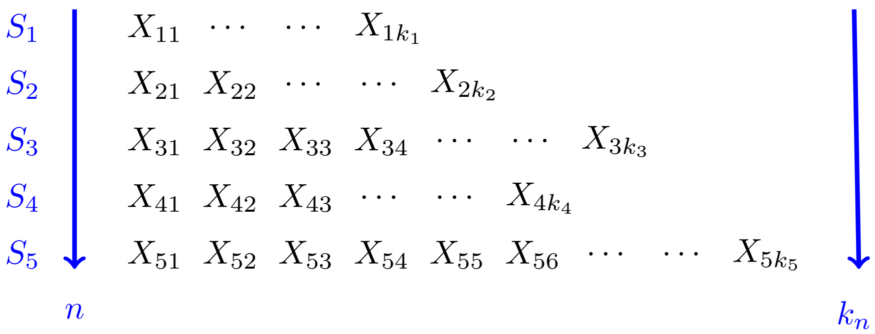


Figure 3.1: Beginning of a triangular array of permutation statistics on all symmetric groups. The row lengths k_n have no specific regularity.

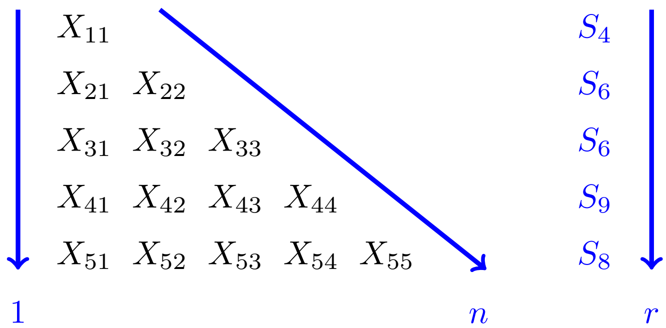


Figure 3.2: Beginning of a uniform triangular array of permutation statistics with an exemplary choice of the first five symmetric groups. The choices $r(n)$ of the ranks have no specific regularity.

In what follows, we use the first framework since it allows simpler notation. From there, it is trivial to state the analogous results within the second framework. So, we have a sequence of finite Coxeter groups $(W_n)_{n\in\mathbb{N}}$ with $\mathrm{rk}(W_n) = n$ and a triangular array $(X_{nj})_{j=1,\dots,k_n}$, where for each $n \in \mathbb{N}$, the $X_{n1}, \dots, X_{nk_n}$ are independent samples of X_{inv} or X_{des} on W_n. We suppose that the triangular array contains only samples of either X_{inv} or X_{des}, but not of both. We expect that k_n must satisfy an asymptotic upper bound to avoid degeneracy of extreme values, for the reasons given in Remark 1.2.10.

Remark 3.1.3. To substantiate this explicitly, we state an asymptotic rate of $(k_n)_{n\in\mathbb{N}}$ for which the above triangular array $(X_{nj})_{j=1,\dots,k_n}$ has degenerate extreme value limit behavior. Each finite Coxeter group W has a unique *longest element* w_0 for which both $X_{\mathrm{inv}}(w_0)$ and $X_{\mathrm{des}}(w_0)$ are maximal (see, e.g., [8, Proposition 2.3.1]. Let M_0 denote this maximum value. For instance, on the symmetric groups S_n, we have $w_0 = (n, n-1, n-2, \ldots, 1)$, giving $M_0 = n(n-1)/2$ for inversions and $M_0 = n-1$ for descents. Now, let $w_1^{(n)}, \ldots, w_{k_n}^{(n)}$ denote i.i.d. random elements of W_n. Then,

$$\mathbb{P}(M_n = M_0) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{k_n} \{w_i^{(n)} \neq w_0\}\right) = 1 - \left(1 - \frac{1}{|W_n|}\right)^{k_n}.$$

If we choose, e.g., $k_n \geq |W_n|^2$, then $(1 - |W_n|^{-1})^{k_n} \longrightarrow 0$ and $\mathbb{P}(M_n = M_0) \longrightarrow 1$, which means that $(M_n)_{n\in\mathbb{N}}$ cannot be rescaled in order to achieve a non-degenerate limit distribution. △

In the following, we prove that given a suitable upper bound on k_n, the row-wise maximum $M_n := \max\{X_{n1}, \ldots, X_{nk_n}\}$ is attracted to the Gumbel distribution as $n \to \infty$. Section 3.2 introduces the key tools, namely, theorems that provide tail equivalence through large deviation bounds. Section 3.3 presents the new EVLT for X_{inv} and X_{des} on the important families S_n, B_n, D_n as the first main result of this thesis. Section 3.4 postulates the EVLT on arbitrary finite Coxeter groups, and also categorizes it for different subclasses of finite Coxeter groups. Section 3.5 discusses other permutation statistics and presents a universal EVLT which does not rely on any specific properties of permutation statistics except asymptotic normality. Section 3.6 suggests further examples of random inversions and descents in certain structures, for which the transferability of the EVLT is an open question.

3.2 Tail equivalence for non-identically distributed sums

In Theorem 1.2.13 by Nadarajah & Mitov [81], it was demonstrated how to derive an EVLT for binomial distributions from tail equivalence to the standard normal distribution. We aim to achieve tail equivalence in a more general framework to obtain EVLTs for the numbers of inversions and descents. Note that by Corollary 2.2.5, both of these numbers can be decomposed into independent summands. In contrast to the decomposition of binomial distributions into Bernoulli distributions, these summands are not identically distributed, which makes achieving tail equivalence more difficult.

The subject of tail equivalence is closely related to the field of *large deviations theory*. Based on limit theorems such as the strong law of large numbers or the CLT, this theory deals with bounds and quantification for the probabilities of large deviations from the limit. See [34, Chapters 1 and 2] for an introduction. We assume the following framework for all theorems presented in this section.

Framework: Let $X_1, X_2, \ldots$ be an at most countable sequence of centered and independent random variables. For $n \in \mathbb{N}$, let $\mathcal{S}_n := X_1 + \ldots + X_n$. Let $\sigma_k^2 = \mathbb{E}(X_k^2)$ for all $k = 1, \ldots, n$, and let $s_n^2 := \sigma_1^2 + \ldots + \sigma_n^2$. Moreover, let F_n denote the CDF of $\mathcal{S}_n/s_n$.

In analogy to Theorem 1.2.12, we aim to demonstrate tail equivalence between F_n and the CDF Φ of the standard normal distribution, or in symbols,

$$1 - F_n(x) \sim 1 - \Phi(x) \iff \frac{1 - F_n(x)}{1 - \Phi(x)} = 1 + o(1).$$

We want to show this not only for fixed x but also for sequences $x = x(n)$ depending on $n \in \mathbb{N}$. Typically, this works only if x does not grow too fast. Upon additionally assuming that $X_1, X_2, \ldots$ are identically distributed, a seminal result on large deviations of $(1 - F_n)/(1 - \Phi)$ is due to Cramér [29].

Theorem 3.2.1. (cf. Cramér [29])
Under the given framework, assume that $X_1, X_2, \ldots$ are i.i.d. and that the moment generating function of X_1 exists in a neighborhood of the origin. If $x = o(\sqrt{n})$, then

$$\frac{1 - F_n(x)}{1 - \Phi(x)} = \exp\left(\frac{x^3}{\sqrt{n}} \mathcal{L}\left(\frac{x}{\sqrt{n}}\right)\right) (1 + o(1)) ,$$

where $\mathcal{L}(x) = \sum_{k=0}^{\infty} a_k x^k$ is a power series with coefficients depending on the cumulants of X_1.

A similar theorem that omits the assumption of identical distribution was developed by Feller [46]. This theorem imposes boundedness assumptions on the random variables, therefore it is not a generalization of Theorem 3.2.1.

Theorem 3.2.2. (see Feller [46])
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence of constants such that $\lambda_n \longrightarrow 0$ and

$$\forall k = 1, \ldots, n\colon |X_k| < \lambda_n s_n . \tag{3.1}$$

Let $x > 0$ be fixed and assume that

$$\forall n \in \mathbb{N}\colon 0 < \lambda_n x < (3 - \sqrt{5})/4 \approx 0.19 .$$

Then, there is a constant ϑ with $|\vartheta| < 9$ and a power series $Q_n(x) = \sum_{\nu=1}^{\infty} q_{n,\nu} x^\nu$ with coefficients $q_{n,\nu}$ depending on the first $\nu + 2$ moments of X_n so that

$$1 - F_n(x) = \exp\left(-\frac{1}{2} x^2 Q_n(x)\right) \left(1 - \Phi(x) + \vartheta \lambda_n e^{-x^2/2}\right) .$$

If, in particular, $0 < \lambda_n x < 1/12$, then $|q_{n,\nu}| < \frac{1}{7}(12\lambda_n)^\nu$.

Remark 3.2.3. Theorem 3.2.2 concerns finite sequences of random variables and does not include any asymptotic statement. Nevertheless, it can be applied for each $n \in \mathbb{N}$ on a uniform triangular array $(X_{n1}, \ldots, X_{nn})$ to draw asymptotic conclusions. As stated by Feller [46], if it is possible to choose a sequence $(\lambda_n)_{n\in\mathbb{N}}$ with $\lambda_n = O(n^{-1/2})$, and if $x = o(n^{1/6})$, then it follows that

$$\begin{aligned} Q_n(x) = q_{n,1} x + \sum_{\nu=2}^{\infty} q_{n,\nu} x^\nu \le \frac{12}{7} \lambda_n x + O(n^{-2/3}) = O(n^{-1/3}) \\ \implies \exp\left(-\frac{x^2}{2} Q_n(x)\right) \longrightarrow 1 . \end{aligned}$$

Furthermore, $e^{-x^2/2}$ is bounded by 1 and ϑ is a constant, so $\vartheta \lambda_n e^{-x^2/2} \longrightarrow 0$. Thus, whenever the aforementioned conditions are satisfied, we have the desired tail equivalence $1 - F_n(x) \sim 1 - \Phi(x)$. △

In comparison, Cramér's Theorem 3.2.1 allows for the broader regime $x = o(n^{1/2})$, and it has also been generalized to random variables that are independent and not necessarily identically distributed. We now introduce a large deviations theorem of Petrov & Robinson [86], which is, to the best of our knowledge, the weakest known generalization of Theorem 3.2.1.

Under the given framework, let L_j be the cumulant generating function of X_j, that is, $L_j(z) = \log\left(\mathbb{E}\left(e^{zX_j}\right)\right)$. We assume that for some $H > 0$, all functions L_j are analytic within the circle $\{z \in \mathbb{C} \colon |z| < H\}$. Moreover, we assume the existence of constants $(c_j)_{j\in\mathbb{N}}$ such that $\forall |z| < H, j \in \mathbb{N} \colon |L_j(z)| < c_j$ and

$$\limsup_{n\to\infty} \sum_{j=1}^{n} \frac{c_j}{n} < \infty \,. \tag{3.2}$$

At last, we require that the variances s_n^2 grow at least linearly, that is,

$$\liminf_{n\to\infty} \frac{s_n^2}{n} > 0 \,. \tag{3.3}$$

Theorem 3.2.4. (see Petrov & Robinson [86], Theorem 2.1)
Given the conditions (3.2) and (3.3), it holds that for $x = o(\sqrt{n})$,

$$\frac{1 - F_n(x)}{1 - \Phi(x)} = \exp\left(\frac{x^3}{\sqrt{n}} \mathcal{L}_n\left(\frac{x}{\sqrt{n}}\right)\right)(1 + o(1)) \,,$$

where $\mathcal{L}_n(x) = \sum_{k=0}^{\infty} a_{kn} x^k$ is a power series with coefficients a_{kn} expressed in terms of the cumulants of $X_1, \ldots, X_n$ of order up to and including $n + 3$.

This theorem is an advancement of [85, Theorem 1], which imposed the stricter condition of $\limsup_{n\to\infty} \sum_{j=1}^{n} \frac{c_j^{3/2}}{n} < \infty$. However, it is not a generalization of Feller's Theorem 3.2.2.

Remark 3.2.5. For the extended regime $n^{1/6} \ll x \ll n^{1/2}$, it is not trivial to achieve tail equivalence via Theorems 3.2.1 and 3.2.4. To do so, we additionally have to demonstrate

$$\begin{aligned} &\exp\left(\frac{x^3}{\sqrt{n}} \mathcal{L}_n\left(\frac{x}{\sqrt{n}}\right)\right) = 1 + o(1) \\ \iff &\frac{x^3}{\sqrt{n}} \mathcal{L}_n\left(\frac{x}{\sqrt{n}}\right) = o(1) \,. \end{aligned}$$

The term $x^3/\sqrt{n}$ can become as large as $o(n)$. It is controlled only if $x = o(n^{1/6})$, which is the same regime as in Theorem 3.2.2. For broader regimes, we have to control the power series $\mathcal{L}_n$. For $j, k \in \mathbb{N}$, let γ_{kj} be the k-th cumulant of X_j and let

$$\Gamma_{kn} = \sum_{j=1}^{n} \frac{\gamma_{kj}}{n} \,.$$

According to [86, p. 2985], the first coefficient of $\mathcal{L}_n$ is given by

$$a_{0,n} = \frac{\Gamma_{3,n}}{6\Gamma_{2,n}^{3/2}} \,.$$

If $a_{0,n}$ is non-zero, then it is impossible to control $\mathcal{L}_n(x/\sqrt{n})$ for any $n^{1/6} \ll x \ll n^{1/2}$. To obtain tail equivalence from Theorem 3.2.4 within the extended regime $n^{1/6} \ll x \ll n^{1/2}$, it is necessary that $a_{0,n} = O(n^{-1})$. In the intermediate case of $a_{0,n} = o(1)$ and $a_{0,n} = \omega(n^{-1})$, it may still be possible to extend the regime of x at least partially. In that case, further coefficients of $\mathcal{L}_n$ may have to be taken into account.

3.3 Sequences of classical Weyl groups

We now show how to obtain the EVLT for the numbers of inversions and descents on sequences of Coxeter groups using the theorems introduced in the previous section. It is important to distinguish whether dihedral groups are involved or not.

We first consider a sequence of classical Weyl groups $(W_n)_{n\in\mathbb{N}}$ with $\mathrm{rk}(W_n) = n\ \forall n \in \mathbb{N}$, and a triangular array $(X_{nj})_{j=1,\ldots,k_n}$ as described in Section 3.1. Let $X_{\mathrm{inv}}^{(n)}$ and $X_{\mathrm{des}}^{(n)}$ be the number of inversions and descents on W_n, respectively. Let $d_1^{(n)}, \ldots, d_n^{(n)}$ be the degrees of W_n and let $q_1^{(n)}, \ldots, q_i^{(n)}$ be the negatives of the roots of $\mathcal{G}_{\mathrm{des}}(W_n)$. According to Corollary 2.2.5, we write

$$X_{\mathrm{inv}}^{(n)} = \sum_{i=1}^{n} X_{\mathrm{inv}}^{(n,i)}\,, \qquad\qquad X_{\mathrm{des}}^{(n)} = \sum_{i=1}^{n} X_{\mathrm{des}}^{(n,i)}\,,$$

where $X_{\mathrm{inv}}^{(n,i)} \sim U\left\{0, 1, \ldots, d_i^{(n)} - 1\right\}$ and $X_{\mathrm{des}}^{(n,i)} \sim \mathrm{Bin}\left(1, (1 + q_i^{(n)})^{-1}\right)$.

Remark 3.3.1. Since Theorem 3.2.4 permits a broader regime of x than Theorem 3.2.2, it is preferable to apply Theorem 3.2.4 to both X_{inv} and X_{des}. Obviously, both statistics satisfy (3.3). However, it turns out that the condition (3.2) of Theorem 3.2.4 is not satisfied for X_{inv}. For $X_{\mathrm{inv}}^{(n,i)} \sim U\left\{0, 1, \ldots, d_i^{(n)} - 1\right\}$, the cumulant generating function is

$$L_i(z) = \log\left(\frac{1}{d_i^{(n)}} \sum_{k=0}^{d_i^{(n)}-1} e^{zk}\right) = \log\left(\frac{1 - e^{d_i^{(n)} z}}{d_i^{(n)}(1 - e^z)}\right).$$

For some $H > 0$, we have to find c_i such that $L_i(z) < c_i\ \forall |z| < H$. In particular,

$$c_i \geq L_i(H) = \log\left(\frac{1 - (e^H)^{d_i^{(n)}}}{d_i^{(n)}(1 - e^H)}\right).$$

Due to $e^H > 1$, we have that $L_i(H)$ grows linearly in i, as its argument grows exponentially in i. Therefore, $\sum_{j=1}^{n} c_j/n$ grows linearly as well and is not bounded, so the condition (3.2) is violated.

For X_{des}, the condition (3.2) is not violated. However, we have to examine the power series $\mathcal{L}_n$ in order to determine the appropriate regime of x. The second, third, and fourth cumulants of $X_{\mathrm{des}}^{(n,i)} \sim \mathrm{Bin}\left(1, (1 + q_i^{(n)})^{-1}\right) =: \mathrm{Bin}(1, p_i)$ are

$$\begin{aligned}\gamma_{2,i} &= p_i(1 - p_i)\,,\\ \gamma_{3,i} &= p_i(1 - p_i)(1 - 2p_i)\,,\end{aligned}$$

$$\gamma_{4,i} = p_i(1-p_i)(1-6\gamma_{2,i})\,.$$

Recall that $a_{0,n} = \Gamma_{3,n}/\Gamma_{2,n}^{3/2}$. The third cumulant $\gamma_{3,i}$ equals the third central moment of $X_{\text{des}}^{(n,i)}$. Therefore, the sum $\sum_{i=1}^{n} \gamma_{3,i}$ equals the third central moment of $X_{\text{des}}^{(n)}$, which is zero as the distribution of $X_{\text{des}}^{(n)}$ is symmetric for all finite Coxeter groups. In conclusion, we have $a_{0,n} = 0$. However, we need to take the second coefficient of $\mathcal{L}_n$ into account, which is, according to [86, p. 2985]:

$$a_{1,n} = \frac{\Gamma_{4,n}\Gamma_{2,n} - 3\Gamma_{3,n}^2}{24\Gamma_{2,n}^3} = \frac{\Gamma_{4,n}}{24\Gamma_{2,n}^2}.$$

Due to $\gamma_{4,i} = \gamma_{2,i}(1-6\gamma_{2,i})$ and $1-6\gamma_{2,i} \in [-1/2, 1)$ for all $p_i \in (0,1)$, we have $|\gamma_{4,i}| < |\gamma_{2,i}|\ \forall i = 1,\dots,n \implies |\Gamma_{4,n}| < |\Gamma_{2,n}|$, giving $a_{1,n} \leq \Gamma_{2,n}^{-1}/24$. However, due to $\Gamma_{2,n} = n^{-1}\mathrm{Var}(X_{\text{des}}^{(n)}) = \Theta(1)$, this only implies $a_{1,n} = O(1)$. In light of Remark 3.2.5, we can extend the regime of x to $x = o(n^{1/4})$ to ensure tail equivalence for X_{des}. $\triangle$

Since Theorem 3.2.4 cannot be applied to inversions, we need Theorem 3.2.2 to achieve tail equivalence. Indeed, this is successful because the components $X_{\text{inv}}^{(n,i)}$ are bounded and the variance of X_{inv} is of appropriate magnitude. This argument also works for descents, but for these, we can use the broader regime $x = o(n^{1/4})$ according to Remark 3.3.1. We summarize these observations for the numbers of inversions and descents on classical Weyl groups as follows, giving the first main result of this thesis.

Theorem 3.3.2. *Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of classical Weyl groups with $\mathrm{rk}(W_n) = n$ $\forall n \in \mathbb{N}$. Let $(X_{nj})_{j=1,\dots,k_n}$ be a row-wise i.i.d. triangular array with either $X_{n1} \overset{\mathcal{D}}{=} X_{\text{inv}}$ $\forall n \in \mathbb{N}$ or $X_{n1} \overset{\mathcal{D}}{=} X_{\text{des}}$ $\forall n \in \mathbb{N}$, where:*

(a) *If $X_{n1} \overset{\mathcal{D}}{=} X_{\text{inv}}$ $\forall n \in \mathbb{N}$, then we assume $k_n = \exp(o(n^{1/3}))$.*
(b) *If $X_{n1} \overset{\mathcal{D}}{=} X_{\text{des}}$ $\forall n \in \mathbb{N}$, then we assume $k_n = \exp(o(n^{1/2}))$.*

Let $M_n := \max\{X_{n1},\dots,X_{nk_n}\}$. Let $\mu_n := \mathbb{E}(X_{n1})$, $s_n^2 := \mathrm{Var}(X_{n1})$, and

$$\alpha_{k_n} = \frac{1}{\sqrt{2\log k_n}}\,, \qquad \beta_{k_n} = \frac{1}{\alpha_n} - \frac{1}{2}\alpha_n(\log\log k_n + \log(4\pi))\,.$$

Let $a_n := \alpha_{k_n}s_n$ and $b_n := \beta_{k_n}s_n + \mu_n$. Then, we have $\forall x \in \mathbb{R}$:

$$\mathbb{P}(M_n \leq a_n x + b_n) \longrightarrow \exp(-\exp(-x))\,.$$

Proof. Let F_n be the CDF of X_{n1}. By Corollary 2.2.5, we know that X_{n1} is a sum of n independent summands. In the case of $(X_{nj})_{j=1,\dots,k_n}$ being numbers of inversions, applying Theorem 3.2.2 separately for each $n \in \mathbb{N}$ gives

$$1 - F_n(xs_n) = \exp\left(-\frac{1}{2}x^2 Q_n(x)\right)\left(1 - \Phi(x) + \vartheta\lambda_n e^{-x^2/2}\right) \quad \forall n \in \mathbb{N}.$$

The condition $\lambda_n = O(n^{-1/2})$ can be equivalently expressed as $|X_k| = O(n^{-1/2}s_n)$. According to Remark 2.1.18, the degrees of classical Weyl groups are bounded by $2n$, and the values of the centered variables $X_{\text{inv}}^{(n,i)} - \mathbb{E}\left(X_{\text{inv}}^{(n,i)}\right)$ are bounded by n. Furthermore,

$s_n = O(n^{3/2})$ holds due to Theorem 2.2.2. Therefore, the choice of $\lambda_n = O(n^{-1/2})$ is possible. Upon undoing the centering assumed in Theorem 3.2.2, we obtain according to Remark 3.2.3:

$$1 - F_n(\mu_n + s_n y) \sim 1 - \Phi(y)\,, \qquad \text{for } y = o(n^{1/6})\,.$$

Plugging in $y = \alpha_{k_n} x + \beta_{k_n}$, and treating x as a constant, the condition $\alpha_{k_n} x + \beta_{k_n} \ll n^{1/6}$ in Theorem 3.2.2 is satisfied due to $n \gg \log(k_n)^3$ by assumption (a).

In the case of $(X_{nj})_{j=1,\dots,k_n}$ being numbers of descents, Theorem 3.2.4 and Remark 3.3.1 give $1 - F_n(\mu_n + s_n y) \sim 1 - \Phi(y)$ for $y = o(n^{1/4})$, which is satisfied for $y = \alpha_{k_n} x + \beta_{k_n}$ by assumption (b). Hence,

$$k_n(1 - F_n(a_n x + b_n)) = k_n\Big(1 - F_n(\mu_n + s_n(\alpha_{k_n} x + \beta_{k_n}))\Big) \longrightarrow e^{-x},$$

proving the Gumbel attraction of the row-wise maxima M_n in both cases. □

Remark 3.3.3. The application of Theorem 3.2.2 fails when we try to extend the regime of k_n. According to [46], if $\lambda_n = O(n^{-1/2})$ and if x is chosen in a way that $n^{1/6} \ll x \ll n^{1/4}$ in Theorem 3.2.2, then we have

$$1 - F_n(x s_n) \sim \exp\left(-\frac{1}{2} q_{n,1} x^3 (1 - \Phi(x))\right), \tag{3.4}$$

as $Q_n(x) = q_{n,1}x + \sum_{\nu=2}^{\infty} q_{n,\nu} x^{\nu}$ with $q_{n,1} = o(n^{-1/2})$. However, $n^{1/2} \ll x^3 \ll n^{3/4}$, giving

$$\exp\left(-\frac{1}{2} x^2 Q_n(x)\right) = \exp\left(-\frac{1}{2} q_{n,1} x^3 + o(1)\right),$$

from which (3.4) follows. The first coefficient $q_{n,1}$ is explicitly stated by Feller [46, Eq. (2.18)] as

$$q_{n,1} = \frac{1}{3 s_n^3} \sum_{i=1}^{n} \mathbb{E}\left(X_{n,i}^3\right).$$

Considering the number of inversions on classical Weyl groups, we have $s_n^3 = \Theta(n^{9/2})$ and $X_{n,i} \sim U(\{0, 1, \dots, d_i - 1\})$. The third moment of $X_{n,i}$ is

$$\mathbb{E}\left(X_{n,i}^3\right) = \sum_{j=0}^{d_i} \frac{1}{d_i + 1} j^3 = \frac{1}{d_i + 1} \frac{d_i^2 (d_i + 1)^2}{4} = \frac{d_i^2 (d_i + 1)}{4} = \Theta(d_i^3)\,.$$

The degrees of the classical Weyl groups are stated in Remark 2.1.18. We conclude that

$$\sum_{k=1}^{n} \mathbb{E}\left(X_{n,i}^3\right) = \Theta(n^4) \Longrightarrow q_{n,1} = \Theta(n^{-1/2})\,.$$

To eliminate $-(1/2) q_{n,1} x^3$ in (3.4), we need $x^3 = o(n^{1/2}) \Longrightarrow x = o(n^{1/6})$, which contradicts the assumption of $x \gg n^{1/6}$. △

As described in Section 3.1, it is equivalent to consider a uniform row-wise i.i.d. triangular array $(X_{nj})_{j=1,\dots,n}$, where X_{n1} is the number of inversions or descents on some finite Coxeter group, whose rank is chosen in dependence of n and diverges as $n \to \infty$. Hence, the ranks are written as $r = r_n := \mathrm{rk}(W_n)$. Then, by analogy with Theorem 1.2.13, it follows that $r \gg \log(n)^3$ is necessary to satisfy the assumptions of Theorem 3.2.2 with $x_n = \alpha_n x + \beta_n$. As the sequence $(r_n)_{n\in\mathbb{N}}$ is divergent, the arguments in the proof of Theorem 3.3.2 remain valid. Thus, Theorem 3.3.2 can be reformulated as follows:

Corollary 3.3.4. *Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of classical Weyl groups with increasing ranks $r_n = \mathrm{rk}(W_n)$. Let $(X_{nj})_{j=1,\dots,n}$ be a row-wise i.i.d. triangular array with either $X_{n1} \stackrel{\mathcal{D}}{=} X_{\mathrm{inv}}\ \forall n \in \mathbb{N}$ or $X_{n1} \stackrel{\mathcal{D}}{=} X_{\mathrm{des}}\ \forall n \in \mathbb{N}$, and let $M_n := \max\{X_{n1}, \dots, X_{nn}\}$. Assume $r_n \gg \log(n)^3$ in the case of $X_{n1} \stackrel{\mathcal{D}}{=} X_{\mathrm{inv}}\ \forall n \in \mathbb{N}$, and assume $r_n \gg \log(n)^{3/2}$ in the case of $X_{n1} \stackrel{\mathcal{D}}{=} X_{\mathrm{des}}\ \forall n \in \mathbb{N}$. Let $\mu_n := \mathbb{E}(X_{n1})$, $s_n^2 := \mathrm{Var}(X_{n1})$, and*

$$\alpha_n = \frac{1}{\sqrt{2\log n}}\,, \qquad \beta_n = \frac{1}{\alpha_n} - \frac{1}{2}\alpha_n(\log\log n + \log(4\pi))\,.$$

Let $a_n := \alpha_n s_n$ and $b_n := \beta_n s_n + \mu_n$. Then, we have $\forall x \in \mathbb{R}$:

$$\mathbb{P}(M_n \le a_n x + b_n) \longrightarrow \exp(-\exp(-x))\,.$$

All subsequent EVLTs stated in Sections 3.4, 3.5 and Chapters 4, 5 can be modified in the same way.

3.4 Arbitrary finite Coxeter groups

In the previous section, we established the EVLT for X_{inv} and X_{des} on sequences of classical Weyl groups. The EVLT for X_{des} is based only on applying Theorem 3.2.4 to the representation of X_{des} given in Corollary 2.2.5b). These arguments hold true for any sequence finite Coxeter groups where $\mathrm{Var}(X_{\mathrm{des}})$ grows linearly with respect to the rank, which is particularly the case for products of classical Weyl groups. Therefore, we can state:

Theorem 3.4.1. *Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of finite Coxeter groups with $\mathrm{rk}(W_n) = n\ \forall n \in \mathbb{N}$, which satisfies $\mathrm{Var}(X_{\mathrm{des}}) = \Theta(n)$. Let $k_n = \exp(o(n^{1/2}))$, let $(X_{nj})_{j=1,\dots,k_n}$ be a row-wise i.i.d. triangular array with $X_{n1} \stackrel{\mathcal{D}}{=} X_{\mathrm{des}}$ and let $M_n := \max\{X_{n1}, \dots, X_{nk_n}\}$. Let a_n, b_n be as in Theorem 3.3.2. Then,*

$$\mathbb{P}(M_n \le a_n x + b_n) \longrightarrow \exp(-\exp(-x)) \quad \forall x \in \mathbb{R}\,.$$

The EVLT for X_{inv} is based on Theorem 3.2.2. For arbitrary finite Coxeter groups, the condition $|X_k| = O(n^{-1/2}s_n)$ used in the proof of Theorem 3.3.2 is not trivially satisfied. For inversions, the $X_k = X_{\mathrm{inv}}^{(n,i)} - \mathbb{E}\left(X_{\mathrm{inv}}^{(n,i)}\right)$ can be bounded by the maximum degree $d_{\max}$ of W_n. Therefore, this condition is written more descriptively as

$$d_{\max} \lesssim \frac{s_n}{\sqrt{n}}\,. \tag{3.5}$$

Using the method of Theorem 3.3.2, we can state a general EVLT for X_{inv} on sequences of finite Coxeter groups.

Theorem 3.4.2. *Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of finite Coxeter groups with $n = \mathrm{rk}(W_n)$. Let $k_n = \exp(o(n^{1/3}))$, let $(X_{nj})_{j=1,\dots,k_n}$ be a row-wise i.i.d. triangular array with $X_{n1} \stackrel{\mathcal{D}}{=} X_{\mathrm{inv}}$ and let $M_n := \max\{X_{n1}, \dots, X_{nk_n}\}$. Let a_n, b_n be as in Theorem 3.3.2. If the condition (3.5) holds, then*

$$\mathbb{P}(M_n \le a_n x + b_n) \longrightarrow \exp(-\exp(-x)) \quad \forall x \in \mathbb{R}\,.$$

In what follows, we rephrase the condition (3.5) more descriptively for certain products of finite irreducible Coxeter groups.

3.4.1 Sequences of products of classical Weyl groups

Let $W_n = \prod_{i=1}^{l_n} W_{n,i}$, where each component $W_{n,i}$ is a classical Weyl group, and let $n = \mathrm{rk}(W_n) = \mathrm{rk}(W_{n,1}) + \ldots + \mathrm{rk}(W_{n,l_n})$ denote the total rank. Then,

$$\mathrm{Var}\left(X_{\mathrm{inv}}^{W_n}\right) = \sum_{i=1}^{l_n} \mathrm{Var}\left(X_{\mathrm{inv}}^{W_{n,i}}\right).$$

For each n and i, we have $\mathrm{Var}(X_{\mathrm{inv}}^{W_{n,i}}) = \Theta(\mathrm{rk}(W_{n,i}))$. However, the total variance $\mathrm{Var}(X_{\mathrm{inv}}^{W_n})$ is not of cubic order with respect to n. By Corollary 2.2.5a), $\mathrm{Var}(X_{\mathrm{inv}}^{W_n})$ still has an independent sum representation of n summands. The maximum degree $d_{\max} \leq 2\max\{\mathrm{rk}(W_{n,1}), \ldots, \mathrm{rk}(W_{n,l_n})\}$ bounds these summands. Therefore, omitting the factor 2 without asymptotic consequences, the condition (3.5) now reads

$$d_{\max} \lesssim \frac{1}{\sqrt{n}} \sqrt{\mathrm{rk}(W_{n,1})^3 + \ldots + \mathrm{rk}(W_{n,l_n})^3}\,. \tag{3.6}$$

This observation yields:

Theorem 3.4.3. *Let $W_n = \prod_{i=1}^{l_n} W_{n,i}$ be a sequence of products of classical Weyl groups, and let $(X_{nj})_{j=1,\ldots,k_n}$ be a row-wise i.i.d. triangular array with $X_{n1} \overset{\mathcal{D}}{=} X_{\mathrm{inv}}$. Let k_n, M_n, a_n, b_n be as in Theorem 3.3.2. If the condition (3.6) holds, then*

$$\mathbb{P}(M_n \leq a_n x + b_n) \longrightarrow \exp(-\exp(-x)) \quad \forall x \in \mathbb{R}\,.$$

3.4.2 Sequences of products involving dihedral groups

In this section, we consider sequences $W_n = \prod_{i=1}^{n} W_{n,i}$ of finite Coxeter groups consisting of dihedral components and classical Weyl group components. Since all dihedral groups have even-numbered ranks, it is not always feasible to construct a sequence $(W_n)_{n\in\mathbb{N}}$ with ranks covering all natural numbers. However, there is no issue since only the ratio between the ranks $\mathrm{rk}(W_n)$ and the row lengths k_n is significant.

Example 3.4.4. If all components are dihedral, i.e.,

$$W_n = \prod_{i=1}^{h_n} I_2(m_{n,i})$$

for some $(m_{n,i})_{n\in\mathbb{N}, i=1,\ldots,n}$ and a sequence $(h_n)_{n\in\mathbb{N}}$, then $\mathrm{rk}(W_n) = 2h_n$. Therefore, the condition for applying Theorem 3.3.2 is $h_n \gg \log(k_n)^3$.

Remark 3.4.5. Regarding W_n as in Example 3.4.4, it has been stated by Kahle & Stump [66, Corollary 3.2, 4.2] that for products of dihedral groups,

$$\mathrm{Var}(X_{\mathrm{inv}}) = \sum_{i=1}^{h_n} \frac{m_{n,i}^2 + 2}{12}, \qquad \mathrm{Var}(X_{\mathrm{des}}) = \sum_{i=1}^{h_n} \frac{1}{m_{n,i}}\,.$$

Furthermore, $I_2(m_{n,i})$ has degrees $2, m_{n,i}$. Therefore, the degrees of W_n are $2, \ldots, 2, m_{n,1}, \ldots, m_{n,h_n}$ with h_n twos. This information is now used to rephrase the condition (3.5) for mixed products of dihedral groups and classical Weyl groups.

Definition 3.4.6. Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of finite Coxeter groups and write $W_n = G_n \times I_n$, where G_n contains only classical Weyl group components and I_n contains only dihedral components. We use the following notation:

$$r_n := \operatorname{rk}(G_n)\,, \qquad R_n := \operatorname{rk}(W_n) = r_n + 2h_n\,,$$
$$G_n := \prod_{i=1}^{l_n} G_{n,i}\,, \qquad I_n := \prod_{i=1}^{h_n} I_2(m_{n,i})\,,$$
$$r_{\max} := \max\{\operatorname{rk}(G_{n,1}), \ldots, \operatorname{rk}(G_{n,l_n})\}\,, \qquad \mathcal{R}_n^2 := \sum_{i=1}^{l_n} \operatorname{rk}(G_{n,i})^3\,,$$
$$m_{\max} := \max\{m_{n,1}, \ldots, m_{n,h_n}\}\,, \qquad \mathcal{M}_n^2 := \sum_{i=1}^{h_n} m_{n,i}^2\,.$$

Furthermore, we write X_{inv}^G and X_{inv}^I for the number of inversions in the classical Weyl group components and in the dihedral components of W_n, respectively. As $\operatorname{rk}(W_n) = r_n + 2h_n$, the growth condition is that at least one of $r_n \gg \log(k_n)^3$ or $h_n \gg \log(k_n)^3$ holds, i.e., $\log(k_n) \ll (r_n \vee h_n)^{1/3}$.

Remark 3.4.7. Regardless of how G_n is composed, Theorem 2.2.2 gives

$$\mathbb{E}(X_{\text{inv}}^G) = \Theta(r_n^2)\,, \qquad \operatorname{Var}(X_{\text{inv}}^G) = \Theta(r_n^3)\,.$$

Combining this with Remark 3.4.5, we obtain $\operatorname{Var}(X_{\text{inv}}) = \Theta(\mathcal{R}_n^2 + \mathcal{M}_n^2)$. We note that by Theorem 2.1.19,

$$\mathcal{G}_{\text{inv}}(W_n; x) = \prod_{i=1}^{R_n} (1 + x + \ldots + x^{d_i - 1})\,,$$

where the degrees d_i encompass the degrees of the classical Weyl group components (bounded by $2r_{\max}$), h_n twos, and the numbers $m_{n1}, \ldots, m_{nh_n}$ (bounded by $m_{\max}$). For such composed groups, the sufficient condition (3.5) for the Gumbel behavior of X_{inv} is

$$r_{\max} \vee m_{\max} = O\left(\sqrt{R_n^{-1}(\mathcal{R}_n^2 + \mathcal{M}_n^2)}\right). \tag{3.7}$$

These observations are summarized as follows:

Theorem 3.4.8. *Let $W_n = G_n \times I_n$ be a sequence of finite Coxeter groups, where the classical components are pooled in G_n and the dihedral components are pooled in I_n. Let k_n be a sequence of integers satisfying $k_n = \exp\big(o((r_n \vee h_n)^{1/3})\big)$. Let $(X_{nj})_{j=1,\ldots,k_n}$ be a row-wise i.i.d. triangular array with $X_{n1} \stackrel{\mathcal{D}}{=} X_{\text{inv}}$ and let $M_n := \max\{X_{n1}, \ldots, X_{nk_n}\}$. Let a_n, b_n be as in Theorem 3.3.2. If the condition (3.7) holds, then*

$$\mathbb{P}(M_n \le a_n x + b_n) \longrightarrow \exp\big(-\exp(-x)\big) \quad \forall x \in \mathbb{R}\,.$$

For products consisting of only dihedral groups, i.e., $G_n = \emptyset$ and $W_n = \prod_{i=1}^{h_n} I_2(m_{n,i})$, the statement of Theorem 3.4.8 is simplified as follows:

Corollary 3.4.9. *Let* $W_n = \prod_{i=1}^{h_n} I_2(m_{n,i})$ *be a product of dihedral groups and* $k_n = \exp\Big(o\big(h_n^{1/3}\big)\Big)$. *Let* $(X_{nj})_{j=1,\dots,k_n}$, M_n, a_n, b_n *be as in Theorem 3.4.8. If*

$$m_{\max} \lesssim h_n^{-1/2}\mathcal{M}_n\,, \tag{3.8}$$

then $\mathbb{P}(M_n \le a_n x + b_n) \longrightarrow \exp(-\exp(-x))\ \forall x \in \mathbb{R}$.

Remark 3.4.10. The condition (3.8) is not trivial. Writing the orders of the dihedral components as a vector $\mathbf{m}_n = (m_{n,1}, \dots, m_{n,h_n})$, we get

$$\|\mathbf{m}_n\|_\infty \lesssim \frac{1}{\sqrt{n}}\|\mathbf{m}_n\|_2\,,$$

where $\|\cdot\|_\infty$ is the maximum norm and $\|\cdot\|_2$ is the euclidean norm. Since $\|\mathbf{m}_n\|_\infty \ge n^{-1/2}\|\mathbf{m}_n\|_2$ always holds, the condition (3.8) can be stated more precisely as

$$\|\mathbf{m}_n\|_\infty = \Theta\left(\frac{1}{\sqrt{n}}\|\mathbf{m}_n\|_2\right).$$

Remark 3.4.11. Regarding the EVLT for X_{des} on sequences of mixed products of Coxeter groups, we require $\mathrm{Var}(X_{\mathrm{des}}) = \Theta(r_n) + \sum_{i=1}^{h_n} m_{n,i}^{-1} \stackrel{!}{=} \Theta(R_n)$ according to Theorem 3.4.1. This is particularly satisfied if all $m_{n,i}$ are uniformly bounded.

3.5 Universal EVLT for other permutation statistics

The results of the previous section can be summarized as follows: For the number of descents, we have the EVLT on all finite Coxeter groups by Theorem 3.4.1. For the number of inversions, we have the EVLT on classical Weyl groups and composed finite Coxeter groups satisfying the regularity condition (3.5) by Theorem 3.4.2. However, the methods from large deviations theory employed for these results have specific requirements that are not satisfied in many situations. For the theorems introduced in Section 3.2, it is essential to have an independent sum decomposition, which is commonly derived from a factorization of the generating function for permutation statistics. For the two-sided Eulerian statistic X_T introduced in Definition 2.3.4, there is no explicit formula for its generating function, and no independent decomposition of X_T is available so far.

Moreover, each of the theorems in Section 3.2 imposes further technical assumptions. For example, the variances must have a suitable order to satisfy the control condition (3.1) in Theorem 3.2.2. The *number of cycles* on symmetric groups is an example of a permutation statistic whose generating function provides an independent sum decomposition, but the control condition (3.1) is not satisfied. For n-permutations, the number of cycles is commonly written as K_{0n}. According to [3, Eq. 1.27], the generating function of K_{0n} is

$$\mathcal{G}_{K_{0n}}(x) = \prod_{j=1}^{n}\left(1 - \frac{1}{j} + \frac{x}{j}\right),$$

therefore,

$$K_{0n} = \sum_{j=1}^{n} \mathrm{Bin}\left(1, \frac{1}{j}\right), \tag{3.9}$$

which is an independent sum. However, it implies

$$\operatorname{Var}(K_{0n}) = \sum_{j=1}^{n} \frac{j-1}{j^2} = \log(n)(1+o(1))\,.$$

The variance is too small to satisfy (3.1) with $\lambda_n = O(n^{-1/2})$. This also violates the condition (3.3), so Theorem 3.2.4 cannot be applied as well.

If the number of i.i.d. samples k_n in each row of a triangular array grows sufficiently slowly, then the CLT suggests that the rows with large k_n behave similarly to the standard normal distribution. In that case, it turns out that tail equivalence and Gumbel attraction can already be obtained from Berry–Esseen's bound. This gives a weaker but universal EVLT for a very broad class of families of distributions, including other permutation statistics such as the examples mentioned above. In other words, this EVLT strongly narrows down the possible sequences of distributions $(F_n)_{n\in\mathbb{N}}$ for which a triangular array of any size is not attracted to the Gumbel distribution.

Theorem 3.5.1. *Let $F_1, F_2, \ldots$ be a sequence of distributions which satisfy the Berry–Esseen bound*

$$\sup_{x\in\mathbb{R}} \left| \frac{F_n(x) - \mathbb{E}(F_n)}{\sigma(F_n)} - \Phi(x) \right| = O(n^{-1/2})\,,$$

where Φ is the CDF of $\mathrm{N}(0,1)$. Let $(X_{nj})_{j=1,\ldots,k_n}$ be a triangular array with $X_{n1} \sim F_n$ and let M_n, a_n, b_n be as in Theorem 3.3.2. If $k_n = O(n^{\varepsilon})$ for some $\varepsilon < 1/2$, then

$$\mathbb{P}(M_n \le a_n x + b_n) \longrightarrow \exp(-\exp(-x))\,.$$

Proof. Let $Y_n := \sigma(X_{n1})^{-1}\,(X_{n1} - \mathbb{E}(X_{n1}))$ and $Z \sim \mathrm{N}(0,1)$. Then, the Berry–Esseen bound is equivalent to

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(Y_n > x) - \mathbb{P}(Z > x)| = O(n^{-1/2})\,. \tag{3.10}$$

Now, we replace x with $x_n := \alpha_{k_n} x + \beta_{k_n}$ for fixed x, with $\alpha_{k_n}, \beta_{k_n}$ as in Theorem 3.3.2. Tail equivalence means

$$\frac{\mathbb{P}(Y_n > x_n)}{\mathbb{P}(Z > x_n)} = 1 + o(1) \iff 1 + \frac{\mathbb{P}(Y_n > x_n) - \mathbb{P}(Z > x_n)}{\mathbb{P}(Z > x_n)} = 1 + o(1)\,.$$

In light of (3.10), we have to show $\mathbb{P}(Z > x_n) \gg n^{-1/2}$. For monotonicity reasons, we can also assume that $k_n = \omega(n^{\delta})$ for some $\delta > 0$. From Mill's Ratio (see [76]), we can deduce

$$\begin{aligned}
\mathbb{P}(Z > x_n) &= 1 - \Phi(\alpha_{k_n} x + \beta_{k_n}) \\
&\sim \frac{1}{\alpha_{k_n} x + \beta_{k_n}} \varphi(\alpha_{k_n} x + \beta_{k_n}) \\
&= O\left(\frac{1}{\sqrt{\log(n)}}\right) \varphi\left(\frac{x}{\sqrt{2\varepsilon\log(n)}} + \sqrt{2\varepsilon\log(n)} - \frac{\log(4\pi\varepsilon\log(n))}{2\sqrt{2\varepsilon\log(n)}}\right) \\
&= O\left(\frac{1}{\sqrt{\log(n)}}\right) \exp\left(-\varepsilon\log(n) - \frac{1}{2}\log(4\pi\varepsilon\log(n)) + O\left(\frac{\log(\log(n))^2}{\log(n)}\right)\right) \\
&= O\left(\frac{1}{\sqrt{\log(n)}}\right) n^{-\varepsilon}(1+o(1))\,,
\end{aligned}$$

from which it follows that $\mathbb{P}(Z > x_n) \gg n^{-1/2}$. From here, the proof continues the same way as in Theorem 3.3.2. □

Remark 3.5.2. In particular, there is no asymptotically normal permutation statistic whose row-wise maximum is never attracted to the Gumbel distribution. Since permutation statistics are defined on finite probability spaces, the CLT commonly implies the Berry-Esseen bound.

Remark 3.5.3. The Berry-Esseen bound is a bound of the Kolmogorov distance:

$$d_K(Y_n, Z) = \sup_{x\in\mathbb{R}} |\mathbb{P}(Y_n \le x) - \mathbb{P}(Z \le x)|\,,$$

where Y_n and Z are given as in the proof of Theorem 3.5.1. In general, we can state that if $d_K(Y_n, Z) = O(n^{-r})$, with $0 < r \le 1/2$, then Theorem 3.5.1 holds for any $k_n = O(n^{\varepsilon})$ with $0 < \varepsilon < r$. In many situations, the Berry-Esseen bound is obtained not for the Kolmogorov distance, but for the Wasserstein distance:

$$d_W(Y_n, Z) = \int_{\mathbb{R}} |F_n(x) - \Phi(x)| \mathrm{d}x \overset{!}{=} O(n^{-1/2})\,.$$

For instance, [20] shows this bound for the two-sided Eulerian statistic X_T. By a standard argument (see, e.g., [65, Lemma 2]), it holds that $d_K(Y_n, Z) \le 2\sqrt{d_W(Y_n, Z)}$, which means that $d_W(Y_n, Z) = O(n^{-1/2})$ implies $d_K(Y_n, Z) = O(n^{-1/4})$, and Theorem 3.5.1 holds with $k_n = O(n^{\varepsilon})$, $0 < \varepsilon < 1/4$.

3.6 Open questions

Besides the aforementioned examples, there are many more interesting permutation statistics. An elaborate list of these is provided within the database [95]. Moreover, there are other algebraic, combinatorial, and probabilistic structures related to symmetric groups and the numbers of inversions and descents. In many cases, the random quantities are asymptotically normal and satisfy the Berry–Esseen bound, which yields a Gumbel behavior by Theorem 3.5.1 with a low bound on k_n. For each of these cases, it is an open question to obtain a subexponential bound on k_n, or at least one that permits the uniform triangular array. We give a few interesting examples with application interest.

Multisets: Conger & Viswanathan [27] studied the CLT for X_{inv} and X_{des} on permutations of multisets. A *multiset* takes the form $M := \{1^{n_1}, 2^{n_2}, \dots, h^{n_h}\}$ with $n_1, \dots, n_h \in \mathbb{N}$ and $n_1 + \ldots + n_h = n$. The permutations of this multiset are all maps $\pi\colon \{1, \dots, n\} \to M$ with $|\pi^{-1}(i)| = n_i \;\forall i = 1, \dots, h$. The numbers of inversions and descents on multisets are defined analogously to Definitions 2.1.1 and 2.1.2. The CLT on multisets is proved by use of size-bias couplings in [27, Theorems 2.12 and 2.16].

Conjugacy classes: The symmetric groups, as well as any other finite Coxeter groups, can be decomposed into their conjugacy classes. The conjugacy class of any $\pi \in S_n$ is given by $\{\sigma^{-1}\pi\sigma \mid \sigma \in S_n\}$. It is well known that the conjugacy class of $\pi \in S_n$ is identified by its cycle structure, and that there is a one-to-one relation between conjugacy classes and partitions $(n_1, \dots, n_n)$ with $n_1, \dots, n_n \in \{0, \dots, n\}$ and $\sum_{i=1}^{n} i n_i = n$. The asymptotic normality of X_{des} has been studied for certain sequences of conjugacy classes. See, e.g., [50] for conjugacy classes corresponding to large cycles, and more recently, [67] for conjugacy classes where the ratio of fixed points $\alpha_1(n) := n_1/n$ converges to some $\alpha \in [0,1]$.

Parabolic double cosets: For any group G and two subgroups $H, K \subseteq G$, the *double cosets* are the sets $HgK = \{hgk \mid h \in H, k \in K\}$ for fixed $g \in G$, and the set of double cosets is denoted by $H\backslash g/K := \{HgK \mid g \in G\}$. This is a concept similar to conjugacy classes. For a symmetric group S_n, the *parabolic subgroups* are given by $S_\lambda := S_{\lambda_1} \times \ldots \times S_{\lambda_I}$ for all partitions $\lambda = (\lambda_1, \ldots, \lambda_I)$, where $\lambda_1 \geq \ldots \geq \lambda_I \geq 1$ and $\lambda_1 + \ldots + \lambda_I = n$. Paguyo [83] studied the double cosets $S_\lambda \backslash S_n / S_\mu$ for any two parabolic subgroups S_λ, S_μ, and found that for short partitions with $I = o(n)$, the numbers of inversions and descents on a sequence of parabolic double cosets $S_\lambda \backslash S_n / S_\mu$ are asymptotically normal, see [83, Theorems 1.2 and 1.4]. By use of Stein's method and dependency graphs (cf. [60, Theorem 3.5]), these results also give an $O(n^{-1/2})$ bound for the Kolmogorov distance, for which we can apply Theorem 3.5.1.

Mallows distribution: All probabilistic results on finite Coxeter groups presented so far are based on drawing elements of these groups uniformly at random. However, there are other interesting probability distributions and random structures on these groups. Jimmy He [58] studied the CLT for the two-sided Eulerian statistic X_T under the so-called *Mallows distribution* μ_q, $q \in (0, \infty)$. This family of distributions dates back to [74]. See, e.g., [102] for a description of the Mallows model and its application interest. The probability mass function of μ_q is weighted by the number of inversions, i.e., for any element π of a finite Coxeter group,

$$\mathbb{P}_{\mu_q}(\{\pi\}) := \frac{q^{\mathrm{inv}(\pi)}}{Z_n(q)},$$

where $Z_n(q)$ is a normalization constant. If $W = S_n$ is a symmetric group, then $Z_n(q)$ can be explicitly stated as the *q-factorial* $Z_n(q) = [n]_q! := \prod_{k=1}^{n-1}(1 + q + \ldots + q^k)$. Obviously, $q = 1$ gives the uniform distribution. If q is small, then the distribution is concentrated around the neutral element. Furthermore, μ_q and $\mu_{1/q}$ are equidistributed, so it is sufficient to consider $q \in (0, 1]$ (see [58, Proposition 2.6]).

Jimmy He proved the asymptotic normality of X_T and $(X_{\mathrm{des}}, X_{\mathrm{ides}})$ on symmetric groups for fixed q, as well as for some regimes of variable q (see [58, Theorems 1.1 and 1.2]). A major part of this work is devoted to the limiting correlation $\rho := \lim_{n\to\infty} \mathrm{corr}(X_{\mathrm{des}}, X_{\mathrm{ides}})$, which is strictly positive for all $q < 1$. Similar to [27], the CLT is proved by constructing a size-bias coupling. Very recently, this method was transferred by Maxwell Sun [101] to the other classical Weyl groups B_n and D_n. A general classification of the validity of the CLT in the sense of [66] seems feasible, although in the existing proofs, the estimates required for the error terms in Stein's method are very laborious.

4 CLT and extremes of multivariate permutation statistics

In the previous chapter, we successfully proved that the extreme values of inversions or descents are attracted to the Gumbel distribution for most finite Coxeter groups, including the irreducible classical Weyl groups. In contrast to these individual statistics, we now aim to understand the asymptotic behavior of the *joint permutation statistic*

$$\begin{pmatrix} X_{\mathrm{inv}} \\ X_{\mathrm{des}} \end{pmatrix}.$$

In Section 2.3, we already introduced its CLT on the family of symmetric groups by Fang & Röllin [45]. It achieves an $O(n^{-1/2})$ rate of convergence for arbitrarily large tuples of doubly indexed permutation statistics, but it cannot be generalized to other classical Weyl groups.

Corollary 2.2.5 gives a decomposition of both X_{inv} and X_{des} into independent random variables, by decomposing their generating functions into simple polynomials. However, it is not possible to extend this statement to their joint distribution. Inversions and descents are not independent of one another. This is already clear from the fact that descents are special cases of inversions, and also established by the fact that the joint generating function

$$A_n(s,t) := \sum_{\pi \in S_n} s^{\mathrm{inv}(\pi)} t^{\mathrm{des}(\pi)}$$

does *not* factor into the polynomials given in Theorems 2.1.19 and 2.1.20. It is an interesting question whether $A_n(s,t)$ factors at all. We have tested a few values of s, t for $n = 3$ or $n = 4$, but no regularity has been detected so far.

Our goal is to obtain both a CLT and an EVLT for the joint distribution $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$. While Janson's dependency criterion [64, Theorem 2] allows to prove the CLT for X_{inv}, there is no multivariate version that allows to prove the CLT for $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$. Due to the dependence between X_{inv} and X_{des}, we cannot derive an independent decomposition of $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$ from Corollary 2.2.5. Moreover, to the best of our knowledge, there is no multivariate equivalent of the large deviations theorems in Section 3.2 that works for our purposes. Therefore, we cannot reuse the proof methods employed in Chapter 3. We need a different approach.

Recall that by (2.2),

$$X_{\mathrm{inv}} = \sum_{1 \le i < j \le n} \mathbf{1}\{Z_i > Z_j\}, \qquad X_{\mathrm{des}} = \sum_{i=1}^{n-1} \mathbf{1}\{Z_i > Z_{i+1}\}$$

for i.i.d. $Z_1, \ldots, Z_n \sim U(0,1)$. Based on these representations, the dependence between X_{inv} and X_{des} will be tackled by replacing X_{inv} with its so-called *Hájek projection* $\hat{X}_{\mathrm{inv}}$

based on $Z_1, \dots, Z_n$. We will justify that the extremes of $(X_{\text{inv}}, X_{\text{des}})^\top$ can be traced back to those of $(\hat{X}_{\text{inv}}, X_{\text{des}})^\top$. This process will lead to Theorem 4.4.1, the second main result of this thesis.

In Section 4.1, we investigate the Hájek projection of inversions and descents on symmetric groups, and in Section 4.2 we extend these observations to other classical Weyl groups. These groups are also equipped with a new family of probability measures, namely, the so-called p-biased signed permutations. In Section 4.3, we introduce a powerful Gaussian approximation by Chang *et al.* [17], which gives a new proof of the CLT for $(X_{\text{inv}}, X_{\text{des}})^\top$. Compared to [45], this approach achieves a weaker rate of convergence, but it also applies to other classical Weyl groups. The new EVLT for $(X_{\text{inv}}, X_{\text{des}})^\top$ is presented in detail in Section 4.4, which also discusses applications to other permutation statistics. Section 4.5 proposes some open questions suggested by simulations. Section 4.6 gathers the proofs of several lemmas in this chapter and the code of the simulations in Section 4.5.

4.1 The Hájek projection of inversions and descents

Instead of finding an exact decomposition of $(X_{\text{inv}}, X_{\text{des}})^\top$ into independent summands, we use an independent sum that gives a close enough approximation.

Definition 4.1.1. Let $Z_1, \dots, Z_n$ be independent random variables, and let $\mathcal{S}$ be the vector space of all $\sum_{i=1}^n g_i(Z_i)$ with real-valued functions $g_1, \dots, g_n$ and $\mathbb{E}(g_i^2(Z_i)) < \infty$. This is a subspace of the space of all square-integrable random variables. The projection of a random variable X onto $\mathcal{S}$ is called the *Hájek projection* of X (with respect to $Z_1, \dots, Z_n$) and it is explicitly given by

$$\hat{X} := \sum_{k=1}^{n} \mathbb{E}(X \mid Z_k) - (n-1)\mathbb{E}(X)\,.$$

The subtrahend ensures unbiasedness, i.e., $\mathbb{E}(\hat{X}) = \mathbb{E}(X)$. By the factorization lemma (also known as the Doob-Dynkin lemma, see, e.g., [90]), every $\mathbb{E}(X \mid Z_k)$ is a function of only Z_k. Therefore, the Hájek projection gives a sum of independent random variables. In applications, we have a sequence $X_1, X_2, \dots$ of random variables or vectors built on another sequence $Z_1, Z_2, \dots$ of independent variables. To decide whether the Hájek projection is a sufficiently accurate approximation, the following criterion is useful.

Theorem 4.1.2. (cf. [104], Theorem 11.2)
Let $(Z_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables, let $(X_n)_{n\geq 1}$ be another sequence of random variables and let $\hat{X}_n$ be the Hájek projection of X_n with respect to $Z_1, \dots, Z_n$ for each $n \in \mathbb{N}$. If $\operatorname{Var}(\hat{X}_n) \sim \operatorname{Var}(X_n)$ as $n \to \infty$, then

$$\frac{X_n - \mathbb{E}(X_n)}{\sigma(X_n)} = \frac{\hat{X}_n - \mathbb{E}(\hat{X}_n)}{\sigma(\hat{X}_n)} + o_{\mathbb{P}}(1)\,.$$

Thus, Theorem 4.1.2 states that if $\operatorname{Var}(\hat{X}_n)$ is approximately equal to $\operatorname{Var}(X_n)$, then the standardizations of X_n and $\hat{X}_n$ have the same asymptotics, since their difference vanishes in probability. If the condition of Theorem 4.1.2 is satisfied, then we also say that *the Hájek approximation is successful.* In particular, if $(\hat{X}_n)_{n\in\mathbb{N}}$ satisfies a CLT, then Theorem 4.1.2 guarantees that $(X_n)_{n\in\mathbb{N}}$ also satisfies a CLT.

In what follows, for a random variable X, we write Y for its standardization, that is, $Y = (X - \mathbb{E}(X))/\sigma(X)$. All occurring variances are guaranteed to be finite. In particular, Y_{inv} and Y_{des} are the standardizations of X_{inv} and X_{des}, respectively, and $\hat{Y}_{\mathrm{inv}}$ is that of $\hat{X}_{\mathrm{inv}}$. The dependence of these variables on the underlying Coxeter group or its rank is suppressed unless needed for clarification.

Next, we provide the Hájek projection of X_{inv} and verify the condition of $\mathrm{Var}(X_{\mathrm{inv}}) \sim \mathrm{Var}(\hat{X}_{\mathrm{inv}})$ stated in Theorem 4.1.2.

Lemma 4.1.3. *The Hájek projection $\hat{X}_{\mathrm{inv}}$ of X_{inv} is given by*

$$\hat{X}_{\mathrm{inv}} = \frac{n(n-1)}{4} + \sum_{k=1}^{n} (n - 2k + 1) Z_k \,,$$

and it holds that $\mathrm{Var}(X_{\mathrm{inv}}) \sim \mathrm{Var}(\hat{X}_{\mathrm{inv}})$ *as* $n \to \infty$.

Proof. By definition, the Hájek projection of X_{inv} based on $Z_1, \ldots, Z_n$ is

$$\hat{X}_{\mathrm{inv}} = \sum_{k=1}^{n} \mathbb{E}(X_{\mathrm{inv}} \mid Z_k) - (n-1)\mathbb{E}(X_{\mathrm{inv}})\,.$$

We first consider the conditional means $\mathbb{E}(X_{\mathrm{inv}} \mid Z_k)$ for $k = 1, \ldots, n$ and get

$$\mathbb{E}(X_{\mathrm{inv}} \mid Z_k) = \sum_{1 \le i < j \le n} \mathbb{P}(Z_i > Z_j \mid Z_k) = \sum_{1 \le i < j \le n} \begin{cases} 1/2, & k \notin \{i, j\} \\ Z_k, & k = i \\ 1 - Z_k, & k = j \end{cases} \,.$$

We fix $k \in \{1, \ldots, n\}$ and analyze the frequency of the three cases listed on the right-hand side. As $\{1, \ldots, n\} \setminus \{k\}$ has cardinality $n-1$, there are $\binom{n-1}{2}$ subsets $\{i, j\} \subseteq \{1, \ldots, n\} \setminus \{k\}$. For each of these, we have $\mathbb{P}(Z_i > Z_j \mid Z_k) = \mathbb{P}(Z_i > Z_j) = 1/2$ for independence reasons.

The non-trivial contributions arise in the other two cases. In case of $i = k$, there are $n-k$ indices j with $j > k$, for which we have $\mathbb{P}(Z_k > Z_j \mid Z_k) = \mathbb{P}(Z_j < Z_k \mid Z_k) = Z_k$, since $Z_k \sim U(0,1)$. Likewise, in case of $j = k$, there are $k-1$ indices i with $i < k$, which gives $\mathbb{P}(Z_i > Z_k \mid Z_k) = \mathbb{P}(Z_i > Z_k) = 1 - Z_k$. These contributions are illustrated in Figure 4.1.

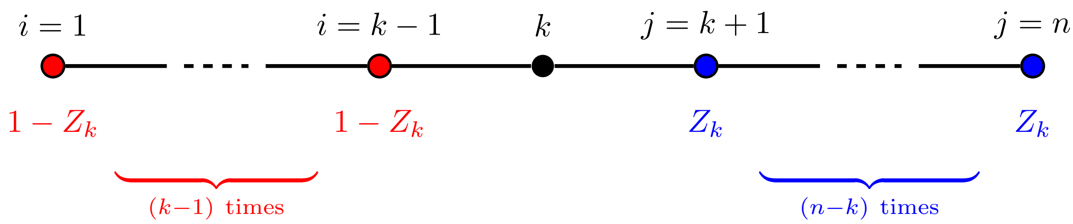


Figure 4.1: Display of the non-trivial contributions to $\mathbb{E}(X_{\mathrm{inv}} \mid Z_k)$ stemming from (i, k), $i = 1, \ldots, k-1$ (red), and (k, j), $j = k+1, \ldots, n$ (blue).

Therefore, we obtain

$$\begin{aligned}\mathbb{E}(X_{\mathrm{inv}} \mid Z_k) &= \frac{1}{2}\binom{n-1}{2} + (n-k)Z_k + (k-1)(1-Z_k) \\ &= \frac{1}{2}\binom{n-1}{2} + (n-2k+1)Z_k + (k-1)\,,\end{aligned}$$

from which we deduce that

$$\begin{aligned}\hat{X}_{\mathrm{inv}} &= \sum_{k=1}^{n} \mathbb{E}(X_{\mathrm{inv}} \mid Z_k) - (n-1)\mathbb{E}(X_{\mathrm{inv}}) \\ &= \frac{n}{2}\binom{n-1}{2} + \sum_{k=1}^{n}(n-2k+1)Z_k + \sum_{k=1}^{n}(k-1) - \frac{n-1}{2}\binom{n}{2} \\ &= \frac{1}{2}\binom{n}{2} + \sum_{k=1}^{n}(n-2k+1)Z_k\,. \qquad (4.1)\end{aligned}$$

The $(n-2k+1)Z_k$ are multiples of $U(0,1)$ that add up pairwise to $U(-j,j)$ with j being odd if n is even and vice versa, as displayed in Figure 4.2.

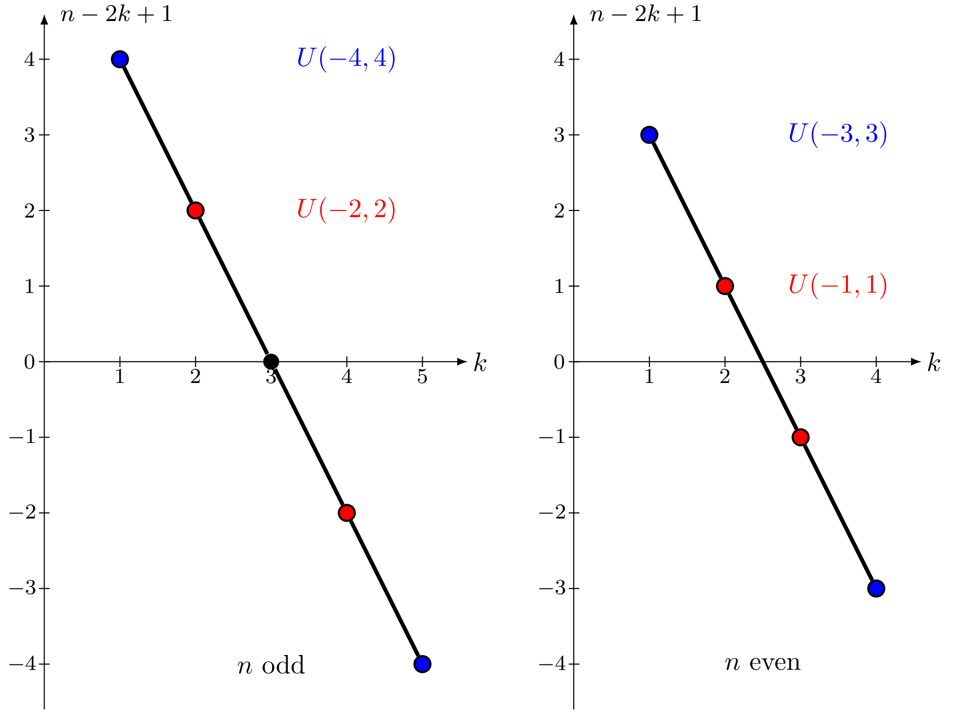


Figure 4.2: Display of the coefficients of Z_k appearing in $\hat{X}_{\mathrm{inv}}$. They add up in pairs to some $U(-j,j)$ with j even (left) or j odd (right). These pairs are highlighted in red and blue, respectively.

According to (4.1) and the independence of the Z_k, the variance of the Hájek projection is given by

$$\mathrm{Var}(\hat{X}_{\mathrm{inv}}) = \sum_{k=1}^{n} \mathrm{Var}\big((n-2k+1)Z_k\big)\,.$$

Due to $\mathrm{Var}(Z_k) = 1/12$, the result is

$$
\begin{aligned}
\mathrm{Var}(\hat{X}_{\mathrm{inv}}) &= \frac{1}{12}\sum_{k=1}^{n}(2k-n-1)^2 \\
&= \frac{1}{12}\sum_{k=1}^{n}\left(4k^2+(n+1)^2-4k(n+1)\right) \\
&= \frac{1}{12}\left(4\sum_{k=1}^{n}k^2+n(n+1)^2-4(n+1)\frac{n(n+1)}{2}\right) \\
&= \frac{1}{12}\left(4\frac{n(n+1)(2n+1)}{6}+n(n+1)^2-2(n+1)^2 n\right) \\
&= \frac{1}{12}\left(\frac{4}{3}n^3-n^3+O(n^2)\right) \\
&= \frac{1}{36}n^3-\frac{1}{3}n\,.
\end{aligned}
$$

By Theorem 2.2.2, we have $\mathrm{Var}(X_{\mathrm{inv}}) = \dfrac{1}{36}n^3 + \dfrac{3n^2-5n}{72}$ and therefore $\mathrm{Var}(X_{\mathrm{inv}}) \sim \mathrm{Var}(\hat{X}_{\mathrm{inv}})$ as $n \to \infty$. □

Remark 4.1.4. Interestingly, this approach fails for the descent statistic X_{des}. Repeating the considerations in the proof of Lemma 4.1.3 for descents, we first obtain

$$
\mathbb{E}(X_{\mathrm{des}} \mid Z_k) = \sum_{i=1}^{n-1}\mathbb{P}(Z_i > Z_{i+1} \mid Z_k) = \sum_{i=1}^{n-1}\begin{cases} 1/2, & k \notin \{i, i+1\} \\ Z_k, & k = i \\ 1-Z_k, & k = i+1 \end{cases}\,.
$$

Now, except for the boundary cases $k = 1$ and $k = n$, the summands for $k = i$ and $k = i+1$ are each used exactly once, so the Z_k in their sum $Z_k + (1 - Z_k)$ cancel out, leaving a constant. In total, we obtain

$$
\hat{X}_{\mathrm{des}} = Z_1 + (1 - Z_n) + \mathrm{const} \Longrightarrow \mathrm{Var}(\hat{X}_{\mathrm{des}}) = \frac{2}{12}\,,
$$

so the variance is not of the linear order of $\mathrm{Var}(X_{\mathrm{des}})$ stated in Theorem 2.2.3, which means Theorem 4.1.2 is not applicable. Hence, we do not obtain a fully independent sum decomposition of $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$. However, the success of the Hájek approximation for inversions is still sufficient for our needs, due to the following observation.

Obviously, $X_{\mathrm{des}} = \sum_{i=1}^{n-1}\mathbf{1}\{Z_i > Z_{i+1}\}$ is a sum of m-dependent random variables (to be precise, $m = 1$). The representation of inversions is not m-dependent, but its Hájek projection provides a close independent sum approximation. Therefore, our findings in this chapter will be based on the following consequence of Theorem 4.1.2 and Lemma 4.1.3:

Corollary 4.1.5. *For the standardized random vector $(Y_{\mathrm{inv}}, Y_{\mathrm{des}})^\top$ and the standardized Hájek projection $\hat{Y}_{\mathrm{inv}}$, we have*

$$
\begin{pmatrix} Y_{\mathrm{inv}} \\ Y_{\mathrm{des}} \end{pmatrix} = \begin{pmatrix} \hat{Y}_{\mathrm{inv}} \\ Y_{\mathrm{des}} \end{pmatrix} + o_{\mathbb{P}}(1)\,.
$$

A decomposition of $(\hat{X}_{\text{inv}}, X_{\text{des}})^\top$ *into* 1*-dependent summands is given by*

$$\begin{pmatrix} \hat{X}_{\text{inv}} \\ X_{\text{des}} \end{pmatrix} = \sum_{k=1}^{n-1} \begin{pmatrix} (n-2k+1)Z_k \\ \mathbf{1}\{Z_k > Z_{k+1}\} \end{pmatrix} + \begin{pmatrix} (1-n)Z_n + n(n-1)/4 \\ 0 \end{pmatrix}.$$

Accordingly, a 1*-dependent decomposition of* $(\hat{Y}_{\text{inv}}, Y_{\text{des}})^\top$ *can be found by standardization.*

4.2 Signed and even-signed permutation groups

Now, we obtain the Hájek approximation statement from Lemma 4.1.3 and Corollary 4.1.5 for the signed and even-signed permutation groups B_n and D_n. Recall the introduction of these groups in Definitions 2.1.11 and 2.1.12, as well as the counting of inversions and descents described in Remarks 2.1.15 and 2.1.16. Moreover, recall the notations $X^B_{\text{inv}}, X^B_{\text{des}}, X^D_{\text{inv}}$ and X^D_{des}, and the representations given in Remark 2.2.1, in particular,

$$X^B_{\text{inv}} = \sum_{1\le i<j\le n} \mathbf{1}\{Z_i > Z_j\} + \sum_{1\le i<j\le n} \mathbf{1}\{-Z_i > Z_j\} + \sum_{i=1}^{n} \mathbf{1}\{Z_i < 0\}\,, \tag{4.2a}$$

$$X^D_{\text{inv}} = \sum_{1\le i<j\le n} \mathbf{1}\{Z_i > Z_j\} + \sum_{1\le i<j\le n} \mathbf{1}\{-Z_i > Z_j\}\,. \tag{4.2b}$$

The third sum, which appears exclusively in X^B_{inv}, has only one summation index and turns out to be negligible for the asymptotic order of $\text{Var}(X^B_{\text{inv}})$. In fact, according to Theorem 2.2.2, both $\text{Var}(X^B_{\text{inv}})$ and $\text{Var}(X^D_{\text{inv}})$ satisfy

$$\text{Var}(X^B_{\text{inv}}), \text{Var}(X^D_{\text{inv}}) = \frac{n^3}{9} + O(n^2)\,.$$

Lemma 4.2.1. (see Section 4.6.1 for the proof)
Let X_{inv} *denote the number of inversions on either* B_n *or* D_n*, and let* $\hat{X}_{\text{inv}}$ *denote its Hájek projection. Then, again,* $\text{Var}(X_{\text{inv}}) \sim \text{Var}(\hat{X}_{\text{inv}})$*, and after standardization,*

$$Y_{\text{inv}} = \hat{Y}_{\text{inv}} + o_{\mathbb{P}}(1)\,.$$

Remark 4.2.2. Recall that by Remark 2.2.1,

$$X^B_{\text{des}} = \sum_{k=1}^{n-1} \mathbf{1}\{Z_k > Z_{k+1}\} + \mathbf{1}\{Z_1 < 0\}\,, \tag{4.3a}$$

$$X^D_{\text{des}} = \sum_{k=1}^{n-1} \mathbf{1}\{Z_k > Z_{k+1}\} + \mathbf{1}\{-Z_2 > Z_1\}\,. \tag{4.3b}$$

Again, this gives an m-dependent representation of X_{des} with $m = 1$. So overall, there is an m-dependent representation of $(\hat{X}_{\text{inv}}, X_{\text{des}})^\top$. Furthermore, since X^B_{des} and X^D_{des} are constructed similarly to X_{des} on S_n, the Hájek approximation of X^B_{des} or X^D_{des} fails for the same reasons as given in Remark 4.1.4.

4.2.1 Asymptotic vanishing of correlation

The covariance matrix of $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$ is not the identity matrix. However, we now show that on all classical Weyl groups, the correlation of X_{inv} and X_{des} vanishes in the limit. This means that if a normal limit of $(Y_{\mathrm{inv}}, Y_{\mathrm{des}})^\top$ exists, then it is the bivariate standard normal distribution. Moreover, we show that the asymptotic vanishing of correlation applies to $(\hat{Y}_{\mathrm{inv}}, Y_{\mathrm{des}})^\top$ as well. Then, in light of Remark 4.2.2 and Theorem 1.3.10, we can work towards proving both a CLT and an EVLT for $(X_{\mathrm{inv}}, X_{\mathrm{des}})$ using standard Gaussian approximation. In order to proceed, we need the covariance of $\hat{X}_{\mathrm{inv}}$ and X_{des}. In this process, we also calculate $\mathrm{Cov}(X_{\mathrm{inv}}, X_{\mathrm{des}})$, since to the best of our knowledge, this result is not available in the literature.

Lemma 4.2.3. (see Sections 4.6.2 and 4.6.3 for the proof)
Let W be one of the classical Weyl groups S_n, B_n, or D_n, and consider the random variables $X_{\mathrm{inv}}, X_{\mathrm{des}}, \hat{X}_{\mathrm{inv}}$ on W. Then,

(a) $$\mathrm{Cov}(X_{\mathrm{inv}}, X_{\mathrm{des}}) = \begin{cases} (n-1)/4, & W = S_n \\ n/4, & W = B_n, D_n \end{cases}.$$

(b) $$\mathrm{Cov}(\hat{X}_{\mathrm{inv}}, X_{\mathrm{des}}) = \begin{cases} (n-1)/6, & W = S_n \\ (n-1)/6 + 1/4, & W = B_n \\ (n-2)/12, & W = D_n \end{cases}.$$

Corollary 4.2.4. *Since $\mathrm{Var}(X_{\mathrm{inv}})\mathrm{Var}(X_{\mathrm{des}}) = \Theta(n^4)$ according to Theorems 2.2.2 and 2.2.3, and since the same holds true if $\mathrm{Var}(X_{\mathrm{inv}})$ is replaced with $\mathrm{Var}(\hat{X}_{\mathrm{inv}})$, we conclude from Lemma 4.2.3 that*

$$\mathrm{corr}(X_{\mathrm{inv}}, X_{\mathrm{des}}) = \frac{\mathrm{Cov}(X_{\mathrm{inv}}, X_{\mathrm{des}})}{\sigma(X_{\mathrm{inv}})\sigma(X_{\mathrm{des}})} = \Theta(1/n)\,,$$
$$\mathrm{corr}(\hat{X}_{\mathrm{inv}}, X_{\mathrm{des}}) = \Theta(1/n)\,, \qquad n \to \infty\,.$$

4.2.2 Signs with random bias

So far, we have assumed that all elements of a signed permutation group are drawn uniformly at random, which implies an equally probable choice of positive and negative signs. On the groups B_n and D_n, we can use a biased choice of the signs, saying that each entry $k \in \{1, \dots, n\}$ receives a negative sign with probability $p \in [0,1]$ and a positive sign with probability $1-p$. This gives rise to a family of probability measures on B_n and D_n, in which the case $p = 1/2$ corresponds to the uniform distribution, while in the case $p = 0$, all mass is on the symmetric group $S_n \subseteq B_n, D_n$. In a probabilistic sense, we obtain a continuous transition between the symmetric groups and the other classical Weyl groups.

Definition 4.2.5. Let $p \in [0,1]$ and $q := 1-p$. Then, the group of *p-biased signed permutations* is the group B_n equipped with the probability measure according to the above, i.e., for any $\pi \in B_n$ (using the convention $0^0 := 1$),

$$\mathbb{P}(\{\pi\}) = \frac{1}{n!} p^{\mathrm{neg}(\pi)} q^{n-\mathrm{neg}(\pi)}\,,$$

where $\mathrm{neg}(\pi)$ denotes the number of negative signs in π. Therefore, the entries of π can be represented by random variables $Z_1, \dots, Z_n \sim U \cdot R(p)$, where $U \sim U(0,1)$ and $R(p)$ is independent of U with

$$\mathbb{P}(R(p) = -1) = p\,, \qquad \mathbb{P}(R(p) = 1) = q\,.$$

For the distribution of $U \cdot R(p)$, we simply write $\mathrm{GR}(p)$ (generalized Rademacher with parameter p). The CDF of this distribution is

$$F_p(z) = \begin{cases} pz + p, & -1 \le z \le 0 \\ qz + p, & 0 \le z \le 1 \end{cases}.$$

Accordingly, the Lebesgue density is

$$f_p(z) = p\mathbf{1}\{-1 < z < 0\} + q\mathbf{1}\{0 < z < 1\}\,.$$

A corresponding probability distribution on the even-signed permutation group D_n is obtained by first choosing the unsigned permutation $|\pi| \in S_n$ uniformly at random, and then assigning $n-1$ signs for the entries $\pi(1), \dots, \pi(n-1)$ with p-bias, and finally specifying the sign of $\pi(n)$ so that there is an even number of negative signs.

Example 4.2.6. Note the special cases $\mathrm{GR}(0) = U(0,1)$, $\mathrm{GR}(1/2) = U(-1,1)$ and $\mathrm{GR}(1) = U(-1,0)$. Figure 4.3 illustrates F_p for $p = 1/4$ and $p = 3/4$.

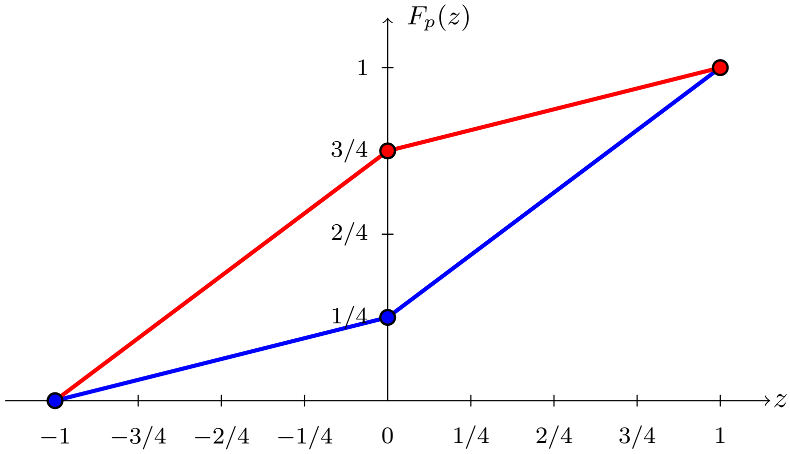


Figure 4.3: Probability distribution functions of generalized Rademacher distributions for $p = 1/4$ (blue) and $p = 3/4$ (red).

Remark 4.2.7. The representations (4.2a), (4.2b) for inversions and (4.3a), (4.3b) for descents are still valid. We recalculate the mean and the variance of the number of inversions for all p. First, we observe that for any $i < j$,

$$\begin{aligned}
\mathbb{P}(-Z_i > Z_j) &= \underbrace{\mathbb{P}(-Z_i > Z_j, Z_i > 0, Z_j > 0)}_{=0} + \underbrace{\mathbb{P}(-Z_i > Z_j, Z_i < 0, Z_j < 0)}_{=p^2} \\
&\quad + \mathbb{P}(-Z_i > Z_j, Z_i > 0, Z_j < 0) + \mathbb{P}(-Z_i > Z_j, Z_i < 0, Z_j > 0) \\
&= p^2 + \mathbb{P}(Z_i > 0, Z_j < 0, Z_i < |Z_j|) + \mathbb{P}(Z_i < 0, Z_j > 0, |Z_i| > Z_j) \\
&= p^2 + \frac{1}{2}pq + \frac{1}{2}pq = p^2 + pq = p\,.
\end{aligned}$$

Then, it follows directly from (4.2a) and (4.2b) that

$$\mathbb{E}\left(X_{\mathrm{inv}}^{B}\right) = \binom{n}{2}\left(p+\frac{1}{2}\right) + np\,, \qquad \mathbb{E}\left(X_{\mathrm{inv}}^{D}\right) = \binom{n}{2}\left(p+\frac{1}{2}\right).$$

Lemma 4.2.8. (see Section 4.6.4 for the proof)
On the p-biased (even-)signed permutation groups, we have

$$\begin{aligned}
\mathrm{Var}\left(X_{\mathrm{inv}}^{B}\right) &= \left(-\frac{1}{3}p^2+\frac{1}{3}p+\frac{1}{36}\right)n^3 + \left(-\frac{1}{2}p^2+\frac{1}{2}p+\frac{1}{24}\right)n^2 \\
&\quad + \left(-\frac{1}{6}p^2+\frac{1}{6}p-\frac{5}{72}\right)n\,, \\
\mathrm{Var}\left(X_{\mathrm{inv}}^{D}\right) &= \left(-\frac{1}{3}p^2+\frac{1}{3}p+\frac{1}{36}\right)n^3 + \left(\frac{1}{2}p^2-\frac{1}{2}p+\frac{1}{24}\right)n^2 \\
&\quad + \left(-\frac{1}{6}p^2+\frac{1}{6}p-\frac{5}{72}\right)n\,.
\end{aligned}$$

In particular, this corresponds with Theorem 2.2.2 if $p=0$ or $p=1/2$. For the variance of the Hájek projection, we fortunately get the same leading term for all p.

Lemma 4.2.9. (see Section 4.6.5 for the proof)
On the p-biased (even-)signed permutation groups, we also have

$$\mathrm{Var}(\hat{X}_{\mathrm{inv}}) = \left(-\frac{1}{3}p^2+\frac{1}{3}p+\frac{1}{36}\right)n^3 + O(n^2)\,,$$

so Theorem 4.1.2 applies again.

The leading term, as a function of p, has no roots in $[0,1]$ and assumes its global maximum at $p=1/2$, which is the unbiased case. This means that the order of $\mathrm{Var}(X_{\mathrm{inv}})$ and $\mathrm{Var}(\hat{X}_{\mathrm{inv}})$ is guaranteed to be cubic.

From Lemma 4.2.9, we obtain an extension of Corollary 4.1.5, which we present as a general statement for all three families of classical Weyl groups.

Corollary 4.2.10. *Let W be a classical Weyl group of rank n. Let*

$$Z_0 := \begin{cases} -\infty, & W = S_n \\ 0, & W = B_n \\ -Z_2, & W = D_n \end{cases}.$$

Then,

$$\begin{pmatrix}\hat{X}_{\mathrm{inv}}\\ X_{\mathrm{des}}\end{pmatrix} = \begin{pmatrix}\mathbb{E}(X_{\mathrm{inv}}\mid Z_1)\\ \mathbf{1}\{Z_0>Z_1\}\end{pmatrix} + \ldots + \begin{pmatrix}\mathbb{E}(X_{\mathrm{inv}}\mid Z_{n-1})\\ \mathbf{1}\{Z_{n-2}>Z_{n-1}\}\end{pmatrix} + \begin{pmatrix}\mathbb{E}(X_{\mathrm{inv}}\mid Z_n)-(n-1)\mathbb{E}(X_{\mathrm{inv}})\\ \mathbf{1}\{Z_{n-1}>Z_n\}\end{pmatrix}$$

is a 1-dependent decomposition of $(\hat{X}_{\mathrm{inv}}, X_{\mathrm{des}})^\top$, in analogy to Corollary 4.1.5. On B_n and D_n, this applies with any sign bias.

Lemma 4.2.11. (see Section 4.6.6 for the proof)
The statement of Corollary 4.2.4 extends to the groups B_n and D_n with any sign bias. To be precise,

$$\begin{aligned}
\mathrm{Cov}(X_{\mathrm{inv}}^B, X_{\mathrm{des}}^B) &= \mathrm{Cov}(X_{\mathrm{inv}}^D, X_{\mathrm{des}}^D) = \frac{n-1}{4} + pq\,, \\
\mathrm{Cov}(\hat{X}_{\mathrm{inv}}^B, X_{\mathrm{des}}^B) &= \frac{n-1}{6} + pq\,, \\
\mathrm{Cov}(\hat{X}_{\mathrm{inv}}^D, X_{\mathrm{des}}^D) &= (n-1)\left(-\frac{2}{3}p^2 + \frac{p}{6} + \frac{1}{6}\right) - \frac{5}{6}pq - \frac{p}{4}\,.
\end{aligned}$$

4.2.3 Products of classical Weyl groups

At last, we investigate whether the Hájek projection of X_{inv} works on *products of classical Weyl groups*. Let $W = \prod_{i=1}^{l} W_i$ be such a product, where each W_i is one of the groups S_n, B_n, or D_n. Then, by Remark 2.2.4, we know that

$$X_{\mathrm{inv}}^W = \sum_{i=1}^{l} X_{\mathrm{inv}}^{W_i}$$

is a sum of independent random variables, yielding

$$\mathrm{Var}(X_{\mathrm{inv}}^W) = \sum_{i=1}^{l} \mathrm{Var}(X_{\mathrm{inv}}^{W_i})\,.$$

Let $X_{\mathrm{inv}}^{W_i}$ be constructed from $Z_1^{(i)}, \dots, Z_{n_i}^{(i)}$, where n_i denotes the number of elements on which the group W_i acts, and each $Z_j^{(i)}$ is $\mathrm{GR}(p_i)$ for some $p_i \in [0,1]$, and the entire collection of all $Z_j^{(i)}$ is independent. Setting $n := n_1 + \ldots + n_l$, the overall Hájek projection $\hat{X}_{\mathrm{inv}}^W$ of X_{inv}^W is

$$\hat{X}_{\mathrm{inv}}^W = \sum_{i=1}^{l} \sum_{j=1}^{n_i} \mathbb{E}\left(X_{\mathrm{inv}}^W \mid Z_j^{(i)}\right) - (n-1)\mathbb{E}(X_{\mathrm{inv}}^W)\,.$$

The conditional mean $\mathbb{E}\left(X_{\mathrm{inv}}^W \mid Z_j^{(i)}\right)$ can be decomposed further into

$$\mathbb{E}\left(X_{\mathrm{inv}}^W \mid Z_j^{(i)}\right) = \sum_{k=1}^{l} \mathbb{E}\left(X_{\mathrm{inv}}^{W_k} \mid Z_j^{(i)}\right).$$

If $k \neq i$, then $X_{\mathrm{inv}}^{W_k}$ is independent of $Z_j^{(i)}$, which means $\mathbb{E}\left(X_{\mathrm{inv}}^{W_k} \mid Z_j^{(i)}\right) = \mathbb{E}(X_{\mathrm{inv}}^{W_k})$ is constant in this case. We therefore obtain

$$\mathrm{Var}(\hat{X}_{\mathrm{inv}}^W) = \sum_{i=1}^{l} \underbrace{\sum_{j=1}^{n_i} \mathrm{Var}\left(\mathbb{E}\left(X_{\mathrm{inv}}^{W_i} \mid Z_j^{(i)}\right)\right)}_{=\mathrm{Var}(\hat{X}_{\mathrm{inv}}^{W_i})} = \sum_{i=1}^{l} \mathrm{Var}(\hat{X}_{\mathrm{inv}}^{W_i})\,.$$

For any W_i, we have $\mathrm{Var}(X_{\mathrm{inv}}^{W_i}) \sim \mathrm{Var}(\hat{X}_{\mathrm{inv}}^{W_i})$. Furthermore, all variances are cubic as seen in Theorem 2.2.2 and Lemma 4.2.9, i.e., we have

$$\mathrm{Var}(X_{\mathrm{inv}}^{W_i}), \mathrm{Var}(\hat{X}_{\mathrm{inv}}^{W_i}) = c_i n_i^3 + O(n_i^2)\,,$$

where $c_i = -\frac{1}{3}p_i^2 + \frac{1}{3}p_i + \frac{1}{36}$. It is seen from the proofs of Lemma 4.1.3 and Lemma 4.2.9 that $\mathrm{Var}(X_{\mathrm{inv}})$ and $\mathrm{Var}(\hat{X}_{\mathrm{inv}})$ differ in their quadratic terms, i.e., $\mathrm{Var}(X_{\mathrm{inv}}) - \mathrm{Var}(\hat{X}_{\mathrm{inv}}) = \Theta(n^2)$. So, we can write

$$\begin{aligned}\mathrm{Var}(X_{\mathrm{inv}}^W) &= \sum_{i=1}^{l} c_i n_i^3 + \alpha_i n_i^2 + O(n_i)\,,\\ \mathrm{Var}(\hat{X}_{\mathrm{inv}}^W) &= \sum_{i=1}^{l} c_i n_i^3 + \beta_i n_i^2 + O(n_i)\,,\end{aligned} \tag{4.4}$$

with $\alpha_i \neq \beta_i$ for all i.

Consider a sequence $(W_n)_{n\in\mathbb{N}}$ of products as introduced above, assuming that the number l of components remains bounded. Then, it is seen that the equivalence of $\mathrm{Var}(X_{\mathrm{inv}}^W)$ and $\mathrm{Var}(\hat{X}_{\mathrm{inv}}^W)$ is preserved, since the cubic terms are equal and cannot be dominated by the quadratic terms.

Corollary 4.2.12. *Due to the above considerations,* $\mathrm{Var}(X_{\mathrm{inv}}^W) \sim \mathrm{Var}(\hat{X}_{\mathrm{inv}}^W)$ *and* $Y_{\mathrm{inv}}^W = \hat{Y}_{\mathrm{inv}}^W + o_{\mathbb{P}}(1)$ *hold for finite products of classical Weyl groups.*

4.3 High-dimensional Gaussian approximation and CLT

In the following, we first establish a CLT and then an EVLT for $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$ by using the 1-dependent decomposition of $(\hat{X}_{\mathrm{inv}}, X_{\mathrm{des}})^\top$ given in Corollary 4.2.10, and by applying a recent CLT for m-dependent triangular arrays by Chang *et al.* [17]. In the proof of the univariate EVLT for X_{inv} and X_{des} (see Theorems 3.3.2, 3.4.1, and 3.4.2), it was essential to establish tail equivalence to the standard normal distribution. In other words, we obtained a Gaussian approximation of the tails of X_{inv} and X_{des}. Accordingly, we need a Gaussian approximation of the tail of the joint bivariate distribution.

The classical CLT states that for a sequence or a triangular array $(X_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}^d$ and an appropriate d-variate normal distribution $\mathcal{N}$,

$$\sup_{\mathbf{x}\in\mathbb{R}^d} |\mathbb{P}(X_n \le \mathbf{x}) - \mathbb{P}(\mathcal{N} \le \mathbf{x})| = \sup_{A=(-\infty,\mathbf{x}]^d} |\mathbb{P}(X_n \in A) - \mathbb{P}(\mathcal{N} \in A)|$$

vanishes as $n \to \infty$ under suitable conditions. Even stronger results are obtained if the system of negative orthants $(-\infty, \mathbf{x}]^d$ is extended to, e.g., the class of all hyperrectangles or even all convex sets. These two classes also contain the positive orthants $[\mathbf{x}, \infty)^d$ which are pivotal for EVLTs. The analysis of the rate of convergence on larger systems has mostly been performed for sums of independent random vectors. See, e.g., [6] for an overview. Due to a significant application interest (see, e.g., [18, 107]), it was also investigated what happens with triangular arrays that grow in dimension. To explain this formally, we consider uniform triangular arrays $\left(X_t^{(n)}\right)_{t=1,\dots,n}$ where the entries $X_1^{(n)}, \dots, X_n^{(n)}$ are *centered* random vectors in $\mathbb{R}^{\mathfrak{p}}$, and $\mathfrak{p} = \mathfrak{p}(n)$ is allowed to grow in n. This implicitly covers the case of fixed dimensions as well, by repeating the components of a vector. Now, write

$$X^{(n)} := \sum_{t=1}^{n} X_t^{(n)}, \qquad \Sigma^{(n)} := \mathrm{Var}(X^{(n)})\,. \tag{4.5}$$

For this *high-dimensional context,* we highlight the work of Chernozhukov, Chetverikov, and Kato [22] as a seminal work giving a Gaussian approximation for the *tails* of sums of independent random vectors. In [23], the authors extended their work from positive orthants to hyperrectangles and sparsely convex sets. In recent years, there has been a steady improvement in the rate of convergence and the growth of dimension, see [24, 30, 44, 68, 70]. These results can be equivalently formulated in terms of bounding

$$\left|\mathbb{P}\left(X^{(n)} > c_{1-\alpha}\right) - \alpha\right|,$$

where $c_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of $\|\mathcal{N}\|_\infty$ for the Gaussian counterpart $\mathcal{N} \sim \mathrm{N}(0, \Sigma^{(n)})$. As this Gaussian distribution is not known in statistical applications, it is also of interest to obtain analogous bounds where $c_{1-\alpha}$ is the quantile of a bootstrap approximation of $\mathcal{N}$. For this, see [24, 25, 35, 36, 73].

While the aforementioned works focused on independent random vectors, the work of Chang *et al.* [17] addresses several frameworks of dependence within $X_1^{(n)}, \ldots, X_n^{(n)}$. This includes m-dependence, which exactly fits our interests in light of Corollary 4.1.5 and Remark 4.2.2. The benefit of [17] is the approximation of

$$r_n(\mathcal{A}^{\mathrm{re}}) := \sup_{A \in \mathcal{A}^{\mathrm{re}}} \left|\mathbb{P}(X^{(n)} \in A) - \mathbb{P}(\mathcal{N} \in A)\right|, \tag{4.6}$$

where $\mathcal{N} \sim \mathrm{N}(0, \Sigma^{(n)})$, and $\mathcal{A}^{\mathrm{re}}$ is the system of all hyperrectangles, including infinite bounds, i.e.,

$$\mathcal{A}^{\mathrm{re}} := \left\{\{\mathbf{w} \in \mathbb{R}^{\mathfrak{p}} \colon \mathbf{a} \leq \mathbf{w} \leq \mathbf{b}\} \mid \mathbf{a}, \mathbf{b} \in [-\infty, \infty]^{\mathfrak{p}}\right\}.$$

So, by taking the supremum in (4.6) over all negative orthants $(-\infty, \mathbf{x}]$, this will allow to extend the bivariate CLT for $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$ beyond symmetric groups. But it will also allow for Gaussian approximation of tails by taking the supremum in (4.6) over all positive orthants $[\mathbf{x}, \infty)$. Furthermore, Chang *et al.* elaborate in [17, Section 2.1] that the asymptotics of $r_n(\mathcal{A}^{\mathrm{re}})$ can be reduced to those of

$$r_n := \sup_{\substack{\mathbf{u} \in \mathbb{R}^{\mathfrak{p}} \\ \nu \in [0,1]}} \left|\mathbb{P}\left(\sqrt{\nu} X^{(n)} + \sqrt{1-\nu}\mathcal{N} \leq \mathbf{u}\right) - \mathbb{P}(\mathcal{N} \leq \mathbf{u})\right|.$$

The following two conditions are imposed on $X_t^{(n)} = \left(X_{t,1}^{(n)}, \ldots, X_{t,\mathfrak{p}}^{(n)}\right)^\top$, $t = 1, \ldots, n$.

Condition 1: There exists a sequence of constants $\mathfrak{B}_n \geq 1$ and a universal constant $\gamma \geq 1$ so that for all $j = 1, \ldots, \mathfrak{p}$:

$$\mathbb{E}\left(\exp\left(\left|X_{t,j}^{(n)}\right|^\gamma \mathfrak{B}_n^{-\gamma}\right)\right) \leq 2.$$

Condition 2: There exists a constant $K > 0$ so that for all $n \in \mathbb{N}$ and $j = 1, \ldots, \mathfrak{p}$:

$$\min_{j=1,\ldots,\mathfrak{p}} \mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n X_{t,j}^{(n)}\right) \geq K.$$

Under these two conditions, Chang *et al.* [17] provide a bound of r_n for random vectors $X^{(n)} = X_1^{(n)} + \ldots + X_n^{(n)}$ with sparse dependency structure. A *dependency graph* G_n

on the node set $\{1,\dots,n\}$ consists of all edges (i,j) for which $X_i^{(n)}$ and $X_j^{(n)}$ are dependent. Let Δ_n be the maximum degree of G_n and let Δ_n^* be the maximum degree of the 2-reachability graph of G_n. If Δ_n and Δ_n^* are not too large (i.e., if the graphs G_n are not too dense as $n \to \infty$), then the Gaussian approximation error r_n can be bounded as follows:

Theorem 4.3.1. (see [17], Theorem 2)
Let $(X_t^{(n)})_{t=1,\dots,n}$ be a triangular array of high-dimensional random vectors, i.e., for fixed $n \in \mathbb{N}$, we have $X_1^{(n)},\dots,X_n^{(n)} \in \mathbb{R}^{\mathfrak{p}}$ with $\mathfrak{p} = \mathfrak{p}(n) \gg n^\kappa$ for some constant $\kappa > 0$. Let Δ_n, Δ_n^* be as above. Under the conditions 1 and 2, it holds that

$$r_n \lesssim \frac{\mathfrak{B}_n(\Delta_n\Delta_n^*)^{1/3}\log(\mathfrak{p})^{7/6}}{n^{1/6}}\,.$$

The same statement applies to $r_n(\mathcal{A}^{\mathrm{re}})$.

If $\mathfrak{p}$ remains fixed, we can artificially repeat the vector components (say, n times) and therefore, the requirement $\mathfrak{p} \gg n^\kappa$ can be removed. Moreover, if the triangular array $(X_t^{(n)})_{t=1,\dots,n}$ is m-dependent, then Δ_n and Δ_n^* are both bounded in the way of $\Delta_n \le m$ and $\Delta_n^* \le 2m$. We obtain the following corollary.

Corollary 4.3.2. *Let $(X_t^{(n)})_{t=1,\dots,n}$ be a triangular array of mean zero random vectors in fixed dimension $\mathfrak{p}$ and suppose that each row $X_1^{(n)},\dots,X_n^{(n)}$ is m-dependent with a global constant $m \in \mathbb{N}$. Under the conditions 1 and 2, it holds that*

$$r_n(\mathcal{A}^{\mathrm{re}}) \lesssim \frac{\mathfrak{B}_n\log(n)^{7/6}}{n^{1/6}}\,, \qquad n \to \infty\,.$$

Remark 4.3.3. Condition 1 means sub-Gaussianity, i.e., by Markov's inequality,

$$\forall u > 0\colon \quad \mathbb{P}\left(\left|X_{t,j}^{(n)}\right| > u\right) \le 2\exp(-u^\gamma \mathfrak{B}_n^{-\gamma})\,.$$

As stated by Chang *et al.* [17, p. 5], we can choose $\gamma = 2$ and $\mathfrak{B}_n = O(1)$ for sub-Gaussian variables, especially for bounded variables like $X_{\mathrm{des}}, X_{\mathrm{inv}}$, and the Hájek projection $\hat{X}_{\mathrm{inv}}$. Condition 2 implies non-degeneracy, which is obviously true in our setting. Therefore, we can establish a joint CLT for inversions and descents.

Theorem 4.3.4. *For the families S_n, B_n, D_n of classical Weyl groups, the joint distribution of $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$ satisfies the CLT. In detail,*

$$(Y_{\mathrm{inv}}, Y_{\mathrm{des}})^\top = \left(\frac{X_{\mathrm{inv}} - \mathbb{E}(X_{\mathrm{inv}})}{\sigma(X_{\mathrm{inv}})}, \frac{X_{\mathrm{des}} - \mathbb{E}(X_{\mathrm{des}})}{\sigma(X_{\mathrm{des}})}\right)^\top \xrightarrow{\mathcal{D}} \mathrm{N}_2(0, \mathrm{I}_2)\,.$$

Proof. Due to Corollary 4.2.10 and Slutsky's theorem, it is sufficient to show that $(\hat{Y}_{\mathrm{inv}}, Y_{\mathrm{des}})^\top \xrightarrow{\mathcal{D}} \mathrm{N}_2(0, \mathrm{I}_2)$. On the symmetric groups, we have by Corollary 4.1.5 that

$$\begin{pmatrix} \hat{X}_{\mathrm{inv}} - \mathbb{E}(\hat{X}_{\mathrm{inv}}) \\ X_{\mathrm{des}} - \mathbb{E}(X_{\mathrm{des}}) \end{pmatrix} = \sum_{k=1}^{n-1} \begin{pmatrix} (n-2k+1)(Z_k - 1/2) \\ \mathbf{1}\{Z_k > Z_{k+1}\} - 1/2 \end{pmatrix} + \begin{pmatrix} (1-n)(Z_n - 1/2) \\ 0 \end{pmatrix}$$

is a sum of 1-dependent random vectors with mean zero. Setting

$$X_k^{(n)} := \begin{pmatrix} (n-2k+1)(Z_k - 1/2)/\sigma(\hat{X}_{\mathrm{inv}}) \\ (\mathbf{1}\{Z_k > Z_{k+1}\} - 1/2)/\sigma(X_{\mathrm{des}}) \end{pmatrix}, \qquad k = 1,\dots,n-1,$$

$$X_n^{(n)} := \begin{pmatrix} (1-n)(Z_n - 1/2)/\sigma(\hat{X}_{\mathrm{inv}}) \\ 0 \end{pmatrix},$$

we obtain the representation $(\hat{Y}_{\mathrm{inv}}, Y_{\mathrm{des}})^\top = \sum_{k=1}^n X_k^{(n)} =: X^{(n)}$. On the other classical Weyl groups, we find an analogous representation by Corollary 4.2.10. The covariance matrix of $X^{(n)}$ (see (4.5)) is given by $\Sigma^{(n)} = \begin{pmatrix} 1 & \rho_n \\ \rho_n & 1 \end{pmatrix}$, where $\rho_n := \operatorname{corr}(\hat{X}_{\mathrm{inv}}, X_{\mathrm{des}})$. An application of Corollary 4.3.2 with $\mathcal{N}_n \sim \mathrm{N}(0, \Sigma^{(n)})$ yields that

$$\sup_{\mathbf{u}\in\mathbb{R}^2} |\mathbb{P}(X^{(n)} \le \mathbf{u}) - \mathbb{P}(\mathcal{N}_n \le \mathbf{u})| \le r_n(\mathcal{A}^{\mathrm{re}}) = O\left(n^{-1/6}\log(n)^{7/6}\right).$$

In combination with the fact that $\rho_n \longrightarrow 0$ (see Lemma 4.2.11), this establishes

$$(\hat{Y}_{\mathrm{inv}}, Y_{\mathrm{des}})^\top \xrightarrow{\mathcal{D}} \mathrm{N}_2(0, \mathrm{I}_2),$$

completing the proof of the theorem. □

This CLT can be straightforwardly extended to finite products of classical Weyl groups, since the applicability of the Hájek projection for such products has been clarified in Corollary 4.2.12.

Corollary 4.3.5. *Any sequence of finite products of classical Weyl groups satisfies the CLT for* $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$.

4.4 The extreme value asymptotics of $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$

In what follows, we use the Gaussian approximation of Theorem 4.3.1 to prove that $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$ is in the max-domain of attraction of the bivariate Gumbel distribution with independent marginals:

$$\Lambda_2(\mathbf{x}) = \exp\left(-e^{-x} - e^{-y}\right), \qquad \mathbf{x} = (x, y) \in \mathbb{R}^2.$$

For this, we will draw connections to the bivariate standard normal distribution, similar to the univariate case (cf. Theorem 3.3.2). The bivariate standard normal distribution is attracted by Λ_2 according to Theorem 1.3.10. Recall the notations $\boldsymbol{\alpha}_n, \boldsymbol{\beta}_n$ therefrom, as well as the notations of (4.5) and (4.6). Let

$$r_n(\mathcal{A}^{\mathrm{ext}}) := \sup_{\mathbf{u}\in\mathbb{R}^p} |\mathbb{P}(X^{(n)} \ge \mathbf{u}) - \mathbb{P}(\mathcal{N} \ge \mathbf{u})| \le r_n(\mathcal{A}^{\mathrm{re}}).$$

Theorem 4.3.1 gives an upper bound of $r_n(\mathcal{A}^{\mathrm{ext}})$. In our setting, it is only valid for the Hájek approximation $(\hat{Y}_{\mathrm{inv}}, Y_{\mathrm{des}})^\top$. It is not immediately clear that the same Gaussian approximation also applies to the original standardized statistic $(Y_{\mathrm{inv}}, Y_{\mathrm{des}})^\top$.

Following the conventions introduced in Sections 3.1, we write $X_{n1}, \ldots, X_{nk_n}$ for the n-th row of the triangular array, with a sequence $(k_n)_{n\in\mathbb{N}}$ of positive integers tending to infinity. Now, all $X_{n1}, \ldots, X_{nk_n}$ are i.i.d. samples of $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$ drawn on a classical Weyl group W_n of rank n. Moreover, "$*$" denotes component-wise multiplication.

The connection of $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$ to the bivariate standard normal distribution is drawn directly by replacing X_{inv} with $\hat{X}_{\mathrm{inv}}$ and using Slutsky's theorem. The following EVLT for $(X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$ is the second main result of this thesis.

Theorem 4.4.1. *Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of classical Weyl groups with* $\mathrm{rk}(W_n) = n$ $\forall n \in \mathbb{N}$. *Let $(X_{nj})_{j=1,\dots,k_n}$ be a row-wise i.i.d. triangular array with $X_{n1} \stackrel{\mathcal{D}}{=} (X_{\mathrm{inv}}, X_{\mathrm{des}})^\top$, and assume $k_n \log k_n = o(n)$. Let $M_n := \max\{X_{n1}, \dots, X_{nk_n}\}$ be the row-wise maximum. Let $\mu_n := \mathbb{E}(X_{n1})$ and let $s_n := (\sigma(X_{\mathrm{inv}}), \sigma(X_{\mathrm{des}}))$. Let $a_n := s_n * \boldsymbol{\alpha}_{k_n}$ and let $b_n := s_n * \boldsymbol{\beta}_{k_n} + \mu_n$. Then,*

$$\forall \mathbf{x} \in \mathbb{R}^2: \quad \mathbb{P}(M_n \le a_n * \mathbf{x} + b_n) \longrightarrow \Lambda_2(\mathbf{x})\,.$$

Proof. For each $j = 1, \dots, k_n$, write $X_{nj} = (X_{\mathrm{inv}}^{(j)}, X_{\mathrm{des}}^{(j)})^\top$. Further, let $\hat{X}_{\mathrm{inv}}^{(1)}, \dots, \hat{X}_{\mathrm{inv}}^{(k_n)}$ be i.i.d. copies of $\hat{X}_{\mathrm{inv}}$ on W_n. To simplify the notation, let

$$\begin{aligned} M_{n,\mathrm{inv}} &:= \frac{\max_{j=1,\dots,k_n} X_{\mathrm{inv}}^{(j)} - \mathbb{E}(X_{\mathrm{inv}})}{\sigma(X_{\mathrm{inv}})}\,, \\ \hat{M}_{n,\mathrm{inv}} &:= \frac{\max_{j=1,\dots,k_n} \hat{X}_{\mathrm{inv}}^{(j)} - \mathbb{E}(X_{\mathrm{inv}})}{\sigma(\hat{X}_{\mathrm{inv}})}\,, \\ M_{n,\mathrm{des}} &:= \frac{\max_{j=1,\dots,k_n} X_{\mathrm{des}}^{(j)} - \mathbb{E}(X_{\mathrm{des}})}{\sigma(X_{\mathrm{des}})}\,. \end{aligned}$$

Accordingly, we have to show that $\forall x, y \in \mathbb{R}$:

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_{n,\mathrm{inv}} - \beta_{k_n}}{\alpha_{k_n}} \le x, \frac{M_{n,\mathrm{des}} - \beta_{k_n}}{\alpha_{k_n}} \le y\right) = \Lambda(x)\Lambda(y)\,.$$

Note that $\alpha_{k_n}^{-1} \sim \sqrt{2\log k_n}$. Therefore, by Slutsky's theorem, the claim immediately follows from

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\hat{M}_{n,\mathrm{inv}} - \beta_{k_n}}{\alpha_{k_n}} \le x, \frac{M_{n,\mathrm{des}} - \beta_{k_n}}{\alpha_{k_n}} \le y\right) = \Lambda(x)\Lambda(y)\,, \qquad x, y \in \mathbb{R}\,, \tag{4.7}$$

and

$$\sqrt{\log k_n}\,|M_{n,\mathrm{inv}} - \hat{M}_{n,\mathrm{inv}}| \xrightarrow{\mathbb{P}} 0\,, \qquad n \to \infty\,. \tag{4.8}$$

We start with the proof of (4.8) and get

$$\begin{aligned} |M_{n,\mathrm{inv}} - \hat{M}_{n,\mathrm{inv}}| &\le \max_{j=1,\dots,k_n} \left| \frac{X_{\mathrm{inv}}^{(j)} - \mathbb{E}(X_{\mathrm{inv}})}{\sigma(X_{\mathrm{inv}})} - \frac{\hat{X}_{\mathrm{inv}}^{(j)} - \mathbb{E}(X_{\mathrm{inv}})}{\sigma(\hat{X}_{\mathrm{inv}})} \right| \\ &= \max_{j=1,\dots,k_n} \left| \frac{X_{\mathrm{inv}}^{(j)} - \hat{X}_{\mathrm{inv}}^{(j)}}{\sigma(X_{\mathrm{inv}})} + \left(\hat{X}_{\mathrm{inv}}^{(j)} - \mathbb{E}(X_{\mathrm{inv}})\right) \frac{\sigma(\hat{X}_{\mathrm{inv}}) - \sigma(X_{\mathrm{inv}})}{\sigma(X_{\mathrm{inv}})\sigma(\hat{X}_{\mathrm{inv}})} \right|\,. \end{aligned}$$

Thus, we obtain for $\varepsilon > 0$:

$$\begin{aligned} &\mathbb{P}\Big(\sqrt{\log k_n}\,|M_{n,\mathrm{inv}} - \hat{M}_{n,\mathrm{inv}}| > 2\varepsilon\Big) \le \mathbb{P}\Big(\sqrt{\log k_n} \max_{j=1,\dots,k_n} \left| \frac{X_{\mathrm{inv}}^{(j)} - \hat{X}_{\mathrm{inv}}^{(j)}}{\sigma(X_{\mathrm{inv}})} \right| > \varepsilon\Big) \\ &+ \mathbb{P}\Big(\sqrt{\log k_n} \max_{j=1,\dots,k_n} \left| \left(\hat{X}_{\mathrm{inv}}^{(j)} - \mathbb{E}(X_{\mathrm{inv}})\right) \frac{\sigma(\hat{X}_{\mathrm{inv}}) - \sigma(X_{\mathrm{inv}})}{\sigma(X_{\mathrm{inv}})\sigma(\hat{X}_{\mathrm{inv}})} \right| > \varepsilon\Big) =: P_1 + P_2\,. \end{aligned}$$

Using the union bound and Markov's inequality, we have

$$P_1 \le k_n\,\mathbb{P}\Big(|X_{\mathrm{inv}} - \hat{X}_{\mathrm{inv}}| > \frac{\sigma(X_{\mathrm{inv}})\varepsilon}{\sqrt{\log k_n}}\Big) \le k_n \frac{\log k_n}{\mathrm{Var}(X_{\mathrm{inv}})\varepsilon^2} \mathbb{E}|X_{\mathrm{inv}} - \hat{X}_{\mathrm{inv}}|^2$$

$$
\begin{aligned}
&= \frac{k_n \log k_n}{\mathrm{Var}(X_{\mathrm{inv}})\varepsilon^2}\Big(\mathrm{Var}(X_{\mathrm{inv}}) + \mathrm{Var}(\hat{X}_{\mathrm{inv}}) - 2\mathrm{Cov}(X_{\mathrm{inv}}, \hat{X}_{\mathrm{inv}})\Big) \\
&= \frac{k_n \log k_n}{\varepsilon^2}\Big(1 - \frac{\mathrm{Var}(\hat{X}_{\mathrm{inv}})}{\mathrm{Var}(X_{\mathrm{inv}})}\Big)\,, \qquad (4.9)
\end{aligned}
$$

where the last equation follows from the fact that $\mathrm{Cov}(X_{\mathrm{inv}}, \hat{X}_{\mathrm{inv}}) = \mathrm{Var}(\hat{X}_{\mathrm{inv}})$ (see the proof of [104, Theorem 11.1]). From Theorem 2.2.2 and Lemma 4.1.3, we conclude that

$$
\frac{\mathrm{Var}(\hat{X}_{\mathrm{inv}})}{\mathrm{Var}(X_{\mathrm{inv}})} = 1 + \Theta(1/n)\,,
$$

from which it follows that

$$
P_1 = k_n \log k_n O(1/n)\,, \qquad n \to \infty\,,
$$

which tends to zero by the assumption on k_n. Repeating the above considerations for P_2 and noting that

$$
\begin{aligned}
\left(\frac{\sigma(\hat{X}_{\mathrm{inv}}) - \sigma(X_{\mathrm{inv}})}{\sigma(X_{\mathrm{inv}})}\right)^2 &= \frac{\mathrm{Var}(X_{\mathrm{inv}}) + \mathrm{Var}(\hat{X}_{\mathrm{inv}}) - 2\sigma(X_{\mathrm{inv}})\sigma(\hat{X}_{\mathrm{inv}})}{\mathrm{Var}(X_{\mathrm{inv}})} \\
&\sim \frac{\mathrm{Var}(X_{\mathrm{inv}}) + \mathrm{Var}(\hat{X}_{\mathrm{inv}}) - 2\mathrm{Var}(\hat{X}_{\mathrm{inv}})}{\mathrm{Var}(X_{\mathrm{inv}})} = 1 - \frac{\mathrm{Var}(\hat{X}_{\mathrm{inv}})}{\mathrm{Var}(X_{\mathrm{inv}})}
\end{aligned}
$$

yields

$$
\begin{aligned}
P_2 &\le \frac{k_n \log k_n}{\varepsilon^2}\left(\frac{\sigma(\hat{X}_{\mathrm{inv}}) - \sigma(X_{\mathrm{inv}})}{\sigma(X_{\mathrm{inv}})}\right)^2 \underbrace{\mathbb{E}\left(\frac{\hat{X}^{(j)}_{\mathrm{inv}} - \mathbb{E}(X_{\mathrm{inv}})}{\sigma(\hat{X}_{\mathrm{inv}})}\right)^2}_{=1} \\
&= k_n \log k_n O(1/n) = o(1)\,, \qquad n \to \infty\,, \qquad (4.10)
\end{aligned}
$$

which completes the proof of (4.8). It remains to prove (4.7). By analogy with the proof of Theorem 4.3.4, Corollary 4.2.10 allows us to find centered 1-dependent random vectors $X_k^{(n,j)}$, $k = 1, \dots, n-1$, such that

$$
\begin{pmatrix} \hat{Y}^{(j)}_{\mathrm{inv}} \\ Y^{(j)}_{\mathrm{des}} \end{pmatrix} = \sum_{k=1}^{n} X_k^{(n,j)}.
$$

Again, the covariance matrix of $(\hat{Y}^{(j)}_{\mathrm{inv}}, Y^{(j)}_{\mathrm{des}})^\top$ is given by $\Sigma^{(n)} = \begin{pmatrix} 1 & \rho_n \\ \rho_n & 1 \end{pmatrix}$, where $\rho_n := \mathrm{corr}(\hat{X}_{\mathrm{inv}}, X_{\mathrm{des}})$. For a centered normal random vector $\mathcal{N}_n = (N_1, \dots, N_{2k_n})^\top$ whose covariance matrix is block-diagonal with all k_n diagonal blocks equal to $\Sigma^{(n)}$, we write

$$
P_n(x,y) := \mathbb{P}\left(\alpha_{k_n}^{-1}\Big(\max_{j=1,\dots,k_n} N_{2j-1} - \beta_{k_n}\Big) \le x, \alpha_{k_n}^{-1}\Big(\max_{j=1,\dots,k_n} N_{2j} - \beta_{k_n}\Big) \le y\right), \qquad x, y \in \mathbb{R}.
$$

We can also write

$$
\begin{aligned}
&\mathbb{P}(\alpha_n(\hat{M}_n - \alpha_n) \le x, \alpha_n(M_{n,\mathrm{des}} - \alpha_n) \le y) \\
&= \mathbb{P}\left(\alpha_n(\hat{Y}^{(1)}_{\mathrm{inv}}, \dots, \hat{Y}^{(k_n)}_{\mathrm{inv}})^\top - \boldsymbol{\alpha}_n \le \mathbf{x},\ \alpha_n(Y^{(1)}_{\mathrm{des}}, \dots, Y^{(k_n)}_{\mathrm{des}})^\top - \boldsymbol{\alpha}_n \le \mathbf{y}\right),
\end{aligned}
$$

with $\mathbf{x} = (x, \ldots, x)^\top$, $\mathbf{y} = (y, \ldots, y)^\top$, $\boldsymbol{\alpha}_n = (\alpha_n, \ldots, \alpha_n)^\top \in \mathbb{R}^{k_n}$. An application of Corollary 4.3.2 then yields, as $n \to \infty$,

$$\begin{aligned}&\left|\mathbb{P}\left(\frac{\hat{M}_{n,\text{inv}} - \beta_{k_n}}{\alpha_{k_n}} \le x, \frac{M_{n,\text{des}} - \beta_{k_n}}{\alpha_{k_n}} \le y\right) - P_n(x, y)\right| \\ &= O\left(n^{-1/6} \log(k_n)^{7/6}\right) = o(1)\,.\end{aligned}$$

Finally, since $\rho_n \to 0$ (see Corollary 4.2.4 and Lemma 4.2.11), the extreme value behavior of bivariate Gaussian random vectors expressed in Theorem 1.3.10 gives

$$P_n(x, y) \overset{n\to\infty}{\longrightarrow} \Lambda(x)\Lambda(y)\,,$$

completing the proof of (4.7). Combining (4.7) and (4.8) gives the claim. □

Remark 4.4.2. Due to the Hájek approximation error, the upper bound on the row-wise number of samples k_n in Theorem 4.4.1 is a lot stricter than in the univariate case (cf. Theorem 3.3.2). In particular, this excludes the uniform triangular array. On the other hand, this new EVLT can be transferred to other individual and joint permutation statistics, and for some, we can achieve almost the same regime of k_n as in Theorem 3.3.2.

In Section 3.5, we explained that the methods of Theorem 3.3.2 are restrictive for other permutation statistics (e.g., if there is no available factorization of the generating function). In this process, we also introduced a universal EVLT 3.5.1 for any asymptotically normal permutation statistic satisfying the Berry–Esseen bound. However, the upper bound k_n in Theorem 3.5.1 still excludes the uniform triangular array and falls short compared to the subexponential bound in Theorem 3.3.2. In the following framework, we describe the requirements to prove the EVLT for a univariate or joint permutation statistic with the methods of Theorem 4.4.1, in order to bypass these restrictions.

Let $\mathcal{W}$ be the system of classical Weyl groups or a subsystem thereof (e.g., the family of symmetric groups), and let $(X_n)_{n\in\mathbb{N}}$ be a permutation statistic on $\mathcal{W}$ in one or two dimensions. Formally, $(X_n)_{n\in\mathbb{N}}$ is a collection of random variables $X_n\colon W_n \to \mathbb{N}^d$ with $W_n \in \mathcal{W}$, $\mathrm{rk}(W) = n$, and with $d \in \{1, 2\}$ fixed. Moreover, we assume that there is a representation

$$X_n = \sum_{i=1}^{n} f_i(Z_1, \ldots, Z_n) =: \sum_{i=1}^{n} X_n^{(i)} \tag{4.11}$$

for some independent sequence $Z_1, Z_2, \ldots$ of random variables and functions $f_n\colon \mathbb{R}^n \to \mathbb{N}^d$, such that the following is satisfied:

- If $d = 1$, then we assume that X_n is m-dependent (i.e., that all blocks $X_n^{(1)}, \ldots, X_n^{(n)}$ are m-dependent) for some $m \in \mathbb{N}$ chosen independently of n. Besides X_{des}, an example of such a permutation statistic is the number of peaks or valleys.
- If $d = 2$, then one component must be m-dependent. The other component must be m-dependent as well, or satisfy the condition of Theorem 4.1.2, i.e., $\mathrm{Var}(X_n) \sim \mathrm{Var}(\hat{X}_n)$, where $\hat{X}_n$ is the Hájek projection of X_n based on (4.11). In the latter case, the Hájek approximation error needs to be controlled by establishing proper bounds in (4.9) and (4.10). In light of Theorem 1.3.10, it is required that the correlation between the two components is bounded away from 1, but this is commonly trivial to verify.

Theorem 4.4.3. *Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of classical Weyl groups with* $\mathrm{rk}(W_n) = n$ $\forall n \in \mathbb{N}$. *Let X_n be a permutation statistic as described above. Let $(X_{nj})_{j=1,\dots,k_n}$ be a row-wise i.i.d. triangular array with $X_{n1} \overset{\mathcal{D}}{=} X_n$, where:*

(a) *If X_n is m-dependent, we assume $k_n = \exp\big(o(n^{1/7})\big)$.*

(b) *If X_n consists of two components, one of which is m-dependent and the other is not, but satisfies the condition of Theorem 4.1.2, then we assume $k_n \log(k_n) = o(n)$.*

Let $M_n := \max\{X_{n1}, \dots, X_{nk_n}\}$ be the row-wise maximum. Let $\mu_n := \mathbb{E}(X_n)$ and $s_n := \sigma(X_n)$, which is taken component-wise in case of $d = 2$. If $d = 1$, let $a_n := s_n \alpha_{k_n}$ and $b_n := s_n \beta_{k_n} + \mu_n$. Then,

$$\forall x \in \mathbb{R}: \quad \mathbb{P}(M_n \le a_n x + b_n) \longrightarrow \Lambda(x)\,.$$

If $d = 2$, let $a_n := s_n * \boldsymbol{\alpha}_{k_n}$ and $b_n := s_n * \boldsymbol{\beta}_{k_n} + \mu_n$. Then,

$$\forall \mathbf{x} \in \mathbb{R}^d: \quad \mathbb{P}(M_n \le a_n * \mathbf{x} + b_n) \longrightarrow \Lambda_2(\mathbf{x})\,.$$

Proof. In case of (b), the proof is identical to that of Theorem 4.4.1. In case of (a), we only need to show (4.7), while we replace $(\hat{M}_{n,\mathrm{inv}}, M_{n,\mathrm{des}})$ with the standardized maximum of X_n. We can apply Theorem 4.3.1 with $\mathfrak{p}(n) = n \vee k_n$ iterations of X_n. Therefore, we need to ensure that

$$n^{-1/6} \log(n \vee k_n)^{7/6} = o(1)\,,$$

which exactly corresponds to the stated condition of $k_n = \exp\big(o(n^{1/7})\big)$. The claim follows the same way as in the proof of Theorem 4.4.1. $\square$

In conclusion, for independent and m-dependent permutation statistics, the high-dimensional Gaussian approximation allows to obtain a subexponential bound on k_n, improving Theorem 3.5.1. In particular, this applies to the following permutation statistics:

Corollary 4.4.4. *Let $Z_1, \dots, Z_n \sim U(0,1)$ be i.i.d. The following three permutation statistics are in the MDA of the Gumbel distribution, given a triangular array with row lengths satisfying $k_n = \exp\big(o(n^{1/7})\big)$:*

- *the number of peaks $X_p := \sum_{i=2}^{n-1} \mathbf{1}\{Z_i > Z_{i-1}, Z_{i+1}\}$,*
- *the number of valleys $X_v := \sum_{i=2}^{n-1} \mathbf{1}\{Z_{i-1}, Z_{i+1} > Z_i\}$,*
- *the number of cycles K_{0n} introduced in Section 3.5, since it has the independent decomposition given in (3.9).*

Regarding products of classical Weyl groups, we have established the CLT for products with a bounded number of components. To obtain the EVLT, we additionally have to control the bounds (4.9) and (4.10). It turns out that these bounds do not impose any further restrictions for k_n.

Theorem 4.4.5. *For fixed $l \in \mathbb{N}$, let $W_n = \prod_{i=1}^{l} W_{n,i}$ be products of finite Coxeter groups with ranks $n_1 \ge \dots \ge n_l$ sorted in decreasing order. Then, the statement of Theorem 4.4.1 applies to $(W_n)_{n\in\mathbb{N}}$.*

Proof. The proof of Theorem 4.4.1 carries over almost seamlessly, we only need to check (4.9) and (4.10), which reduces to bounding

$$1 - \frac{\operatorname{Var}(\hat{X}_{\text{inv}})}{\operatorname{Var}(X_{\text{inv}})}\,. \tag{4.12}$$

We can rephrase (4.4) as

$$\operatorname{Var}(\hat{X}_{\text{inv}}^{W_n}) = \sum_{i=1}^{l} c_i n_i^3 + \alpha n^2 + O(n)\,, \qquad \operatorname{Var}(X_{\text{inv}}^{W_n}) = \sum_{i=1}^{l} c_i n_i^3 + \beta n^2 + O(n)\,.$$

Then,

$$(4.12) = 1 - \frac{\sum_{i=1}^{l} c_i n_i^3 + \alpha n^2 + O(n)}{\sum_{i=1}^{l} c_i n_i^3 + \beta n^2 + O(n)}\,.$$

Regardless of whether the residual $\beta n^2 + O(n)$ is positive or negative, we can bound (4.12) in both directions. If the residual is positive, this reads

$$\begin{aligned}
(4.12) &\geq 1 - \frac{\sum_{i=1}^{l} c_i n_i^3 + \alpha n^2 + O(n)}{\sum_{i=1}^{l} c_i n_i^3} = \frac{\alpha n^2 + O(n)}{\sum_{i=1}^{l} c_i n_i^3} = O\left(\frac{1}{n}\right),\\
(4.12) &= \frac{(\beta-\alpha) n^2 + O(n)}{\sum_{i=1}^{l} c_i n_i^3 + \beta n^2 + O(n)} \leq \frac{(\beta-\alpha) n^2 + O(n)}{\sum_{i=1}^{l} c_i n_i^3} = O\left(\frac{1}{n}\right).
\end{aligned}$$

Therefore, we obtain the same bound for (4.12) as in the proof of Theorem 4.4.1. □

4.5 Summary and open problems

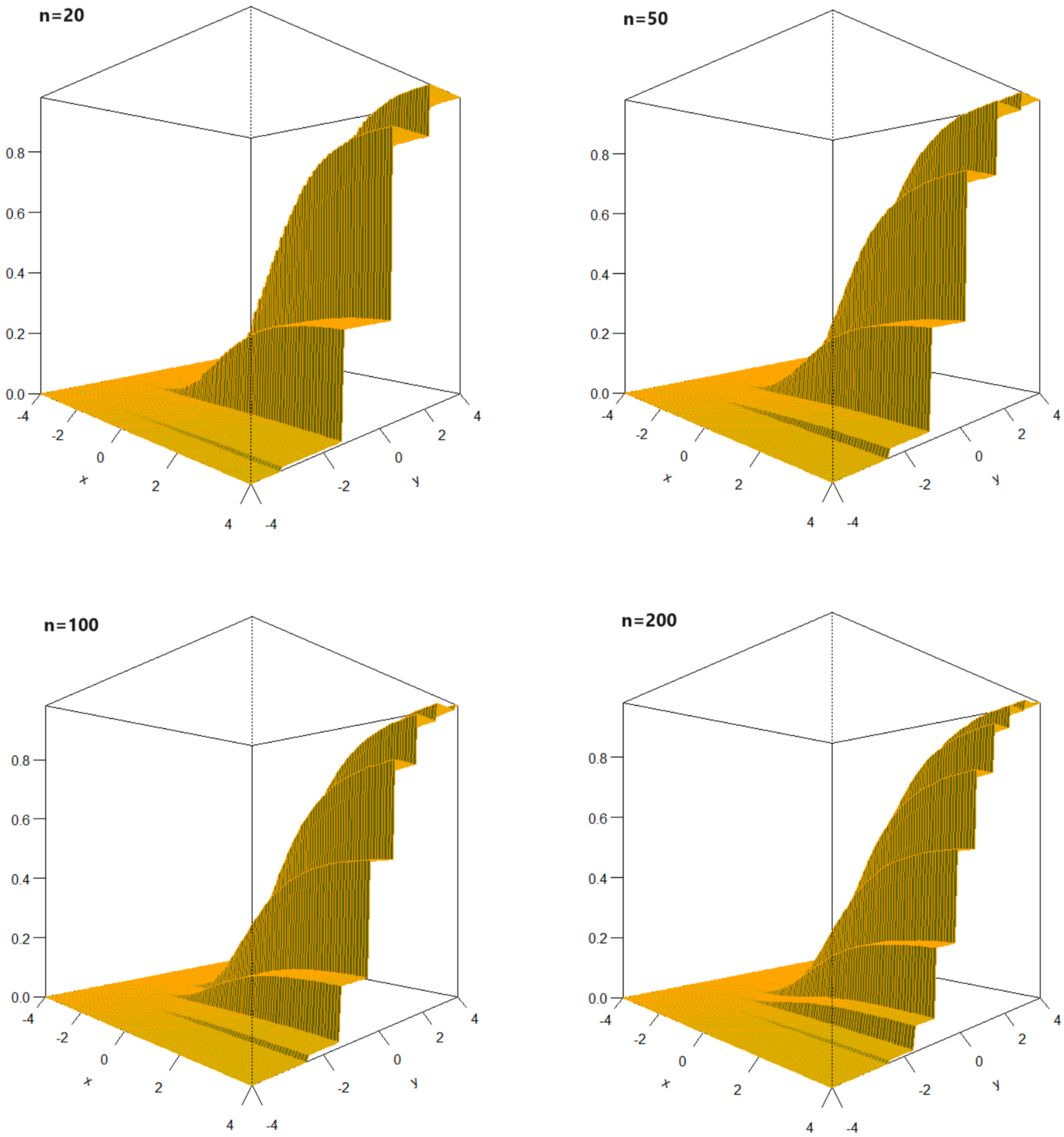
The main result of this chapter is the extreme value behavior of the joint distribution $(X_{\text{inv}}, X_{\text{des}})^\top$, which is stated in Theorem 4.4.1. Besides, the asymptotic normality of $(X_{\text{inv}}, X_{\text{des}})^\top$ on classical Weyl groups shown in Theorem 4.3.4 gives a significant extension of [45]. We benefited from the fact that the number of inversions X_{inv} can be suitably approximated by its Hájek projection, allowing to apply Gaussian approximation theory for m-dependent random vectors. There are several open questions related to the sharpness of the upper bound on k_n, and also to other permutation statistics with more complex dependency structures.

4.5.1 Extension of the upper bound on k_n

Due to replacing X_{inv} with its Hájek projection, Theorem 4.4.1 requires an overly strict upper bound on the row lengths k_n. In particular, the uniform triangular array $(X_{nj})_{j=1,\ldots,n}$ is not covered by Theorem 4.4.1. However, we suppose that the upper bound $k_n \log(k_n) = o(n)$ is far from exhaustive. In fact, we conjecture that the dependency structure between X_{inv} and X_{des} still allows for an exponential bound on k_n as seen in Theorems 3.3.2 and 4.4.3.

Conjecture 4.5.1. *There is a constant $\gamma > 0$ such that the statement of Theorem 4.4.1 holds true for a triangular array $(X_{nj})_{j=1,\ldots,k_n}$ with $X_{n1} \overset{\mathcal{D}}{=} (X_{\text{inv}}, X_{\text{des}})^\top$ and $k_n = \exp\left(o(n^\gamma)\right)$.*

In particular, regarding the uniform triangular array $(X_{nj})_{j=1,\ldots,n}$ with $X_{n1} \stackrel{\mathcal{D}}{=} (X_{\text{inv}}, X_{\text{des}})^\top$ on the symmetric group S_n, the weak convergence of $(M_n - b_n)/a_n$ to Λ_2 with M_n, a_n, b_n, Λ_2 as in Theorem 4.4.1 was observed in simulations performed with RStudio 2023.06.1 [89]. The code details are sketched in Section 4.6.7. For the underlying symmetric groups, we chose the sizes $n \in \{20, 50, 100, 200, 500, 1000\}$, and created 10000 independent replications of $(M_n - b_n)/a_n$ in order to plot their empirical distribution function. These plots are shown in Figure 4.4, with the x-axis referring to the inversion component and the y-axis referring to the descent component. In comparison, the two-dimensional Gumbel distribution Λ_2 is displayed in Figure 1.5. Typically, ELVTs exhibit slow rates of convergence, e.g., in [71, Section 2.4], it is argued that the convergence in Theorem 1.1.11 cannot be faster than $\Theta(\log(n)^{-1})$ even if α_n, β_n are replaced. Therefore, it is not surprising that there is only little similarity to Λ_2 for small symmetric groups. However, Figure 4.4 illustrates that the empirical distribution function indeed approaches Λ_2 as the underlying symmetric groups become large.



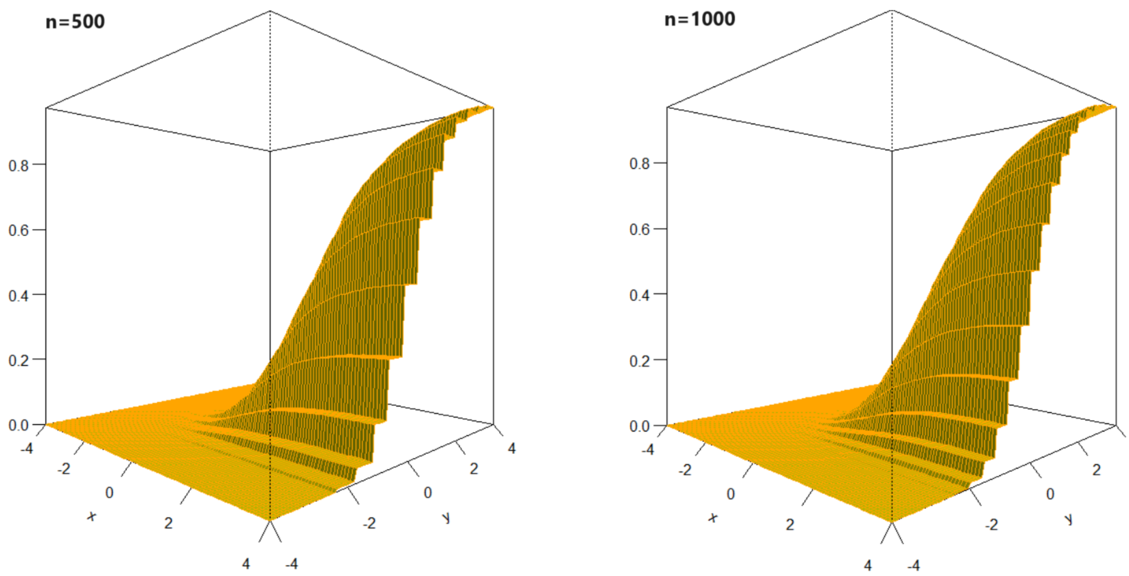


Figure 4.4: Plots of empirical distribution functions of $(M_n - b_n)/a_n$ for the symmetric groups S_n, with M_n, a_n, b_n as given in Theorem 4.4.1 and $n \in \{20, 50, 100, 200, 500, 1000\}$. These empirical distribution functions converge to Λ_2 as $n \to \infty$.

This simulation demonstrates the potential to improve the upper bound on k_n in Theorem 4.4.1. For the extension to other m-dependent statistics given in Theorem 4.4.3, the bound $k_n = \exp\left(o(n^{1/7})\right)$ results from the rate of convergence in Theorem 4.3.1. The error rate of this Gaussian approximation is $n^{-1/6}\log(\mathfrak{p})^{7/6}$, and since the dimension $\mathfrak{p}$ also affects the error rate, we can choose to replace $\mathfrak{p}$ with k_n to yield an exponential bound on k_n. Research on Gaussian approximation for high-dimensional m-dependent random vectors to improve the error bound in Theorem 4.3.1 has evolved even further. Most recently, Bong *et al.* [11] proved approximation theorems in the way of Theorem 4.3.1 with a main error term of $n^{-1/2}\log(n\mathfrak{p})^{1/2}$, and a logarithmic error term only depending on $\mathfrak{p}$. There are several additional parts in this logarithmic term that can be treated as constants in our setting. Despite the authors saying that the conditions of these approximation theorems essentially capture non-degeneracy and existence of third moments, it is very laborious to verify these conditions by hand. If this can be done for the statistics listed in Corollary 4.4.4, then by [11, Theorem 3.2], the upper bound on k_n can be improved to $k_n = n^{-1}\exp\left(o(n^{1/4})\right)$.

Conjecture 4.5.2. *The statement of Theorem* 4.4.3(a) *holds true if the bound* $k_n = n^{-1}\exp\left(o(n^{1/4})\right)$ *is assumed.*

4.5.2 Weakly dependent permutation statistics

In Definition 2.3.4, we introduced the two-sided Eulerian statistic $X_T = X_{\text{des}} + X_{\text{ides}}$. Its asymptotic normality on finite Coxeter groups was shown in [15, 48]. Moreover, it was first shown by Vatutin [105] that on the symmetric groups S_n, X_{des} and X_{ides} are asymptotically uncorrelated, i.e., $\rho_n(X_{\text{des}}, X_{\text{ides}}) \longrightarrow 0$. Vatutin proved the asymptotic normality of $(X_{\text{des}}, X_{\text{ides}})^\top$ through extensive analytic arguments on characteristic functions, without providing probabilistic insight into their dependency structure. Later, it was shown in [20, p. 8] and [66, Propositions 5.6–5.8] that $\rho_n(X_{\text{des}}, X_{\text{ides}}) = O(1/n)$ for classical Weyl groups. [66, Theorem 5.1] implies that $\rho_n(X_{\mathrm{des}}, X_{\mathrm{ides}}) \longrightarrow 0$ for all finite Coxeter groups.

Theorem 4.3.1 cannot be applied to $X_T = X_{\mathrm{des}} + X_{\mathrm{ides}}$ since no m-dependent decomposition is available, and the dependency graph is too dense as well. The dependency structure also does not satisfy other conditions for which Gaussian approximation theorems exist. For example, [17, Theorem 1] provides a Gaussian approximation similar to Theorem 4.3.1 for the framework of strong mixing dependence (cf. Definition 1.2.1), also called *α-mixing dependence*. For triangular arrays, the property of uniform α-mixing dependence is defined as follows:

Definition 4.5.3. Let $(X_{nj})_{j=1,\dots,k_n}$ be a triangular array. For each $n \in \mathbb{N}$ and each $j = 1, \dots, k_n$, let $\mathcal{F}^-_{nj}$ be the σ-field generated by $X_{n1}, \dots, X_{nj}$ and let $\mathcal{F}^+_{nj}$ be the σ-field generated by $X_{nj}, X_{n,j+1}, \dots, X_{nk_n}$. For $k \in \{1, \dots, k_n - 1\}$, the n-th *α-mixing coefficient at lag k* is

$$\alpha_n(k) := \sup_{j=1,\dots,k_n-k} \sup_{A \in \mathcal{F}^-_{nj}, B \in \mathcal{F}^+_{n,j+k}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|\,,$$

and for $k \geq k_n$, we put $\alpha_n(k) = 0$. The overall k-th α-mixing coefficient is

$$\alpha_k := \sup_{n \in \mathbb{N}} \alpha_n(k).$$

We say the triangular array $(X_{nj})_{j=1,\dots,k_n}$ is *α-mixing* if $\alpha(k) \longrightarrow 0$ as $k \to \infty$. △

Remark 4.5.4. Note that for a triangular array consisting of indicator variables $X_{ni} = \mathbf{1}\{A_{ni}\}$ with a collection of events $(A_{ni})_{i=1,\dots,k_n}$, the α-mixing coefficients are equivalently expressed as

$$\alpha_{n,k} := \sup_{j=1,\dots,k_n-k} \sup_{\substack{\mathcal{A} \subseteq \{1,\dots,j\} \\ \mathcal{B} \subseteq \{j+k,\dots,k_n\}}} \left| \mathbb{P}\left(\bigcap_{i \in \mathcal{A} \cup \mathcal{B}} A^{\mathrm{c}}_{ni} \right) - \mathbb{P}\left(\bigcap_{i \in \mathcal{A}} A^{\mathrm{c}}_{ni} \right) \mathbb{P}\left(\bigcap_{i \in \mathcal{B}} A^{\mathrm{c}}_{ni} \right) \right|,$$

meaning in this case, we can focus on the dependence between any two blocks $(A_{ni})_{i \in \mathcal{A}}$, $(A_{nj})_{j \in \mathcal{B}}$ with distance at least k. △

We now check if the two-sided Eulerian statistic X_T forms an α-mixing triangular array on the family of symmetric groups (the arguments are similar for other classical Weyl groups). Let $k_n = 2(n-1)$ and write

$$X_T^{(n)} = \sum_{i=1}^{n-1} Y_i + \sum_{j=1}^{n-1} \widetilde{Y}_j\,,$$

where $Y_i := \mathbf{1}\{\pi(i) > \pi(i+1)\}$ and $\widetilde{Y}_j := \mathbf{1}\{\pi^{-1}(j) > \pi^{-1}(j+1)\}$ for uniformly random $\pi \in S_n$, in accordance with the notation in [66, Section 5]. The blocks $\{Y_1, \dots, Y_{n-1}\}$ and $\{\widetilde{Y}_1, \dots, \widetilde{Y}_n\}$ each are 1-dependent and therefore do not cause any issues by themselves. However, regarding the dependence between these two blocks, the α-mixing coefficients are not influenced by the lag k but rather by the size n of the underlying symmetric group.

Consider two blocks $\mathcal{A}, \mathcal{B}$ according to Remark 4.5.4. In the simplest case of $\mathcal{A}, \mathcal{B}$ being singletons, we consider the events $\mathbf{1}\{Y_i = 1\}$ and $\mathbf{1}\{\widetilde{Y}_j = 1\}$ for some $i, j \in \{1, \dots, n-1\}$. To determine their dependence, we need to distinguish by the intersection of $\{\pi(i), \pi(i+1)\}$

and $\{j, j+1\}$. If these two sets are disjoint, then $\mathbf{1}\{Y_i = 1\}$ and $\mathbf{1}\{\widetilde{Y}_j = 1\}$ are independent. By a simple counting argument, this is the case for $(n-2)(n-3)(n-2)!$ permutations. So, there are $n! - (n-2)(n-3)(n-2)! = (4n-6)(n-2)! = \Theta((n-1)!)$ remaining permutations for which $\mathbf{1}\{Y_i = 1\}$ and $\mathbf{1}\{\widetilde{Y}_j = 1\}$ are dependent. For the details on computing $\mathbb{P}(Y_i = 1, \widetilde{Y}_j = 1)$ in these subcases, we refer to [66, p. 450]. It turns out that a lower bound for any mixing coefficient $\alpha_n(k)$ is given by $\Theta(1/n)$.

These arguments still hold for blocks $\mathcal{A}, \mathcal{B}$ of fixed size. If the size of the blocks is unbounded as $n \to \infty$, then the probabilities of the respective intersections diminish accordingly. While these arguments are technically more complicated, it can be supposed that the triangular array of $(X_T^{(n)})_{n\in\mathbb{N}}$ is α-mixing with a decay rate of $\alpha(k) = \Theta(k^{-1})$. However, the conditions of [17, Theorem 1] require that the mixing coefficients $\alpha(k)$ decay at an *exponential* rate, i.e., $\alpha(k) = O(e^{-\gamma})$ for some $\gamma > 0$. While this may be regarded as a mild condition for high-dimensional time series and other applications, it is unfortunately too strict for dependency structures in permutation statistics.

By analogy with the simulation of Section 4.5.1, we created 10000 independent replications of $(M_n - b_n)/a_n$ on the symmetric group S_n with $n = 1000$, where M_n is the maximum of 1000 samples of X_T. Figure 4.5 shows the empirical distribution functions of these replications in comparison with the univariate Gumbel distribution, again suggesting convergence as $n \to \infty$.

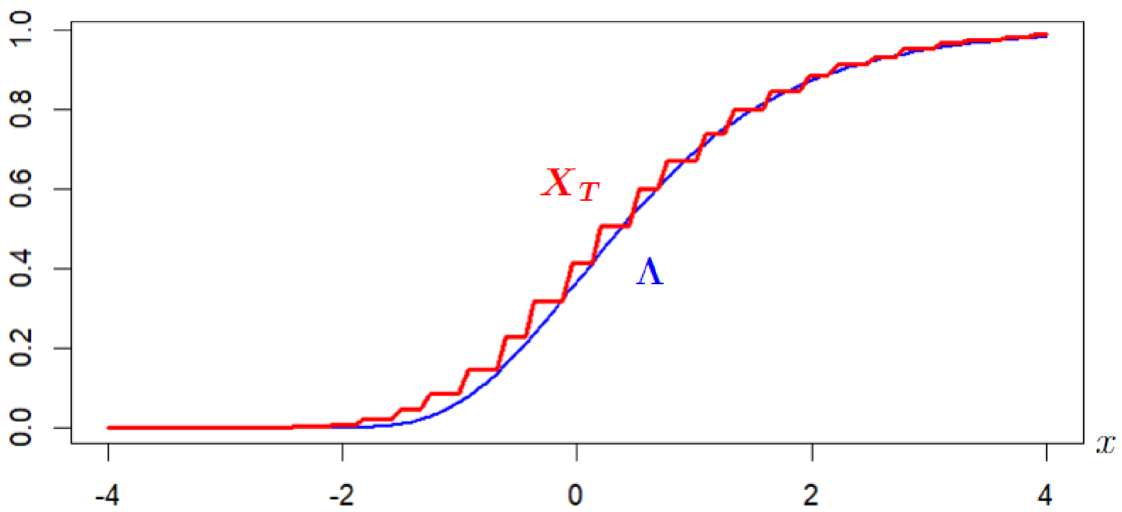


Figure 4.5: Plots of the standard Gumbel distribution function Λ and the simulated empirical distribution function of $\mathbb{P}(M_n \leq a_n x + b_n)$ for $x \in [-4, 4]$ and M_n, a_n, b_n stemming from X_T, as given in Conjecture 4.5.5.

So, if it is possible to find a high-dimensional Gaussian approximation of α-mixing triangular arrays with an $O(k^{-1})$ decay of $(\alpha(k))_{k\in\mathbb{N}}$, then we propose the following:

Conjecture 4.5.5. *Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of classical Weyl groups with* $\mathrm{rk}(W_n) = n$ $\forall n \in \mathbb{N}$, *and let $(X_{nj})_{j=1,\dots,k_n}$ be a row-wise i.i.d. triangular array with $X_{n1} \overset{\mathcal{D}}{=} X_T$ on W_n. Let M_n, a_n, b_n be as in Theorem 4.4.3. If $k_n = \exp(o(n^\gamma))$ for some constant $\gamma > 0$, then*

$$\forall x \in \mathbb{R} \colon \mathbb{P}(M_n \leq a_n x + b_n) \longrightarrow \exp(-\exp(-x)).$$

A corresponding statement in two dimensions is true for the triangular array $(X_{nj})_{j=1,\dots,k_n}$ *with* $X_{n1} \overset{\mathcal{D}}{=} (X_{\mathrm{des}}, X_{\mathrm{ides}})$ *on* W_n*.*

4.6 Technical proofs and simulation code

Throughout the following proofs, the symbol $\sum_{i<j}$ is a shorthand notation for $\sum_{1 \le i < j \le n}$.

4.6.1 Proof of Lemma 4.2.1

We first prove the claim for the even-signed permutation groups D_n and proceed as in Lemma 4.1.3. Recall that

$$\hat{X}^D_{\mathrm{inv}} = \sum_{k=1}^{n} \mathbb{E}(X^D_{\mathrm{inv}} \mid Z_k) - (n-1)\mathbb{E}(X^D_{\mathrm{inv}}),$$

with $Z_k \sim U(-1,1)$ and X^D_{inv} defined by (4.2b), giving

$$\begin{aligned} \mathbb{E}(X^D_{\mathrm{inv}} \mid Z_k) &= \sum_{i<j} \mathbb{E}\big(\mathbf{1}\{Z_i > Z_j\} + \mathbf{1}\{-Z_i > Z_j\} \mid Z_k\big) \\ &= \sum_{i<j} \mathbb{P}(Z_i > Z_j \mid Z_k) + \mathbb{P}(-Z_i > Z_j \mid Z_k). \end{aligned} \tag{4.13}$$

Write $U_i := |Z_i|$ and $f(Z_i, Z_j) := \mathbf{1}\{Z_i > Z_j\} + \mathbf{1}\{-Z_i > Z_j\}$ for $i < j$. A straightforward case distinction gives:

- $Z_j > 0 \Longrightarrow f(Z_i, Z_j) = \mathbf{1}\{U_i > U_j\}$,
- $Z_j < 0 \Longrightarrow f(Z_i, Z_j) = \mathbf{1}\{U_i < U_j\} + 1$,

as $f(Z_i, Z_j) = f(|Z_i|, Z_j)$ does not depend on the sign of Z_i. To compute (4.13), we only need to consider the $n-k$ tuples (k,j) and the $k-1$ tuples (i,k), since the remaining tuples are independent of Z_k and produce constants that do not contribute to the variance.

Recall that for $k < j$, we have $\mathbb{P}(U_k > U_j \mid U_k) = U_k$ and $\mathbb{P}(U_k < U_j \mid U_k) = 1 - U_k$. Therefore, we write

$$\mathbb{E}(X^D_{\mathrm{inv}} \mid Z_k) = \sum_{i=1}^{k-1} \mathbb{E}\big(f(Z_i, Z_k) \mid Z_k\big) + \sum_{j=k+1}^{n} \mathbb{E}\big(f(Z_k, Z_j) \mid Z_k\big) + \mathrm{const},$$

where

$$\begin{aligned} \mathbb{E}\big(f(Z_i, Z_k) \mid Z_k\big) &= \mathbf{1}\{Z_k > 0\}(1 - U_k) + \mathbf{1}\{Z_k < 0\}(1 + U_k), \\ \mathbb{E}\big(f(Z_k, Z_j) \mid Z_k\big) &= \mathbb{P}(Z_j < 0)U_k + \mathbb{P}(Z_j > 0)(1 + 1 - U_k). \end{aligned}$$

Overall,

$$\begin{aligned} \mathbb{E}(X^D_{\mathrm{inv}} \mid Z_k) = (k-1)\big(&\mathbf{1}\{Z_k > 0\}(1 - U_k) + \mathbf{1}\{Z_k < 0\}(1 + U_k) \\ &+ (n-k)\underbrace{\left(\frac{1}{2}U_k + \frac{1}{2}(2 - U_k)\right)}_{=\,1} + \mathrm{const} \end{aligned}$$

$$= (k-1)\mathbf{1}\{Z_k > 0\}(1-U_k) + (k-1)\mathbf{1}\{Z_k < 0\}(1+U_k) + \text{const}\,. \quad (4.14)$$

To compute $\mathrm{Var}(\hat{X}^D_{\mathrm{inv}}) = \sum_{k=1}^n \mathrm{Var}(\mathbb{E}(X^D_{\mathrm{inv}} \mid Z_k))$, we focus on the non-constant parts in (4.14). To use the standard formula $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, where X is not affected by constant summands, we first compute

$$\begin{aligned}
&\mathbb{E}\left(\left((k-1)\mathbf{1}\{Z_k > 0\}(1-U_k) + (k-1)\mathbf{1}\{Z_k < 0\}(1+U_k)\right)^2\right) \\
&= \mathbb{E}((k-1)^2(1-U_k)^2\mathbf{1}\{Z_k > 0\} + (k-1)^2(1+U_k)^2\mathbf{1}\{Z_k < 0\}) \\
&\quad + \underbrace{2\mathbb{E}((k-1)(1-U_k)\mathbf{1}\{Z_k > 0\}(k-1)(1+U_k)\mathbf{1}\{Z_k < 0\})}_{=\,0}\,,
\end{aligned}$$

because both $Z_k > 0$ and $Z_k < 0$ cannot occur simultaneously. Next, we obtain

$$\begin{aligned}
&\mathbb{E}((k-1)^2(1-U_k)^2\mathbf{1}\{Z_k > 0\} + (k-1)^2(1+U_k)^2\mathbf{1}\{Z_k < 0\}) \\
&= \frac{1}{2}(k-1)^2 \underbrace{\mathbb{E}((1-U_k)^2)}_{=\,1/3} + \frac{1}{2}(k-1)^2 \underbrace{\mathbb{E}((1+U_k)^2)}_{=\,7/3} \\
&= \frac{4}{3}(k-1)^2.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
&\Big[\mathbb{E}((k-1)(1-U_k)\mathbf{1}\{Z_k > 0\} + (k-1)(1+U_k)\mathbf{1}\{Z_k < 0\})\Big]^2 \\
&= \left(\frac{1}{4}(k-1) + \frac{3}{4}(k-1)\right)^2 = (k-1)^2,
\end{aligned}$$

and therefore,

$$\mathrm{Var}(\mathbb{E}(X^D_{\mathrm{inv}} \mid Z_k)) = \frac{1}{3}(k-1)^2.$$

Summation gives

$$\mathrm{Var}(\hat{X}^D_{\mathrm{inv}}) = \sum_{k=1}^n \frac{1}{3}(k-1)^2 = \sum_{k=0}^{n-1} \frac{1}{3}k^2 = \frac{1}{3}\frac{n(n-1)(2n-1)}{6} = \frac{1}{9}n^3 + O(n^2)\,,$$

as desired. So, we have computed $\mathrm{Var}(\hat{X}^D_{\mathrm{inv}})$ on D_n. On B_n, the calculation is similar. Recall that $X^B_{\mathrm{inv}} = X^D_{\mathrm{inv}} + \sum_{i=1}^n \mathbf{1}\{Z_i < 0\}$, therefore,

$$\begin{aligned}
\mathrm{Var}\left(\mathbb{E}(X^B_{\mathrm{inv}} \mid Z_k)\right) &= \mathrm{Var}\left(\mathbb{E}(X^D_{\mathrm{inv}} \mid Z_k) + \sum_{j=1}^n \mathbb{E}(\mathbf{1}\{Z_j < 0\} \mid Z_k)\right) \\
&= \mathrm{Var}\left(\mathbb{E}(X^D_{\mathrm{inv}} \mid Z_k) + \mathbb{E}(\mathbf{1}\{Z_k < 0\} \mid Z_k) + \text{const}\right) \\
&= \mathrm{Var}\left(\mathbb{E}(X^D_{\mathrm{inv}} \mid Z_k) + \mathbf{1}\{Z_k < 0\}\right) \\
&= \mathrm{Var}\Big((k-1)\mathbf{1}\{Z_k > 0\}(1-U_k) + (k-1)\mathbf{1}\{Z_k < 0\}(1+U_k) \\
&\qquad\qquad + \mathbf{1}\{Z_k < 0\}\Big).
\end{aligned}$$

Using the standard formula again, we have

$$\begin{aligned}
\mathbb{E}\left(\mathbb{E}(X_{\mathrm{inv}}^B \mid Z_k)^2\right) &= \frac{4}{3}(k-1)^2 + \mathbb{E}(\mathbf{1}\{Z_k < 0\}) + 2\mathbb{E}((k-1)(1+U_k)\mathbf{1}\{Z_k < 0\}) \\
&= \frac{4}{3}(k-1)^2 + \frac{1}{2} + \frac{3}{2}(k-1)\,, \\
\mathbb{E}\left(\mathbb{E}(X_{\mathrm{inv}}^B \mid Z_k)\right)^2 &= \left(\frac{1}{4}(k-1) + \frac{3}{4}(k-1) + \frac{1}{2}\right)^2 = \left((k-1) + \frac{1}{2}\right)^2 \\
&= (k-1)^2 + (k-1) + \frac{1}{2}\,.
\end{aligned}$$

In conclusion,

$$\operatorname{Var}\left(\mathbb{E}(X_{\mathrm{inv}}^B \mid Z_k)\right) = \frac{1}{3}(k-1)^2 + \frac{1}{2}(k-1)\,,$$

and in total,

$$\begin{aligned}
\operatorname{Var}(\hat{X}_{\mathrm{inv}}^B) &= \frac{1}{3}\frac{n(n-1)(2n-1)}{6} + \frac{1}{2}\frac{n(n-1)}{2} \\
&= \frac{1}{9}n^3 + \frac{1}{12}n^2 - \frac{7}{36}n = \frac{1}{9}n^3 + O(n^2)\,.
\end{aligned}$$

The claim follows from Theorem 4.1.2. □

4.6.2 Proof of Lemma 4.2.3a)

On the symmetric groups, we compute $\operatorname{Cov}(X_{\mathrm{inv}}, X_{\mathrm{des}})$ from (2.2), i.e., we have

$$\begin{aligned}
\operatorname{Cov}(X_{\mathrm{inv}}, X_{\mathrm{des}}) &= \sum_{i<j}\sum_{k=1}^{n-1} \operatorname{Cov}(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_k > Z_{k+1}\}) \\
&= \sum_{i<j}\sum_{\substack{k\in\{i-1,i,j-1,j\} \\ 1\le k\le n-1}} \operatorname{Cov}(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_k > Z_{k+1}\})\,,
\end{aligned}$$

where we used that if $k \notin \{i-1, i, j-1, j\}$, then the events $\{Z_i > Z_j\}$ and $\{Z_k > Z_{k+1}\}$ are independent, and therefore $\operatorname{Cov}(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_k > Z_{k+1}\}) = 0$. In what follows, we analyze the case $k \in \{i-1, i, j-1, j\}$, first assuming that all these numbers are distinct. Moreover, we temporarily ignore the boundary cases of $i = 1$ (where $k = i-1$ is outside the range of $\{1, \dots, n\}$) and $j = n$ (where $k = n$ is within the range but the variable Z_{k+1} compared with Z_k is not). This gives four possible constellations:

- type I: $k+1 = i$ and $j > k+2$,
- type II: $k = i$ and $j > k+1$,
- type III: $k+1 = j$ and $i < k$,
- type IV: $k = j$ and $i < k-1$.

For type I, we have

$$\operatorname{Cov}(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_{i-1} > Z_i\}) = \mathbb{P}(Z_i > Z_j, Z_{i-1} > Z_i) - \underbrace{\mathbb{P}(Z_i > Z_j)\mathbb{P}(Z_{i-1} > Z_i)}_{=1/4}\,,$$

while for type II,

$$\operatorname{Cov}(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_i > Z_{i+1}\}) = \mathbb{P}(Z_i > Z_j, Z_i > Z_{i+1}) - \mathbb{P}(Z_i > Z_j)\mathbb{P}(Z_i > Z_{i+1})\,.$$

For three distinct real numbers a, b, c, there are six possible orderings, all of which are equally likely in case of a, b, c being uniform variables. The event $\{a > b,\, a > c\}$ found in type II is the union of $\{a > b > c\}$ and $\{a > c > b\}$, while the event $\{a > b,\, b > c\}$ found in type I is only a redundant writing of $\{a > b > c\}$. Therefore, in type II,

$$\mathrm{Cov}(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_i > Z_{i+1}\}) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}\,,$$

while in type I,

$$\mathrm{Cov}(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_{i-1} > Z_i\}) = \frac{1}{6} - \frac{1}{4} = -\frac{1}{12}\,.$$

The types III and IV are handled the same way. For type III, we have

$$\mathrm{Cov}(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_{j-1} > Z_j\}) = \frac{1}{12}\,,$$

and for type IV,

$$\mathrm{Cov}(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_j > Z_{j+1}\}) = -\frac{1}{12}\,.$$

So, in the case of $j \neq i+1$ and $i \neq 1, j \neq n$, the inner sum

$$\sum_{k=1}^{n-1} \mathrm{Cov}(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_k > Z_{k+1}\})$$

consists of two canceling pairs of $1/12$ and $-1/12$, and vanishes altogether. Figure 4.6 displays the passage of k over the indices $1, \dots, n$ and the positions of the positive and negative covariances.

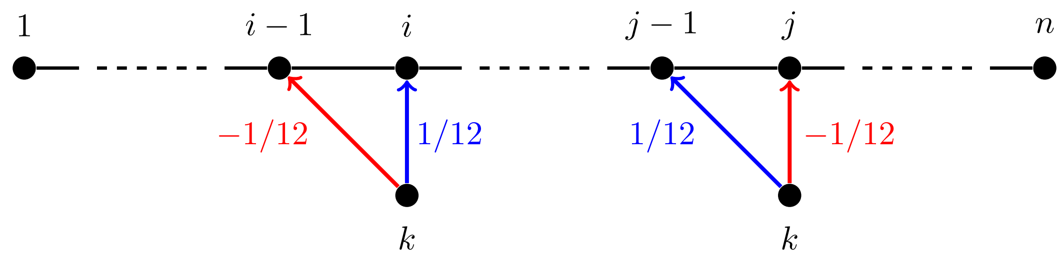


Figure 4.6: Canceling pairs of positive and negative covariances between $\mathbf{1}\{Z_i > Z_j\}$ and $\mathbf{1}\{Z_k > Z_{k+1}\}$ as k passes over $1, \dots, n-1$ in the non-exceptional case of $1 < i < i+1 < j < n$. The covariance is zero for all other values of k.

With the help of this figure, we can also see what happens if $1 < i < i+1 < j < n$ does not hold:

- If i and j are subsequent, i.e., $j = i+1$, then the two positive contributions in Figure 4.6 collide. Moreover, for $k = i$ we obtain $\mathrm{Cov}(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_k > Z_{k+1}\}) = \mathrm{Cov}(\mathbf{1}\{Z_i > Z_{i+1}\}, \mathbf{1}\{Z_i > Z_{i+1}\}) = \mathrm{Var}(\mathbf{1}\{Z_i > Z_{i+1}\}) = 1/4$.
- If $i = 1$, then the leftmost negative contribution in Figure 4.6 disappears.
- If $j = n$, then the rightmost negative contribution in Figure 4.6 disappears.

As these situations are not mutually exclusive, we obtain the following list of exceptional cases and their contributions $C_{ij} := \sum_{k=1}^{n-1} \mathrm{Cov}(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_k > Z_{k+1}\})$:

- (A1): $j = i+1$, but neither $i = 1$ nor $j = n \Longrightarrow C_{ij} = 1/4 - 1/6 = 1/12$,
- (A2): $i = 1$ and $j = 3, \dots, n-1 \Longrightarrow C_{ij} = 1/12$,
- (A3): $j = n$ and $i = 2, \dots, n-2 \Longrightarrow C_{ij} = 1/12$,
- (A4): $i = 1$, $j = 2 \Longrightarrow C_{ij} = 1/4 - 1/12 = 1/6$,
- (A5): $i = n-1$, $j = n \Longrightarrow C_{ij} = 1/6$,
- (A6): $i = 1$, $j = n \Longrightarrow C_{ij} = 1/6$.

As an example, we display the case (A1) in Figure 4.7.

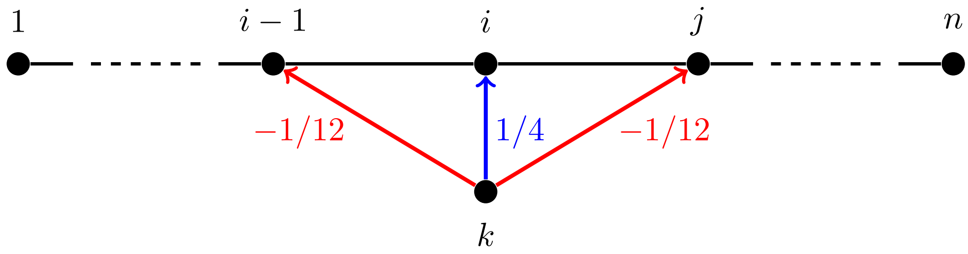


Figure 4.7: Display of covariances between $\mathbf{1}\{Z_i > Z_j\}$ and $\mathbf{1}\{Z_k > Z_{k+1}\}$ in the case of i and $j = i+1$ being subsequent.

Taking the contributions and frequencies of (A1), ..., (A6) into account, we obtain the exact result

$$\mathrm{Cov}(X_{\mathrm{inv}}, X_{\mathrm{des}}) = \underbrace{(n-3)\left(\frac{1}{4} - \frac{1}{6}\right)}_{\text{(A1)}} + \underbrace{2(n-3)\frac{1}{12}}_{\text{(A2, A3)}} + \underbrace{2\left(\frac{1}{4} - \frac{1}{12}\right)}_{\text{(A4, A5)}} + \underbrace{\frac{1}{6}}_{\text{(A6)}} = \frac{n-1}{4}\,.$$

For the groups B_n and D_n, the calculation largely follows the same procedure as that for S_n. Recall that now, $Z_1, \dots, Z_n \sim U(-1,1)$. On D_n, we have by (4.2b) and (4.3b) that

$$\mathrm{Cov}(X^D_{\mathrm{inv}}, X^D_{\mathrm{des}}) = \sum_{i<j} \sum_{k=1}^{n-1} \mathrm{Cov}\left(\mathbf{1}\{Z_i > Z_j\}, \mathbf{1}\{Z_k > Z_{k+1}\}\right) \tag{4.15a}$$

$$+ \sum_{i<j} \sum_{k=1}^{n-1} \mathrm{Cov}\left(\mathbf{1}\{-Z_i > Z_j\}, \mathbf{1}\{Z_k > Z_{k+1}\}\right) \tag{4.15b}$$

$$+ \sum_{i<j} \mathrm{Cov}(\mathbf{1}\{Z_i > Z_j\} + \mathbf{1}\{-Z_i > Z_j\}, \mathbf{1}\{-Z_2 > Z_1\})\,. \tag{4.15c}$$

The contribution of (4.15a) is $(n-1)/4$ as seen above. In (4.15b), we first demonstrate the cancellation in the non-exceptional case when $\{i-1, i, j-1, j\}$ form a set of distinct numbers. In that case, we have

$$\begin{aligned}
&\mathrm{Cov}(\mathbf{1}\{-Z_i > Z_j\}, \mathbf{1}\{Z_i > Z_{i+1}\}) + \mathrm{Cov}(\mathbf{1}\{-Z_i > Z_j\}, \mathbf{1}\{Z_{i-1} > Z_i\}) \\
&= \mathbb{E}(\mathbf{1}\{-Z_i > Z_j\}\mathbf{1}\{Z_i > Z_{i+1}\}) - 1/4 + \mathbb{E}(\mathbf{1}\{-Z_i > Z_j\}\mathbf{1}\{Z_{i-1} > Z_i\}) - 1/4 \\
&= \mathbb{E}(\mathbf{1}\{-Z_i > Z_j\}\mathbf{1}\{Z_i > Z_{i+1}\}) + \mathbb{E}(\mathbf{1}\{-Z_i > Z_j\}\mathbf{1}\{Z_{i+1} > Z_i\}) - 1/2 \\
&= \mathbb{E}(\mathbf{1}\{-Z_i > Z_j\}) - 1/2 = 0\,,
\end{aligned}$$

and accordingly,

$$\mathrm{Cov}(\mathbf{1}\{-Z_i > Z_j\}, \mathbf{1}\{Z_j > Z_{j+1}\}) + \mathrm{Cov}(\mathbf{1}\{-Z_i > Z_j\}, \mathbf{1}\{Z_{j-1} > Z_j\}) = 0\,.$$

However, this cancellation occurs not only in the non-exceptional case, but also in the cumulation of the exceptional cases (A1) – (A6) listed above, except for the covariances resulting from the collision of $j = i+1$ and $k = i$. Therefore, the next step is to compute $\mathrm{Cov}(\mathbf{1}\{-Z_k > Z_{k+1}\}, \mathbf{1}\{Z_k > Z_{k+1}\})$, $k = 1, \dots, n-1$. For $Z_1, Z_2 \sim U(-1,1)$, we compute

$$\begin{aligned}\mathbb{E}(\mathbf{1}\{-Z_1 > Z_2\}\mathbf{1}\{Z_1 > Z_2\}) &= \mathbb{P}(-Z_1 > Z_2, Z_1 > Z_2)\\ &= \mathbb{P}(Z_2 < 0, Z_2 < Z_1 < -Z_2)\\ &= \frac{1}{4}\int_{[-1,1]^2} \mathbf{1}\{y < 0, -|y| < x < |y|\}\mathrm{d}(x,y)\\ &= \frac{1}{4}\int_{-1}^{0}\int_{y}^{-y} 1\mathrm{d}x\mathrm{d}y = \frac{1}{4}\int_{-1}^{0} 2y\mathrm{d}y = \frac{1}{4}\,.\end{aligned}$$

Since we also have $\mathbb{P}(\mathbf{1}\{-Z_1 > Z_2\})\mathbb{P}(\mathbf{1}\{Z_1 > Z_2\}) = 1/4$, it follows that these two events are uncorrelated. The entire line (4.15b) contributes zero. Finally, consider (4.15c). Obviously, this double-indexed sum involves exactly the pairs (i,j) with $i = 1$ or $i = 2$. We get

$$\begin{aligned}(4.15c) &= \sum_{i=1}^{2}\sum_{j=3}^{n} \mathrm{Cov}(\mathbf{1}\{Z_i > Z_j\} + \mathbf{1}\{-Z_i > Z_j\}, \mathbf{1}\{-Z_2 > Z_1\})\\ &\quad + \underbrace{\mathrm{Cov}(\mathbf{1}\{-Z_1 > Z_2\}, \mathbf{1}\{-Z_2 > Z_1\})}_{=1/4} + \underbrace{\mathrm{Cov}(\mathbf{1}\{Z_1 > Z_2\}, \mathbf{1}\{-Z_2 > Z_1\})}_{=0}\\ &= \sum_{j=3}^{n}\Big[\underbrace{\mathrm{Cov}(\mathbf{1}\{Z_1 > Z_j\}, \mathbf{1}\{-Z_2 > Z_1\})}_{=-1/12} + \underbrace{\mathrm{Cov}(\mathbf{1}\{Z_2 > Z_j\}, \mathbf{1}\{-Z_2 > Z_1\})}_{=-1/12}\Big]\\ &\quad + \sum_{j=3}^{n}\Big[\underbrace{\mathrm{Cov}(\mathbf{1}\{-Z_j > Z_1\}, \mathbf{1}\{-Z_2 > Z_1\})}_{=1/12} + \underbrace{\mathrm{Cov}(\mathbf{1}\{-Z_j > Z_2\}, \mathbf{1}\{-Z_2 > Z_1\})}_{=1/12}\Big] + \frac{1}{4}\\ &= -\frac{2}{12}(n-2) + \frac{2}{12}(n-2) + \frac{1}{4} = \frac{1}{4}\,.\end{aligned}$$

Therefore, we obtain the overall result for D_n, which is

$$\mathrm{Cov}(X^D_{\mathrm{inv}}, X^D_{\mathrm{des}}) = \frac{n-1}{4} + \frac{1}{4} = \frac{n}{4}\,.$$

At last, we show that this result holds on B_n as well. By (4.2a) and (4.3a), we have

$$\begin{aligned}\mathrm{Cov}(X^B_{\mathrm{inv}}, X^B_{\mathrm{des}}) &= (4.15a) + (4.15b)\\ &\quad + \sum_{i=1}^{n}\sum_{k=1}^{n-1} \mathrm{Cov}(\mathbf{1}\{Z_i < 0\}, \mathbf{1}\{Z_k > Z_{k+1}\}) && (4.16a)\\ &\quad + \sum_{i<j} \mathrm{Cov}(\mathbf{1}\{Z_i > Z_j\} + \mathbf{1}\{-Z_i > Z_j\}, \mathbf{1}\{Z_1 < 0\}) && (4.16b)\\ &\quad + \sum_{i=1}^{n} \mathrm{Cov}(\mathbf{1}\{Z_i < 0\}, \mathbf{1}\{Z_1 < 0\})\,. && (4.16c)\end{aligned}$$

We now show that (4.16a) and (4.16b) vanish. In (4.16a), the inner sum only involves $k = i-1$ and $k = i$, therefore,

$$(4.16a) = \sum_{i=1}^{n} \mathrm{Cov}(\mathbf{1}\{Z_i < 0\}, \mathbf{1}\{Z_{i-1} > Z_i\}) + \mathrm{Cov}(\mathbf{1}\{Z_i < 0\}, \mathbf{1}\{Z_i > Z_{i+1}\})$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \operatorname{Cov}(\mathbf{1}\{Z_i < 0\}, \mathbf{1}\{Z_{i-1} > Z_i\}) + \operatorname{Cov}(\mathbf{1}\{Z_i < 0\}, \mathbf{1}\{Z_i > Z_{i-1}\}) \\
&= \mathbb{E}(\mathbf{1}\{Z_1 < 0\}) - \frac{1}{2} = 0\,.
\end{aligned}
$$

This cancellation even applies to the boundary terms $i = 1$ and $i = n$. We also get

$$
\begin{aligned}
(4.16b) &= \sum_{j=2}^{n} \operatorname{Cov}(\mathbf{1}\{Z_1 > Z_j\}, \mathbf{1}\{Z_1 < 0\}) + \operatorname{Cov}(\mathbf{1}\{-Z_1 > Z_j\}, \mathbf{1}\{Z_1 < 0\}) \\
&= \underbrace{\sum_{j=2}^{n} \mathbb{P}(Z_1 < 0, Z_1 > Z_j) - \frac{1}{4}}_{=-1/8} + \underbrace{\sum_{j=2}^{n} \mathbb{P}(Z_1 < 0, -Z_1 > Z_j) - \frac{1}{4}}_{=1/8} = 0\,.
\end{aligned}
$$

Finally,

$$
(4.16c) = \sum_{i=1}^{n} \operatorname{Cov}(\mathbf{1}\{Z_i < 0\}, \mathbf{1}\{Z_1 < 0\}) = \operatorname{Var}(\mathbf{1}\{Z_1 < 0\}) = \frac{1}{4}\,,
$$

giving the overall result $\operatorname{Cov}(X^B_{\text{inv}}, X^B_{\text{des}}) = (n-1)/4 + 1/4 = n/4$. □

4.6.3 Proof of Lemma 4.2.3b)

On the symmetric groups, we have by Lemma 4.1.3:

$$
\hat{X}_{\text{inv}} = \sum_{j=1}^{n} (n - 2j + 1) Z_j + \frac{1}{2}\binom{n}{2}\,,
$$

yielding

$$
\begin{aligned}
\operatorname{Cov}(\hat{X}_{\text{inv}}, X_{\text{des}}) &= \operatorname{Cov}\left(\sum_{j=1}^{n} (n-2j+1) Z_j, \sum_{k=1}^{n-1} \mathbf{1}\{Z_k > Z_{k+1}\}\right) \\
&= \sum_{j=2}^{n-1} \sum_{k=1}^{n-1} (n-2j+1) \operatorname{Cov}(Z_j, \mathbf{1}\{Z_k > Z_{k+1}\}) \qquad (4.17a) \\
&\quad + \sum_{j \in \{1,n\}} \sum_{k=1}^{n-1} (n-2j+1) \operatorname{Cov}(Z_j, \mathbf{1}\{Z_k > Z_{k+1}\})\,. \qquad (4.17b)
\end{aligned}
$$

Due to the independence of the Z_j, we get

$$
(4.17a) = \sum_{j=2}^{n-1} (n-2j+1)\big(\operatorname{Cov}(Z_j, \mathbf{1}\{Z_j < Z_{j-1}\}) + \operatorname{Cov}(Z_j, \mathbf{1}\{Z_j > Z_{j+1}\})\big) = 0\,,
$$

where the last equality follows from

$$
\begin{aligned}
&\operatorname{Cov}(Z_j, \mathbf{1}\{Z_j < Z_{j-1}\}) + \operatorname{Cov}(Z_j, \mathbf{1}\{Z_j > Z_{j+1}\}) \\
&= \mathbb{E}(Z_j \mathbf{1}\{Z_j < Z_{j-1}\}) + \mathbb{E}(Z_j \mathbf{1}\{Z_j > Z_{j+1}\}) - \frac{1}{2} \\
&= \mathbb{E}(Z_j \mathbf{1}\{Z_j < Z_{j-1}\}) + \mathbb{E}(Z_j \mathbf{1}\{Z_j > Z_{j-1}\}) - \frac{1}{2} = \mathbb{E}(Z_j) - \frac{1}{2} = 0\,.
\end{aligned}
$$

As $Z_1\mathbf{1}\{Z_1 < Z_2\}$ is a function of two uniform variables with joint density $f\colon \mathbb{R}^2 \to \mathbb{R}$, $(x,y) \mapsto \mathbf{1}\{(x,y) \in [0,1]^2\}$, we can apply Fubini's Theorem to obtain

$$\begin{aligned}\mathbb{E}(Z_1\mathbf{1}\{Z_1 < Z_2\}) &= \int_{[0,1]^2} x\mathbf{1}\{x<y\}\,\mathrm{d}(x,y) = \int_0^1 x\left(\int_0^1 \mathbf{1}\{x<y\}\mathrm{d}y\right)\mathrm{d}x \\ &= \int_0^1 x(1-x)\,\mathrm{d}x = \frac{1}{6}\,.\end{aligned}$$

Therefore, we get

$$\begin{aligned}(4.17\text{b}) &= (n-1)\mathrm{Cov}(Z_1, \mathbf{1}\{Z_1 > Z_2\}) - (n-1)\mathrm{Cov}(Z_n, \mathbf{1}\{Z_{n-1} > Z_n\}) \\ &= (n-1)(\mathbb{E}(Z_1\mathbf{1}\{Z_1 > Z_2\}) - \mathbb{E}(Z_1\mathbf{1}\{Z_1 < Z_2\})) \\ &= (n-1)\left(\frac{1}{2} - 2\mathbb{E}(Z_1\mathbf{1}\{Z_1 < Z_2\})\right) = \frac{n-1}{6}\,,\end{aligned}$$

which shows that $\mathrm{Cov}(\hat{X}_{\mathrm{inv}}, X_{\mathrm{des}}) = (n-1)/6$ on S_n. For the groups B_n and D_n, with $Z_k \sim U(-1,1)$ and the modifications (4.14), (4.3b), the calculation is more extensive but still follows the same procedure. On D_n, we have

$$\begin{aligned}\mathrm{Cov}(\hat{X}^D_{\mathrm{inv}}, X^D_{\mathrm{des}}) = \sum_{j=1}^{n}\sum_{k=1}^{n-1} &(j-1)\mathrm{Cov}((1-U_j)\mathbf{1}\{Z_j > 0\}, \mathbf{1}\{Z_k > Z_{k+1}\}) && (4.18\text{a}) \\ &+ (j-1)\mathrm{Cov}((1+U_j)\mathbf{1}\{Z_j < 0\}, \mathbf{1}\{Z_k > Z_{k+1}\}) && (4.18\text{b}) \\ &+\sum_{j=1}^{n}(j-1)\mathrm{Cov}((1-U_j)\mathbf{1}\{Z_j > 0\}, \mathbf{1}\{-Z_2 > Z_1\}) && (4.18\text{c}) \\ &+ (j-1)\mathrm{Cov}((1+U_j)\mathbf{1}\{Z_j < 0\}, \mathbf{1}\{-Z_2 > Z_1\})\,. && (4.18\text{d})\end{aligned}$$

In the first two lines (4.18a) and (4.18b), there is cancellation of all summands if $j \notin \{1,n\}$ due to previously used arguments. Only $j = n$ is relevant, as the summands are zero for $j = 1$. We have $\mathbb{E}((1-U_1)\mathbf{1}\{Z_1 > 0\})\mathbb{E}(\mathbf{1}\{Z_1 > Z_2\}) = 1/8$, and by Fubini's Theorem, we obtain for (4.18a):

$$\begin{aligned}\mathbb{E}\left((1-U_1)\mathbf{1}\{Z_1 > 0\}\mathbf{1}\{Z_1 > Z_2\}\right) &= \frac{1}{4}\int_{[-1,1]^2}(1-|x|)\mathbf{1}\{x>0\}\mathbf{1}\{x>y\}\mathrm{d}(x,y) \\ &= \frac{1}{4}\int_0^1\int_{-1}^1(1-x)\mathbf{1}\{x>y\}\mathrm{d}y\mathrm{d}x \\ &= \frac{1}{4}\int_0^1(1-x)\left(1+\int_0^1\mathbf{1}\{x>y\}\mathrm{d}y\right)\mathrm{d}x \\ &= \frac{1}{4}\int_0^1(1-x)(1+x)\mathrm{d}x = \frac{1}{6}\,.\end{aligned}$$

Accordingly, for (4.18b), we have $\mathbb{E}((1+U_1)\mathbf{1}\{Z_1 < 0\})\mathbb{E}(\mathbf{1}\{Z_1 > Z_2\}) = 3/8$ and

$$\begin{aligned}\mathbb{E}((1+U_1)\mathbf{1}\{Z_1 < 0\}\mathbf{1}\{Z_1 > Z_2\}) &= \frac{1}{4}\int_{[-1,0]\times[-1,1]}(1+|x|)\mathbf{1}\{x>y\}\mathrm{d}(x,y) \\ &= \frac{1}{4}\int_{-1}^0(1+|x|)(1-|x|)\mathrm{d}x = \frac{1}{6}\,.\end{aligned}$$

Next, (4.18c) and (4.18d) are non-zero only for $j = 1$ and $j = 2$. Since $\mathbf{1}\{-Z_2 > Z_1\}$ equals $\mathbf{1}\{-Z_1 > Z_2\}$, we have

$$(4.18\text{c}) + (4.18\text{d}) = (2n-3)\mathrm{Cov}\big(U_1\mathbf{1}\{Z_1 > 0\} + (2-U_1)\mathbf{1}\{Z_1 < 0\}, \mathbf{1}\{-Z_2 > Z_1\}\big)\,.$$

Again, by Fubini's theorem,

$$\operatorname{Cov}(U_1\mathbf{1}\{Z_1>0\},\mathbf{1}\{-Z_2>Z_1\}) = -\frac{1}{12}\,,$$
$$\operatorname{Cov}((2-U_1)\mathbf{1}\{Z_1<0\},\mathbf{1}\{-Z_2>Z_1\}) = \frac{1}{6}\,.$$

In total,

$$\operatorname{Cov}(\hat{X}^D_{\mathrm{inv}}, X^D_{\mathrm{des}}) = (n-1)\left(\frac{1}{12}-\frac{1}{6}\right) + (2n-3)\left(-\frac{1}{12}+\frac{1}{6}\right) = \frac{n-2}{12}\,.$$

On B_n, we have due to (4.3b):

$$\begin{aligned}
\operatorname{Cov}(\hat{X}^B_{\mathrm{inv}}, X^B_{\mathrm{des}}) = {} & (4.18a) + (4.18b)\\
& + \sum_{j=1}^{n}(j-1)\operatorname{Cov}(U_j\mathbf{1}\{Z_j>0\},\mathbf{1}\{Z_1<0\}) && \text{(4.19a)}\\
& + \sum_{j=1}^{n}(j-1)\operatorname{Cov}((2-U_j)\mathbf{1}\{Z_j<0\},\mathbf{1}\{Z_1<0\}) && \text{(4.19b)}\\
& + \underbrace{\operatorname{Cov}\left(\sum_{k=1}^{n}\mathbf{1}\{Z_k<0\},\sum_{j=1}^{n}\mathbf{1}\{Z_j>Z_{j+1}\}\right)}_{=0}\\
& + \underbrace{\operatorname{Cov}\left(\sum_{k=1}^{n}\mathbf{1}\{Z_k<0\},\mathbf{1}\{Z_1<0\}\right)}_{=1/4}.
\end{aligned}$$

Only $j=1$ is relevant in (4.19a) and (4.19b), and we easily obtain

$$\operatorname{Cov}(U_1\mathbf{1}\{Z_1>0\},\mathbf{1}\{Z_1<0\}) = -\frac{1}{8}\,,$$
$$\operatorname{Cov}((2-U_1)\mathbf{1}\{Z_1<0\},\mathbf{1}\{Z_1<0\}) = \frac{3}{8}\,,$$

giving

$$\operatorname{Cov}(\hat{X}^B_{\mathrm{inv}}, X^B_{\mathrm{des}}) = (n-1)\left(-\frac{1}{12}-\frac{1}{8}+\frac{3}{8}\right)+\frac{1}{4} = \frac{n-1}{6}+\frac{1}{4}$$

as the overall result. □

4.6.4 Proof of Lemma 4.2.8

We follow the calculation of $\operatorname{Var}(X_{\mathrm{inv}})$ in the uniform case provided in [66, Section 3]. So, the main task is to calculate $\mathbb{E}(X^2_{\mathrm{inv}})$. For X^B_{inv}, we recall (4.2a) and use the shorthand notations

$$X^B_{\mathrm{inv}} = \underbrace{\sum_{i<j}\mathbf{1}\{Z_i>Z_j\}}_{=:\,X^+} + \underbrace{\sum_{i<j}\mathbf{1}\{-Z_i>Z_j\}}_{=:\,X^-} + \underbrace{\sum_{i=1}^{n}\mathbf{1}\{Z_i<0\}}_{=:\,X^\circ} = X^+ + X^- + X^\circ.$$

This means we have to compute $\mathbb{E}((X_{\text{inv}}^B)^2)$ by the decomposition

$$\begin{aligned}\mathbb{E}((X_{\text{inv}}^B)^2) = \mathbb{E}((X^+)^2) + \mathbb{E}((X^-)^2) + 2\mathbb{E}(X^+X^-) \\ + \mathbb{E}((X^\circ)^2) + 2\mathbb{E}(X^+X^\circ) + 2\mathbb{E}(X^-X^\circ)\,. \end{aligned} \tag{4.20}$$

The first term $\mathbb{E}((X^+)^2)$ is invariant under p, since it only involves events $\mathbf{1}\{Z_i > Z_j\}$ for which $\mathbb{P}(Z_i > Z_j) = 1/2$, even if the involved Z_i, Z_j are not uniformly distributed. Therefore, we obtain $\mathbb{E}((X^+)^2)$ from [66, Section 3]:

$$\mathbb{E}((X^+)^2) = \frac{1}{2}\binom{n}{2} + \frac{1}{4}\binom{n}{2}\binom{n-2}{2} + \frac{5}{3}\binom{n}{3}\,.$$

Next, we turn to

$$\mathbb{E}((X^-)^2) = \sum_{i<j}\sum_{k<l}\mathbb{P}(-Z_i > Z_j, -Z_k > Z_l)\,.$$

For the $\binom{n}{2}\binom{n-2}{2}$ choices of distinct i, j, k, l, we have that $\mathbb{P}(-Z_i > Z_j, -Z_k > Z_l) = p^2$ due to independence, and for the $\binom{n}{2}$ cases of $(i, j) = (k, l)$, we simply get $\mathbb{P}(-Z_i > Z_j) = p$. The set of triples with exactly two of the indices colliding needs to be analyzed similar to [66, p. 443f.]. Note that the cases $i = k$ and $j = l$ are counted twice. For instance, in the case of $i = l$, we perform a case distinction based on the signs of Z_i, Z_j, and Z_k:

$$\begin{aligned}\mathbb{P}(-Z_i > Z_j, -Z_k > Z_i) &= \mathbb{P}(-Z_i > Z_j, -Z_k > Z_i, Z_i > 0) \\ &\quad + \mathbb{P}(-Z_i > Z_j, -Z_k > Z_i, Z_i < 0) \\ &= \mathbb{P}(-Z_i > Z_j, -Z_k > Z_i, Z_i > 0, Z_j < 0, Z_k < 0) \\ &\quad + \mathbb{P}(Z_i < 0, Z_j < 0, Z_k < 0) \\ &\quad + \mathbb{P}(Z_i < 0, Z_j > 0, Z_k < 0, -Z_i > Z_j) \\ &\quad + \mathbb{P}(Z_i < 0, Z_j < 0, Z_k > 0, -Z_k > Z_i) \\ &\quad + \mathbb{P}(Z_i < 0, Z_j > 0, Z_k > 0, -Z_i > Z_j, -Z_k > Z_i) \\ &= \frac{1}{3}p^2q + p^3 + \frac{1}{2}p^2q + \frac{1}{2}p^2q + \frac{1}{3}pq^2 \\ &= \frac{1}{3}p(2p+1)\,.\end{aligned}$$

It turns out that each of the six triples gives this contribution. So, overall,

$$\mathbb{E}((X^-)^2) = \binom{n}{2}p + \binom{n}{2}\binom{n-2}{2}p^2 + \binom{n}{3}2p(2p+1)\,.$$

For $\mathbb{E}(X^+X^-)$, each of the disjoint quadruples gives a contribution of $p/2$. For the colliding pairs, we have to compute

$$\begin{aligned}&\mathbb{P}(Z_i > Z_j, -Z_i > Z_j) \\ &= \underbrace{\mathbb{P}(Z_i > Z_j, -Z_i > Z_j, Z_i > 0, Z_j > 0)}_{=0} + \underbrace{\mathbb{P}(Z_i > Z_j, -Z_i > Z_j, Z_i < 0, Z_j > 0)}_{=0} \\ &\qquad + \mathbb{P}(Z_i > Z_j, -Z_i > Z_j, Z_i < 0, Z_j < 0) + \mathbb{P}(Z_i > Z_j, -Z_i > Z_j, Z_i > 0, Z_j < 0) \\ &= \mathbb{P}(Z_i < 0, Z_j < 0, Z_i > Z_j) + \mathbb{P}(Z_i > 0, Z_j < 0, -Z_i > Z_j) = \frac{p^2}{2} + \frac{pq}{2} = \frac{p}{2}\,.\end{aligned}$$

For the triples, we repeat the procedure described above. For the cases of $j = k$ and $j = l$, we get

$$\mathbb{P}(Z_i > Z_j, -Z_j > Z_k) = -\frac{1}{6}p(2p-5)\,.$$

For the cases $i = l$ and $i = k$, the calculation differs slightly:

$$\begin{aligned}
\mathbb{P}(Z_i > Z_j, -Z_i > Z_l) &= \mathbb{P}(Z_i > Z_j, -Z_i > Z_l, Z_i > 0, Z_l < 0)\\
&\quad + \mathbb{P}(Z_i > Z_j, -Z_i > Z_l, Z_i < 0, Z_j < 0)\\
&= \mathbb{P}(Z_i > 0, Z_j < 0, Z_l < 0, -Z_i > Z_l)\\
&\quad + \mathbb{P}(Z_i > 0, Z_j > 0, Z_l < 0, Z_j < Z_i < -Z_l)\\
&\quad + \mathbb{P}(Z_i < 0, Z_j < 0, Z_l < 0, Z_i > Z_j)\\
&\quad + \mathbb{P}(Z_i < 0, Z_j < 0, Z_l > 0, Z_j < Z_i < -Z_l)\\
&= \frac{1}{2}p^2q + \frac{1}{6}pq^2 + \frac{1}{2}p^3 + \frac{1}{6}p^2q\\
&= \frac{1}{6}p(2p+1)\,.
\end{aligned}$$

Overall, we obtain

$$\begin{aligned}
\mathbb{E}(X^+X^-) &= \binom{n}{2}\frac{p}{2} + \binom{n}{2}\binom{n-2}{2}\frac{p}{2} + 3\binom{n}{3}\left(\frac{1}{6}p(2p+1) - \frac{1}{6}p(2p-5)\right)\\
&= \binom{n}{2}\frac{p}{2} + \binom{n}{2}\binom{n-2}{2}\frac{p}{2} + 3\binom{n}{3}p\,.
\end{aligned}$$

The remaining three terms in (4.20) are easily calculated as

$$\begin{aligned}
\mathbb{E}(X^+X^\circ) &= \sum_{i<j}\sum_{k=1}^{n}\mathbb{P}(Z_i > Z_j, Z_k < 0)\\
&= 3\binom{n}{3}\frac{p}{2} + \sum_{i<j}\mathbb{P}(Z_i > Z_j, Z_i < 0) + \sum_{i<j}\mathbb{P}(Z_i > Z_j, Z_j < 0)\\
&= 3\binom{n}{3}\frac{p}{2} + \binom{n}{2}(p^2 + pq) = 3\binom{n}{3}\frac{p}{2} + \binom{n}{2}p\,,\\
\mathbb{E}(X^-X^\circ) &= \sum_{i<j}\sum_{k=1}^{n}\mathbb{P}(-Z_i > Z_j, Z_k < 0)\\
&= 3\binom{n}{3}p^2 + 2\binom{n}{2}\left(p^2 + \frac{pq}{2}\right) = 3\binom{n}{3}p^2 + 2\binom{n}{2}(p^2 + p)\,,\\
\mathbb{E}((X^\circ)^2) &= \sum_{i,j=1}^{n}\mathbb{P}(Z_i < 0, Z_j < 0) = 2\binom{n}{2}p^2 + np\,.
\end{aligned}$$

Summing all terms in (4.20) and subtracting the square of the mean gives the claim for B_n. On D_n, we omit the parts involving X° and get the corresponding result. □

4.6.5 Proof of Lemma 4.2.9

We follow the proof of Lemma 4.2.1, and again we first prove the claim on the even-signed permutation groups D_n. For any bias p, it still holds that

$$\mathbb{E}(X_{\text{inv}}^D \mid Z_k) = \sum_{i=1}^{k-1} \mathbb{E}(f(Z_i, Z_k) \mid Z_k) + \sum_{j=k+1}^{n} \mathbb{E}(f(Z_k, Z_j) \mid Z_k)\,,$$

with $\mathbb{E}\big(f(Z_i, Z_k) \mid Z_k\big)$ and $\mathbb{E}\big(f(Z_k, Z_j) \mid Z_k\big)$ as stated in the proof of Lemma 4.2.1. In the p-biased case, we have

$$\begin{aligned}\mathbb{E}\big(f(Z_k, Z_j) \mid Z_k\big) &= \mathbb{P}(Z_j < 0)U_k + \mathbb{P}(Z_j > 0)(1 + 1 - U_k) \\ &= (1-p)U_k + p(2 - U_k) = -2pU_k + U_k + 2p\,.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}(X_{\text{inv}}^D \mid Z_k) = (k-1)\big(\mathbf{1}\{Z_k > 0\}(1 - U_k) + \mathbf{1}\{Z_k < 0\}(1 + U_k)\big) \\ + (n-k)(-2pU_k + U_k + 2p) + \text{const}\,.\end{aligned}$$

Ignoring the constant part with a slight abuse of notation, we now compute

$$\begin{aligned}&\mathbb{E}\left(\mathbb{E}(X_{\text{inv}}^D \mid Z_k)^2\right) \\ &= (n-k)^2\mathbb{E}\left((-2pU_k + U_k + 2p)^2\right) && \text{(4.21a)} \\ &+ (k-1)^2\mathbb{E}\left(\big(\mathbf{1}\{Z_k > 0\}(1 - U_k) + \mathbf{1}\{Z_k < 0\}(1 + U_k)\big)^2\right) && \text{(4.21b)} \\ &+ 2(k-1)(n-k)\mathbb{E}\Big((-2pU_k + U_k + 2p)\big(\mathbf{1}\{Z_k > 0\}(1 - U_k) \\ &\qquad + \mathbf{1}\{Z_k < 0\}(1 + U_k)\big)\Big)\,. && \text{(4.21c)}\end{aligned}$$

The sign of Z_k is independent of U_k by construction. Therefore,

$$\begin{aligned}\text{(4.21a)} &= (n-k)^2\mathbb{E}\left((4p^2 - 4p + 1)U_k^2 + 4p(-2p + 1)U_k + 4p^2\right) \\ &= (n-k)^2\left(\frac{4}{3}p^2 + \frac{2}{3}p + \frac{1}{3}\right), \\ \text{(4.21b)} &= (k-1)^2\left(q\mathbb{E}((1 - U_k))^2 + p\mathbb{E}((1 + U_k)^2)\right) \\ &= (k-1)^2\left(\frac{q}{3} + \frac{7}{3}p\right) = (k-1)^2\left(2p + \frac{1}{3}\right), \\ \text{(4.21c)} &= 2(k-1)(n-k)\Big((1-p)\mathbb{E}((1 - U_k)(-2pU_k + U_k + 2p)) \\ &\qquad + p\mathbb{E}((1 + U_k)(-2pU_k + U_k + 2p))\Big) \\ &= 2(k-1)(n-k)\left((1-p)\left(\frac{2}{3}p + \frac{1}{6}\right) + p\left(\frac{4}{3}p + \frac{5}{6}\right)\right) \\ &= 2(k-1)(n-k)\left(\frac{2}{3}p^2 + \frac{4p}{3} + \frac{1}{6}\right).\end{aligned}$$

In total,

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X_{\mathrm{inv}}^{D} \mid Z_k))^2) &= (n-k)^2\left(\frac{4}{3}p^2+\frac{2}{3}p+\frac{1}{3}\right)+(k-1)^2\left(2p+\frac{1}{3}\right)\\ &\quad+2(k-1)(n-k)\left(\frac{2}{3}p^2+\frac{4}{3}p+\frac{1}{6}\right).\end{aligned}$$

We subtract the square of

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X_{\mathrm{inv}}^{D} \mid Z_k)) &= (n-k)\left(p+\frac{1}{2}\right)+(k-1)\left(\frac{1}{2}-\frac{p}{2}+\frac{3}{2}p\right)\\ &= (n-k)\left(p+\frac{1}{2}\right)+(k-1)\left(p+\frac{1}{2}\right)\\ &= \left(p+\frac{1}{2}\right)(n-1).\end{aligned}$$

The total variance of $\hat{X}_{\mathrm{inv}}^{D}$ is

$$\begin{aligned}\operatorname{Var}(\hat{X}_{\mathrm{inv}}^{D}) &= \sum_{k=1}^{n}\mathbb{E}(\mathbb{E}(X_{\mathrm{inv}}^{D}) \mid Z_k)^2)-\left(\mathbb{E}(\mathbb{E}(X_{\mathrm{inv}}^{D}) \mid Z_k)\right)^2\\ &= \sum_{k=1}^{n}\mathbb{E}(\mathbb{E}(X_{\mathrm{inv}}^{D}) \mid Z_k)^2)-n(n-1)^2\left(p+\frac{1}{2}\right)^2,\end{aligned}$$

so, to conclude the proof for D_n, we compute

$$\begin{aligned}\sum_{k=1}^{n}\mathbb{E}(\mathbb{E}(X_{\mathrm{inv}}^{D} \mid Z_k)^2) &= \left(\frac{4}{3}p^2+\frac{2}{3}p+\frac{1}{3}\right)\sum_{k=1}^{n}(k-1)^2+\left(2p+\frac{1}{3}\right)\sum_{k=1}^{n}(n-k)^2\\ &\quad+\left(\frac{4}{3}p^2+\frac{8p}{3}+\frac{1}{3}\right)\sum_{k=1}^{n}(k-1)(n-k)\\ &= \left(\frac{4}{3}p^2+\frac{2}{3}p+\frac{1}{3}\right)\cdot\frac{1}{6}n(n-1)(2n-1)\\ &\quad+\left(2p+\frac{1}{3}\right)\cdot\frac{1}{6}n(n-1)(2n-1)\\ &\quad+\left(\frac{4}{3}p^2+\frac{8p}{3}+\frac{1}{3}\right)\cdot\frac{1}{6}n(n-1)(n-2)\\ &= n^3\left(\frac{2}{3}p^2+\frac{4}{3}p+\frac{5}{18}\right)-n^2\left(\frac{4}{3}p^2-\frac{8}{3}p-\frac{1}{2}\right)\\ &\quad+n\left(\frac{2}{3}p^2+\frac{4}{3}p+\frac{2}{9}\right).\end{aligned}$$

Subtracting $n(n-1)^2\,(p+1/2)^2$ gives the desired leading term stated for $\operatorname{Var}(X_{\mathrm{inv}}^{D})$ in Lemma 4.2.8, i.e.,

$$\operatorname{Var}(X_{\mathrm{inv}}^{D}), \operatorname{Var}(\hat{X}_{\mathrm{inv}}^{D}) = \left(-\frac{1}{3}p^2+\frac{1}{3}p+\frac{1}{36}\right)n^3+O(n^2).$$

On the groups B_n, we achieve the same result since the additional parts in $\operatorname{Var}(\hat{X}_{\mathrm{inv}}^{B})$ yielded by $\sum_{i=1}^{n}\mathbf{1}\{Z_i<0\}$ are asymptotically negligible. Recall that

$$X_{\mathrm{inv}}^{B} = X_{\mathrm{inv}}^{D}+\sum_{i=1}^{n}\mathbf{1}\{Z_i<0\},$$

therefore,

$$\begin{aligned}
\operatorname{Var}\left(\mathbb{E}(X^B_{\text{inv}} \mid Z_k)\right) &= \operatorname{Var}\left(\mathbb{E}(X^D_{\text{inv}} \mid Z_k) + \sum_{j=1}^{n} \mathbb{E}(\mathbf{1}\{Z_j < 0\} \mid Z_k)\right) \\
&= \operatorname{Var}\left(\mathbb{E}(X^D_{\text{inv}} \mid Z_k) + \mathbb{E}(\mathbf{1}\{Z_k < 0\} \mid Z_k) + \text{const}\right) \\
&= \operatorname{Var}\left(\mathbb{E}(X^D_{\text{inv}} \mid Z_k) + \mathbf{1}\{Z_k < 0\}\right) \\
&= \operatorname{Var}\Big((k-1)\mathbf{1}\{Z_k > 0\}(1-U_k) + (k-1)\mathbf{1}\{Z_k < 0\}(1+U_k) \\
&\qquad + (n-k)(-2pU_k + U_k + 2p) + \mathbf{1}\{Z_k < 0\} + \text{const}\Big) .
\end{aligned}$$

Using the standard formula $\operatorname{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ again, we have

$$\begin{aligned}
\mathbb{E}\left(\mathbb{E}(X^B_{\text{inv}} \mid Z_k)^2\right) &= \mathbb{E}\left(\mathbb{E}(X^D_{\text{inv}} \mid Z_k)^2\right) + \mathbb{E}\left(\mathbf{1}\{Z_k < 0\}\right) \\
&\quad + 2\mathbb{E}((k-1)(1+U_k)\mathbf{1}\{Z_k < 0\}) \\
&\quad + 2\mathbb{E}((n-k)(-2pU_k + U_k + 2p)\mathbf{1}\{Z_k < 0\}) \\
&= \mathbb{E}\left(\mathbb{E}(X^D_{\text{inv}} \mid Z_k)^2\right) + p + 3p(n-k) + (k-1)(2p+1)p, \\
\mathbb{E}\left(\mathbb{E}(X^B_{\text{inv}} \mid Z_k)\right)^2 &= \mathbb{E}\left(\mathbb{E}(X^D_{\text{inv}} \mid Z_k)\right)^2 + p^2 + 2\mathbb{E}(\mathbf{1}\{Z_k < 0\})\mathbb{E}\left(\mathbb{E}(X^D_{\text{inv}} \mid Z_k)\right) \\
&= \left(p + \frac{1}{2}\right)^2 (n-1)^2 + p^2 + 2p(n-1)\left(p + \frac{1}{2}\right) .
\end{aligned}$$

In conclusion,

$$\begin{aligned}
\operatorname{Var}(\hat{X}^B_{\text{inv}}) &= \operatorname{Var}(\hat{X}^D_{\text{inv}}) + n(p - p^2) - p(2p+1)\frac{n(n-1)}{2} \\
&= \operatorname{Var}(\hat{X}^D_{\text{inv}}) + O(n^2) .
\end{aligned}$$

The claim follows from Theorem 4.1.2. □

4.6.6 Proof of Lemma 4.2.11

At first, we compute $\operatorname{Cov}(X_{\text{inv}}, X_{\text{des}})$ on B_n and D_n in the p-biased case. Regardless of the bias, it always holds that

$$\operatorname{Cov}(X^D_{\text{inv}}, X^D_{\text{des}}) = (4.15\text{a}) + (4.15\text{b}) + (4.15\text{c})$$

and (4.15a) $= (n-1)/4$. The cancellation arguments for (4.15b) still hold as well, except that $\forall k = 1, \ldots, n-1$:

$$\operatorname{Cov}(\mathbf{1}\{-Z_k > Z_{k+1}\}, \mathbf{1}\{Z_k > Z_{k+1}\}) = \frac{p}{2} - \frac{p}{2} = 0 ,$$

so, interestingly, (4.15b) $= 0$. (4.15c) is calculated in the same way as in the proof of Lemma 4.2.3a), with the only differences that $\operatorname{Cov}(\mathbf{1}\{Z_1 > Z_2\}, \mathbf{1}\{-Z_2 > Z_1\}) = 0$ and $\operatorname{Cov}(\mathbf{1}\{-Z_1 > Z_2\}, \mathbf{1}\{-Z_2 > Z_1\}) = \operatorname{Var}(\mathbf{1}\{-Z_1 > Z_2\}) = pq$. Overall,

$$\operatorname{Cov}(X^D_{\text{inv}}, X^D_{\text{des}}) = \frac{n-1}{4} + pq ,$$

as claimed. For $\mathrm{Cov}(X^B_{\mathrm{inv}}, X^B_{\mathrm{des}}) = (4.15a) + (4.15b) + (4.16a) + (4.16b) + (4.16c)$, we again see that (4.16a) and (4.16b) vanish, and with $(4.16c) = \mathrm{Var}(\mathbf{1}\{Z_1 < 0\}) = pq$, the claim follows as well.

Next, we have to compute $\mathrm{Cov}(\hat{X}^D_{\mathrm{inv}}, X^D_{\mathrm{des}})$ in the p-biased case. A significant change to the unbiased case is that $\hat{X}^D_{\mathrm{inv}}$ now contains $\sum_{k=1}^n (k-1)(2pU_k - U_k + 1)$, which is a negligible constant in the unbiased case. In the p-biased case, we have

$$\begin{aligned}\mathrm{Cov}(\hat{X}^D_{\mathrm{inv}}, X^D_{\mathrm{des}}) = {} & (4.18a) + (4.18b) + (4.18c) + (4.18d) \\ & + \sum_{j=1}^{n}\sum_{k=1}^{n-1}(j-1)\mathrm{Cov}(2pU_j - U_j + 1, \mathbf{1}\{Z_k > Z_{k+1}\}) \qquad (4.22a) \\ & + \sum_{j=1}^{n}(j-1)\mathrm{Cov}(2pU_j - U_j + 1, \mathbf{1}\{-Z_2 > Z_1\}) \,. \qquad (4.22b)\end{aligned}$$

In (4.18a), (4.18b), and (4.22a), there is cancellation if $j \notin \{1, n\}$ due to previously used arguments. Only $j = 1$ is relevant in (4.18a) and (4.18b), while only $j = n$ is relevant in (4.22a). We have $\mathbb{E}(U_j\mathbf{1}\{Z_j > 0\})\mathbb{E}(\mathbf{1}\{Z_j > Z_{j+1}\}) = q/4$, and the joint density of Z_j and Z_{j+1} is

$$f_p(x,y) := f_p(x)f_p(y) = \begin{cases} p^2, & x, y < 0 \\ pq, & x > 0, y < 0 \\ pq, & x < 0, y > 0 \\ q^2, & x, y > 0 \end{cases}.$$

By Fubini's Theorem, we obtain

$$\begin{aligned}\mathbb{E}(U_1\mathbf{1}\{Z_1 > 0\}\mathbf{1}\{Z_1 > Z_2\}) &= \int_{[-1,1]^2} |x|\mathbf{1}\{x > 0\}\mathbf{1}\{x > y\} f_p(x,y)\mathrm{d}(x,y) \\ &= \int_{[0,1]^2} q^2 x\mathbf{1}\{x > y\}\mathrm{d}(x,y) \\ &\quad + \int_{[0,1]\times[-1,0]} pqx\mathbf{1}\{x > y\}\mathrm{d}(x,y) \\ &= q^2\int_0^1 x^2 dx + pq\int_0^1 x\mathrm{d}x = \frac{q^2}{3} + \frac{pq}{2}\,,\end{aligned}$$

and accordingly,

$$\begin{aligned}\mathbb{E}((2-U_1)\mathbf{1}\{Z_1 < 0\}\mathbf{1}\{Z_1 > Z_2\}) &= \int_{-1}^{0}\int_{-1}^{1}(2-|x|)\mathbf{1}\{x > y\}f_p(x,y)\mathrm{d}y\mathrm{d}x \\ &= p^2\int_{-1}^{0}(2+x)(1+x)\mathrm{d}x = \frac{5}{6}p^2, \\ \mathbb{E}((2p-1)U_n\mathbf{1}\{Z_{n-1} > Z_n\}) &= (2p-1)\int_{[-1,1]^2}|x|\mathbf{1}\{x < y\}f_p(x,y)\mathrm{d}(x,y) \\ &= (2p-1)\left(\frac{p^2}{3} + \frac{q^2}{6} + \frac{pq}{2}\right) = (2p-1)\frac{p+1}{6}\,.\end{aligned}$$

Therefore,

$$(4.18a) + (4.18b) + (4.22a) = (n-1)\left(p^2 - p + \frac{1}{6}\right). \qquad (4.23)$$

Moreover, we again have by Fubini's Theorem:

$$\begin{aligned}\mathrm{Cov}(U_1\mathbf{1}\{Z_1>0\},\mathbf{1}\{-Z_1>Z_2\}) &= \frac{pq}{6}-\frac{p}{4}\,,\\ \mathrm{Cov}((2-U_1)\mathbf{1}\{Z_1<0\},\mathbf{1}\{-Z_1>Z_2\}) &= \frac{2}{3}pq\,,\end{aligned}$$

giving

$$(4.18c)+(4.18d)=(2n-3)\left(\frac{5}{6}pq-\frac{p}{4}\right).$$

At last,

$$(4.22b)=(2p-1)\mathrm{Cov}(U_2,\mathbf{1}\{-Z_2>Z_1\})=(2p-1)\left(\frac{p^2}{2}+\frac{pq}{2}-\frac{p}{2}\right)=0.$$

By taking the sum (4.23) + (4.18c) + (4.18d) + (4.22b), the claim follows for D_n. Finally, for B_n, recall that $\mathbb{E}(X^B_{\text{inv}}\mid Z_j)=\mathbb{E}(X^D_{\text{inv}}\mid Z_j)+\mathbf{1}\{Z_j<0\}+\text{const}$. We compute

$$\begin{aligned}\mathrm{Cov}(\hat{X}^B_{\text{inv}},X^B_{\text{des}}) &= (4.23)+\underbrace{\sum_{j=1}^n (j-1)\mathrm{Cov}(U_j\mathbf{1}\{Z_j<0\}+1+U_j\mathbf{1}\{Z_j<0\},\mathbf{1}\{Z_1<0\})}_{=(n-1)pq}\\ &\quad+\underbrace{\sum_{j=1}^n (j-1)\mathrm{Cov}(2pU_j-U_j+1,\mathbf{1}\{Z_1<0\})}_{=0}\\ &\quad+\underbrace{\mathrm{Cov}\left(\sum_{j=1}^n\mathbf{1}\{Z_j<0\},\sum_{k=1}^n\mathbf{1}\{Z_k>Z_{k+1}\}+\mathbf{1}\{Z_1<0\}\right)}_{=pq}\\ &= (4.23)+(n-1)\left(-\frac{pq}{2}\right)+(n-1)pq+pq\\ &= (n-1)(p^2+p+\frac{1}{6}+p-p^2)+pq\,,\end{aligned}$$

from which the claim follows for B_n as well. □

4.6.7 Simulation code

The simulations in Section 4.5.1 examine the limit behavior of the row-wise maxima $M_n=\max\{X_{n1},\dots,X_{nn}\}$, drawn from a uniform triangular array with $X_{n1}\overset{\mathcal{D}}{=}(X_{\text{inv}},X_{\text{des}})^\top$, as well as $X_{n1}\overset{\mathcal{D}}{=}X_T=X_{\text{des}}+X_{\text{ides}}$. We first show how to generate a random permutation of size n and to calculate its number of inversions. For the former, we can use the base function `sample`, while for the latter, we use an efficient recursive algorithm based on the classical *MergeSort* algorithm.

```
xinv <- function(n) {
  arr = sample(n)
  inv = count_inversions(arr)
  return(inv[[2]])
}
```

```
count_inversions <- function(arr) {
  n <- length(arr)
  if (length(arr) <= 1) {
    return(list(arr, 0))
  }
  mid <- length(arr) %/% 2
  left <- count_inversions(arr[1:mid]); right <- count_inversions([(mid + 1):n])
  sort_left <- left[[1]]; sort_right <- right[[1]]
  inv_left <- left[[2]]; inv_right <- right[[2]]
  merged <- numeric(n)
  i <- 1; j <- 1; inv_merge <- 0
  for (k in 1:n) {
    if (i > length(sort_left)) {
      merged[k] <- sort_right[j]
      j <- j + 1
    } else if (j > length(sort_right)) {
      merged[k] <- sort_left[i]
      i <- i + 1
    } else if (sort_left[i] <= sort_right[j]) {
      merged[k] <- sort_left[i]
      i <- i + 1
    } else {
      merged[k] <- sort_right[j]
      j <- j + 1
      inv_merge <- inv_merge + (length(sort_left) - i + 1)
    }
  }
  inversions <- inv_left + inv_right + inv_merge
  return(list(merged, inversions))
}
```

Counting descents and inverse descents is straightforward.

```
xdes <- function(n) {
  arr = sample(n)
  c <- 0
  for (i in 1:(n-1)) {
    if (arr[i] > arr[i+1]) {
      c <- c + 1
    }
  }
  return(c)
}
```

```
xt <- function(n) {
  arr = sample(n)
  invarr <- numeric(n)
  for (i in 1:n) {
    invarr[arr[i]] <- i
  }
  c <- 0
  for (i in 1:(n-1)) {
    if (arr[i] > arr[i+1]) {
      c <- c + 1
    }
    if (invarr[i] > invarr[i+1]) {
      c <- c + 1
    }
  }
  return(c)
}
```

Next, for the simulation on the joint distribution $(X_{\text{inv}}, X_{\text{des}})^\top$, we generate the n-th row of the triangular array and rescale its maximum with the transformation constants a_n, b_n as introduced in Theorem 4.4.1. These consist of the constants $\alpha_n, \beta_n, \boldsymbol{\alpha}_n, \boldsymbol{\beta}_n$ as introduced in Theorems 1.1.11 and 1.3.10, as well as the mean and variance.

```
rowmax <- function(n) {
  inv <- numeric(n); des <- numeric(n)
  for (i in 1:n) {
    inv[i] <- xinv(n); des[i] <- xdes(n)
  }
  maxx <- c(max(inv), max(des))
  alpha <- 1/sqrt(2*log(n))
  beta <- 1/alpha - alpha/2 * log(4*pi*log(n))
  muinv <- n*(n-1)/4; mudes <- (n-1)/2
  mu_n <- c(muinv, mudes)
  varinv <- n^3/36 + n^2/24 - 5*n/72; vardes <- (n+1)/12
  s_n <- c(sqrt(varinv), sqrt(vardes))
  a_n <- alpha * s_n; b_n <- beta * s_n + mu_n
  result <- (maxx - b_n)/a_n
  return(result)
}
```

Accordingly, we use the same mechanism for the simulation on X_T. The only difference is in the mean and variance.

```
rowmax_xt <- function(n) {
  t <- numeric(n)
  for (i in 1:n) {
    t[i] <- xt(n)
  }
  alpha <- 1/sqrt(2*log(n))
  beta <- 1/alpha - alpha/2 * log(4*pi*log(n))
  mu_n <- n-1; var_n <- (n+1)/6 + (n-1)/n
  s_n <- sqrt(var_n)
  a_n <- alpha * s_n; b_n <- beta * s_n + mu_n
  result <- (max(t) - b_n)/a_n
  return(result)
}
```

We decided to create 10,000 replications of $(M_n - b_n)/a_n$. For any selected size n of the underlying symmetric group, we generate these replications and plot their empirical distribution. In the simulation, we chose $n \in \{20, 50, 100, 200, 500, 1000\}$. To obtain the empirical distribution, we require the function `empirical_cdf` from the `mltools` package.

```
simulation <- function(n) {
  all <- matrix(nrow=10000, ncol=2)
  for (i in 1:10000) {
    all[i,] <- rowmax(n)
  }
  dt <- data.table(x=all[,1], y=all[,2])
  a <- seq(-4, 4, length.out = 100); b <- seq(-4, 4, length.out = 100)
  windows(width = 12, height = 12)
  ecdf <- outer(a, b, FUN = function(a, b) empirical_cdf(dt, ubounds = data.table(x=a,y=b))$CDF)
  persp(x, y, ecdf, ticktype="detailed", theta = 45, phi = 0, col = "yellow",
        border = "orange", shade = 0.5)
}
```

The plot of the two-sided Eulerian statistic X_T is slightly easier to implement, since it is one-dimensional. As seen in Figure 4.5, we also added a direct comparison to the Gumbel distribution function. The simulation for X_T was executed only for $n = 1000$.

```
simulation_xt <- function(n) {
  all <- numeric(10000)
  for (i in 1:10000) {
    all[i] <- rowmax_xt(n)
  }
  a <- seq(-4, 4, length.out = 100)
  windows(width = 12, height = 12)
  y <- empirical_cdf(all,a)$CDF
  gumbel <- exp(-exp(-a))
  plot(a, gumbel, type="l", col="blue", lwd=2)
  lines(a,y, type="l", col="red", lwd=3, add=TRUE)
}
```

5 Generalized inversions and descents

In this chapter, we introduce a generalization of inversions and descents on classical Weyl groups. Recall that on the symmetric group S_n, the uniform random numbers of inversions and descents are given by

$$X_{\text{inv}} = \sum_{1 \le i < j \le n} \mathbf{1}\{Z_i > Z_j\}, \qquad X_{\text{des}} = \sum_{i=1}^{n-1} \mathbf{1}\{Z_i > Z_{i+1}\}, \tag{5.1}$$

with independent $Z_1, \dots, Z_n \sim U(0,1)$, see (2.2). A class of generalized inversion statistics $X_{\text{inv}}^{(d)}$ can be constructed by restricting the left-hand sum to pairs (i,j) with $1 \le j - i \le d$, for some $d \in \{1, \dots, n-1\}$. These generalized inversions were first introduced by de Mari & Shayman [32] who used this concept to describe the Betti numbers of Hessenberg subvarieties in regular complex-valued matrices. Likewise, the number of descents X_{des} can be generalized to $X_{\text{des}}^{(d)}$ by counting all $i \in \{1, \dots, n-d\}$ with $Z_i > Z_{i+d}$. Even further, both classes $X_{\text{inv}}^{(d)}, X_{\text{des}}^{(d)}$ can be extended to the other classical Weyl groups B_n and D_n. These concepts were introduced by Meier & Stump [75], who also showed a CLT for both generalized inversions and generalized descents.

We now aim to extend the knowledge gained in the previous two chapters to these new classes of permutation statistics, i.e., we aim to prove Gumbel attraction for both the individual statistics $X_{\text{inv}}^{(d_1)}, X_{\text{des}}^{(d_2)}$ and the joint statistic $(X_{\text{inv}}^{(d_1)}, X_{\text{des}}^{(d_2)})^\top$. Here, d_1 and d_2 can be either fixed or dependent on n. We will investigate the impact of the choice of d for these results. Since there is no closed representation of the generating function of $X_{\text{inv}}^{(d)}$ and $X_{\text{des}}^{(d)}$, we will again deal with the dependency structure by using Hájek projections and high-dimensional Gaussian approximation.

Section 5.1 gives basic definitions and properties of generalized inversions and descents. Section 5.2 is devoted to determining the choices of d for which the Hájek approximation of $\hat{X}_{\text{inv}}^{(d)}$ is successful. In Section 5.3, we deduce the bivariate CLT and the extreme value limit theorems for generalized inversions and descents. Section 5.4 gives the technical proofs of the lemmas in Section 5.2.

5.1 Basic definitions

To simplify notation, we write $\mathfrak{N}_{n,d} := \{(i,j) \in \{1, \dots, n\}^2 \mid 1 \le j - i \le d\}$.

Definition 5.1.1. Let S_n be a symmetric group, let $\pi \in S_n$ and $d \in \{1, \dots, n-1\}$. Then, *d-inversions* are all pairs $(i,j) \in \mathfrak{N}_{n,d}$ with $\pi(i) > \pi(j)$, and *d-descents* are all numbers i with $i \le n - d$ and $\pi(i) > \pi(i+d)$. In this sense, common descents are 1-inversions or 1-descents, and common inversions are $(n-1)$-inversions. Drawing $\pi \in S_n$ uniformly at random, we write $X_{\text{inv}}^{(d)}$ for the number of d-inversions and $X_{\text{des}}^{(d)}$ for the number of

d-descents. By analogy with (2.2), these are expressed as

$$X^{(d)}_{\text{inv}} = \sum_{(i,j)\in\mathfrak{N}_{n,d}} \mathbf{1}\{Z_i > Z_j\}\,, \qquad\qquad X^{(d)}_{\text{des}} = \sum_{i=1}^{n-d} \mathbf{1}\{Z_i > Z_{i+d}\} \tag{5.2}$$

for i.i.d. $Z_1, \ldots, Z_n \sim U(0,1)$. The terms *generalized inversions* and *generalized descents* are umbrella terms for all d-inversions and d-descents, respectively. △

In the literature, there are different terminologies, e.g., in [9, 87], d-inversions as given in Definition 5.1.1 are called d-descents. However, we use the terms and notation provided in [75] throughout this chapter to avoid confusion.

Remark 5.1.2. Obviously, each $k \in \{1, \ldots, n\}$ can be involved in at most $2d$ d-inversions. This bound is redundant if $d > n/2$. In fact, it is an important case distinction whether $d \leq n/2$ or $d > n/2$. In the case of $d \leq n/2$, we split $\{1, \ldots, n\}$ into the subregions

$$K_1 := \{1, \ldots, d\}\,, \qquad K_2 := \{d+1, \ldots, n-d\}\,, \qquad K_3 := \{n-d+1, \ldots, n\}\,,$$

where $K_2 = \emptyset$ if $d = n/2$. For any $k \in K_2$, *all* larger indices $k+1, \ldots, k+d$ and *all* smaller indices $k-1, \ldots, k-d$ allow to form d-inversions. For any $k \notin K_2$, there are *less* than d indices available in one direction, which we call an *overlap*. If $k \in K_1$, then only $k-1 < d$ smaller indices are available for d-inversions. This is a *left-sided overlap*. If $k > n-d$ is large, then there are only $n-k < d$ larger indices available. This is a *right-sided overlap*. The following Figures 5.1 – 5.3 give illustrations, assuming $d < n/2$.

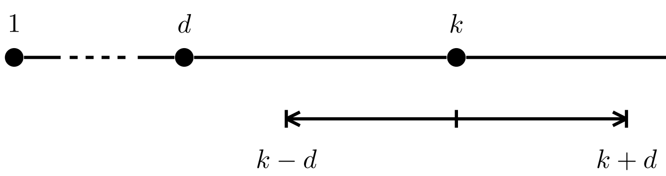


Figure 5.1: A central index $k \in K_2$ for which the whole segment $\{k-d, \ldots, k+d\}$ is contained in $\{1, \ldots, n\}$.

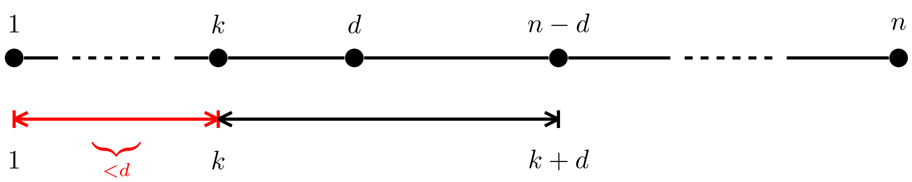


Figure 5.2: A left-positioned index $k \in K_2$ with left-sided overlap, due to $k-d \notin \{1, \ldots, d\}$. The overlap is indicated by the red segment.

Remark 5.1.3. On the contrary, if $d > n/2$, then $n-d < d$ and each index has an overlap to at least one side, and the above partition into three subregions is now written as

$$K_1 := \{1, \ldots, n-d\}\,, \qquad K_2 := \{n-d+1, \ldots, d\}\,, \qquad K_3 := \{d+1, \ldots, n\}\,.$$

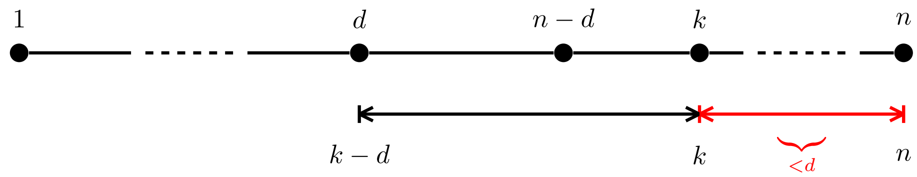


Figure 5.3: A right-positioned index $k \in K_3$ with right-sided overlap, due to $k + d \notin \{1, \dots, n\}$. The overlap is indicated by the red segment.

The indices in K_2 produce a *two-sided overlap*. For simplicity, we call the case of $d \leq n/2$ the *short case* and the case of $d > n/2$ the *long case*. △

The mean and variance of $X_{\mathrm{inv}}^{(d)}$ have been extensively computed by Pike [87]. It is easy to verify that the special cases $d = 1$ and $d = n - 1$ are consistent with the results for common inversions and descents in Theorems 2.2.2 and 2.2.3.

Theorem 5.1.4. (see [87], Theorem 1)
For all $d = 1, \dots, n-1$, it holds that

$$\mathbb{E}\left(X_{\mathrm{inv}}^{(d)}\right) = \frac{2nd - d^2 - d}{4}.$$

Moreover, if $d \leq n/2$, then

$$\mathrm{Var}\left(X_{\mathrm{inv}}^{(d)}\right) = \frac{6nd + 4d^3 + 3d^2 - d}{72}.$$

If $d > n/2$, then

$$\mathrm{Var}\left(X_{\mathrm{inv}}^{(d)}\right) = -\frac{1}{6}d^3 + \left(\frac{1}{3}n - \frac{7}{24}\right)d^2 - \left(\frac{1}{6}n^2 - \frac{5}{12}n + \frac{1}{8}\right)d + \frac{1}{36}n^3 - \frac{1}{12}n^2 + \frac{1}{18}n.$$

The calculation of $\mathrm{Var}\left(X_{\mathrm{inv}}^{(d)}\right)$ is reviewed in [75, Theorem A.1], where the variance of d-descents is provided as well.

Theorem 5.1.5. (see [75], Theorem A.1)
The mean and variance of generalized descents are given as $\mathbb{E}\left(X_{\mathrm{des}}^{(d)}\right) = (n-d)/2$ and

$$\mathrm{Var}\left(X_{\mathrm{des}}^{(d)}\right) = \begin{cases} (n+d)/12, & d \leq n/2 \\ (n-d)/4, & d > n/2 \end{cases}.$$

We now discuss the extension of generalized inversions and descents from symmetric groups to the other classical Weyl groups B_n and D_n. This extension was introduced by Meier & Stump [75] and it is based on the *root poset* of a classical Weyl group. We refer to [75, Section 2] for the details. On the symmetric group S_n, the ordered pairs of indices (i, j) correspond to the positive roots $[ij] := e_i - e_j$, where e_i, e_j are unit vectors in $\mathbb{R}^n$, and the height of $[ij]$ within the root poset is $\mathrm{ht}([ij]) = j - i$.

On the signed permutation group B_n, we also have to consider the positive roots $[\widetilde{ij}] := e_i + e_j$ and $[i] := e_i$ for $1 \leq i < j \leq n$. The heights of these additional roots are

$\mathrm{ht}([\widetilde{ij}]) = i+j$ and $\mathrm{ht}([i]) = i$. The root $[\widetilde{ij}]$ corresponds to the indicator $\mathbf{1}\{-\pi(i) > \pi(j)\}$ appearing in Remark 2.1.15, while $[i]$ corresponds to $\mathbf{1}\{\pi(i) < 0\}$. On the even-signed permutation group D_n, the roots $[i]$ are disregarded, and $[\widetilde{ij}]$ has height $i+j-2$. Hasse diagrams of the root posets of B_n and D_n are illustrated by the examples B_5 and D_6 in Figure 5.4.

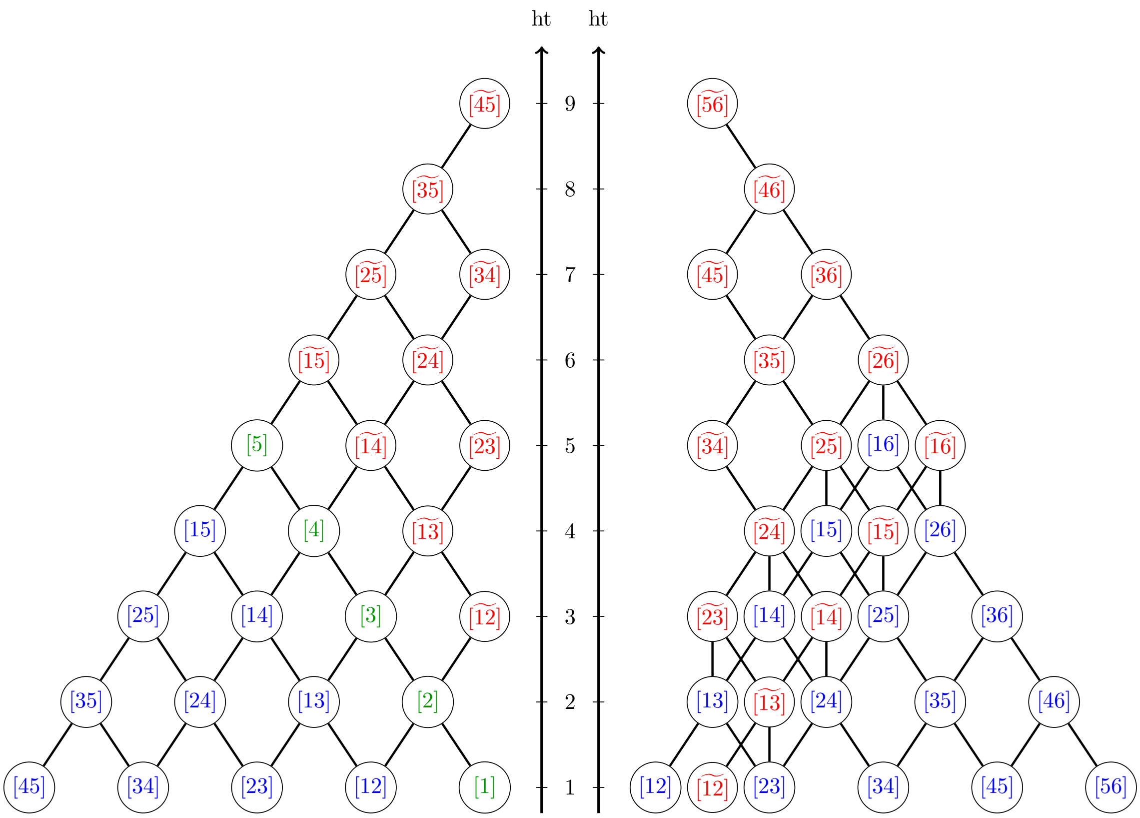


Figure 5.4: Hasse diagrams of the root posets of the signed permutation group B_5 (left) and the even-signed permutation group D_6 (right). The roots $[ij]$ are marked blue, the roots $[\widetilde{ij}]$ are marked red and the roots $[i]$ are marked green. The vertical axis indicates the heights of all positive roots.

Definition 5.1.6. For any classical Weyl group, d-inversions are determined by roots of height at most d, and d-descents are determined by roots of height exactly d, see [75, Definition 2.4]. For symmetric groups, this coincides with Definition 5.1.1. In addition to $\mathfrak{N}_{n,d}$, we introduce

$$\widetilde{\mathfrak{N}}_{n,d} := \{(i,j) \in \{1,\dots,n\}^2 \mid i<j,\ i+j \le d\}.$$

Then, on the signed and even-signed permutation groups, $X_{\mathrm{inv}}^{(d)}$ and $X_{\mathrm{des}}^{(d)}$ can be expressed as follows:

$$X_{\mathrm{inv}}^{B,(d)} = \sum_{(i,j)\in\mathfrak{N}_{n,d}} \mathbf{1}\{Z_i > Z_j\} + \sum_{(i,j)\in\widetilde{\mathfrak{N}}_{n,d}} \mathbf{1}\{-Z_i > Z_j\} + \sum_{i=1}^{n\wedge d} \mathbf{1}\{Z_i < 0\}\,, \tag{5.3a}$$

$$X_{\mathrm{inv}}^{D,(d)} = \sum_{(i,j)\in\mathfrak{N}_{n,d}} \mathbf{1}\{Z_i > Z_j\} + \sum_{(i,j)\in\widetilde{\mathfrak{N}}_{n,d+2}} \mathbf{1}\{-Z_i > Z_j\}\,, \tag{5.3b}$$

$$X_{\mathrm{des}}^{B,(d)} = \sum_{i=1}^{n-d} \mathbf{1}\{Z_i > Z_{i+d}\} + \sum_{i=1}^{\lceil d/2 \rceil - 1} \mathbf{1}\{-Z_i > Z_{d-i}\} + \mathbf{1}\{Z_d < 0\}\,, \tag{5.3c}$$

$$X_{\mathrm{des}}^{D,(d)} = \sum_{i=1}^{n-d} \mathbf{1}\{Z_i > Z_{i+d}\} + \sum_{i=1}^{\lceil d/2 \rceil} \mathbf{1}\{-Z_i > Z_{d+2-i}\}\,. \tag{5.3d}$$

Note that in (5.3c) and (5.3d), indicators are ignored if they involve indices out of bounds. The largest possible choice of d equals the total height of the root poset, namely, $d_{\max}-1$, where $d_{\max}$ denotes the largest degree of the underlying classical Weyl group, see Remark 2.1.18. To precisely compute the variance of $X_{\mathrm{inv}}^{(d)}$ and $X_{\mathrm{des}}^{(d)}$ on the groups B_n and D_n, one needs to distinguish eight cases, as seen in [75, Theorems A.4 and A.13]. However, many of these cases give the same asymptotic quantification, which can be stated as follows:

Lemma 5.1.7. (cf. [75], Theorems A.4 and A.13)
For the generalized inversions and descents on both the groups B_n and D_n, it holds that

$$\mathrm{Var}\left(X_{\mathrm{inv}}^{(d)}\right) = \begin{cases} \frac{1}{36}d^3 + \frac{1}{12}nd + O(d^2), & d \le n/2 \\ \frac{1}{36}d^3 + O(d^2), & n/2 \le d < n\,, \\ -\frac{1}{12}d^3 + \frac{1}{3}nd^2 - \frac{1}{3}n^2 d + \frac{1}{9}n^3 + O(d^2), & d \ge n \end{cases}$$

$$\mathrm{Var}\left(X_{\mathrm{des}}^{(d)}\right) = \begin{cases} \frac{1}{24}d + \frac{1}{12}n + O(1), & d < n \\ -\frac{1}{8}d + \frac{1}{4}n + O(1), & d \ge n \end{cases}.$$

For generalized inversions and descents on all classical Weyl groups, the CLT was investigated by Meier & Stump [75]. Here, it is essential that the number of available pairs to form d-inversions or d-descents continues to grow as $n \to \infty$. For a classical Weyl group W of rank n, let $N_{n,d}^{\le}$ denote the number of positive roots with height at most d, and let $N_{n,d}^{=}$ denote the number of positive roots with height exactly d. In particular, $N_{n,d}^{\le} = |\mathfrak{N}_{n,d}|$ if $W = S_n$, $N_{n,d}^{\le} = |\mathfrak{N}_{n,d}| + |\widetilde{\mathfrak{N}}_{n,d+2}|$ if $W = D_n$, and $N_{n,d}^{\le} = |\mathfrak{N}_{n,d}| + |\widetilde{\mathfrak{N}}_{n,d}| + d$ if $W = B_n$.

Theorem 5.1.8. (see [75], Corollary 2.7 and Theorem 2.9)
Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of classical Weyl groups and let $d = (d_n)_{n\in\mathbb{N}}$ be a sequence of natural numbers with $1 \le d \le d_{\max}-1$. Let $X_{\mathrm{inv}}^{(d)}$ and $X_{\mathrm{des}}^{(d)}$ be the statistics of d-inversions and d-descents on W_n, respectively. Then:

(a) *If $N_{n,d}^{\le} \longrightarrow \infty$, then $X_{\mathrm{inv}}^{(d)}$ satisfies the CLT.*
(b) *If $N_{n,d}^{=} \longrightarrow \infty$, then $X_{\mathrm{des}}^{(d)}$ satisfies the CLT.*

5.2 The Hájek approximation of generalized inversions

In the following process, we investigate whether the previously developed methods are applicable to the individual statistics $X_{\mathrm{inv}}^{(d_1)}, X_{\mathrm{des}}^{(d_2)}$ and the joint statistic $(X_{\mathrm{inv}}^{(d_1)}, X_{\mathrm{des}}^{(d_2)})^\top$, where d_1, d_2 are either fixed or dependent on n. To this end, we first investigate the Hájek projection $\hat{X}_{\mathrm{inv}}^{(d)}$ in detail, and see for which d it satisfies the condition of Theorem 4.1.2. We first consider $\hat{X}_{\mathrm{inv}}^{(d)}$ on symmetric groups and then provide the analogous observations on the other classical Weyl groups.

For common descents, we have seen in Remark 4.1.4 that for nearly every Z_k, there is exactly one index $k+1$ giving one contribution of Z_k to $\mathbb{E}(X_{\mathrm{des}} \mid Z_k)$, and exactly one index $k-1$ giving one contribution of $1-Z_k$, yielding cancellation of the Z_k. The boundary cases are insufficient to produce a linear variance as required. If d is fixed, this problem persists. As described in Remark 5.1.2, the boundaries produce an overlap that helps to avoid complete cancellation and keep some contributions of Z_k. Figure 5.5 illustrates this for $d = n/3$. In conclusion, we need $d = d_n$ sufficiently large so that the overlap gives enough contribution to $\mathrm{Var}(\hat{X}_{\mathrm{inv}}^{(d)})$ in relation to $\mathrm{Var}(X_{\mathrm{inv}}^{(d)})$.

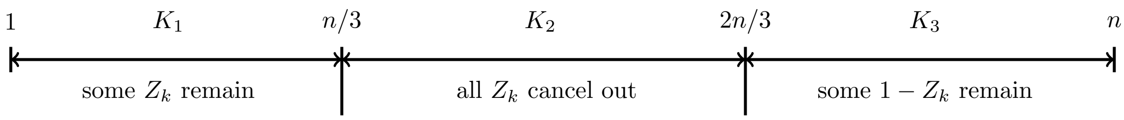


Figure 5.5: Overview of cancellation and remainders of Z_k and $1-Z_k$ in the three subregions K_1, K_2, K_3.

In what follows, we analyze the short case of $d \le n/2$. By definition,

$$\hat{X}_{\mathrm{inv}}^{(d)} = \sum_{k=1}^{n} \mathbb{E}\left(X_{\mathrm{inv}}^{(d)} \mid Z_k\right) - (n-1)\mathbb{E}\left(X_{\mathrm{inv}}^{(d)}\right)$$

and

$$\mathbb{E}\left(X_{\mathrm{inv}}^{(d)} \mid Z_k\right) = \sum_{(i,j)\in\mathfrak{N}_{n,d}} \mathbb{P}(Z_i > Z_j \mid Z_k) = \sum_{(i,j)\in\mathfrak{N}_{n,d}} \begin{cases} 1/2, & k \notin \{i,j\} \\ Z_k, & k = i \\ 1-Z_k, & k = j \end{cases} . \tag{5.4}$$

As already noted in the proof of Lemma 4.1.3, only the pairs (i,j) with $k \in \{i,j\}$ contribute to $\mathrm{Var}\left(\mathbb{E}\left(X_{\mathrm{inv}}^{(d)} \mid Z_k\right)\right)$. We call these contributions the *non-trivial parts* for simplicity. The number of these pairs depends on whether k belongs to K_1, K_2, or K_3. Figure 5.6 visualizes this case distinction for the exemplary choice of $n = 15$ and $d = 4$.
If $k \in K_2$, then the non-trivial parts are

$$\underbrace{Z_k + Z_k + \ldots + Z_k}_{d \text{ times}} + \underbrace{(1-Z_k) + (1-Z_k) + \ldots + (1-Z_k)}_{d \text{ times}} = d\,.$$

This means $\mathbb{E}\left(X_{\mathrm{inv}}^{(d)} \mid Z_k\right)$ is constant due to cancellation, and vanishes when computing the variance. So, in the short case, $\mathrm{Var}\left(\mathbb{E}\left(X_{\mathrm{inv}}^{(d)} \mid Z_k\right)\right)$ originates only from K_1 and K_3.
If $k \in K_1$, then $\mathbb{E}\left(X_{\mathrm{inv}}^{(d)} \mid Z_k\right) = (k-1)(1-Z_k) + dZ_k + \mathrm{const} = (d+1-k)Z_k + \mathrm{const}$.
If $k \in K_3$, then $\mathbb{E}\left(X_{\mathrm{inv}}^{(d)} \mid Z_k\right) = (n-k)Z_k + d(1-Z_k) + \mathrm{const} = (n-k-d)Z_k + \mathrm{const}$.
So, the overall representation of $\hat{X}_{\mathrm{inv}}^{(d)}$ in the short case is

$$\hat{X}_{\mathrm{inv}}^{(d)} = \sum_{k=1}^{n} \omega_d(k) Z_k + \mathrm{const}\,, \qquad \text{with} \quad \omega_d(k) := \begin{cases} d-k+1\,, & k \in K_1 \\ 0, & k \in K_2 \\ n-d-k\,, & k \in K_3 \end{cases} . \tag{5.5}$$

Figure 5.7 illustrates the coefficients $\omega_d(k)$ appearing in (5.5).

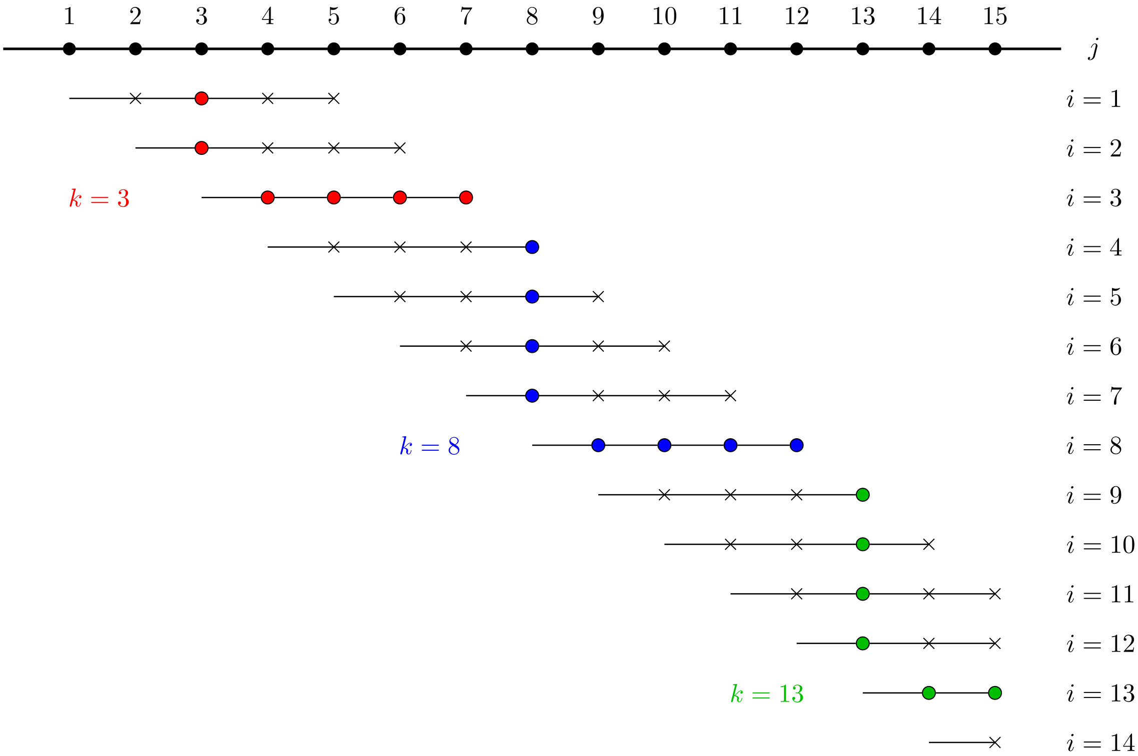


Figure 5.6: Overview of relevant pairs (i, j) for computing the variance of $\mathbb{E}\left(X_{\text{inv}}^{(d)} \mid Z_k\right)$, where $n = 15$ and $d = 4$. For each of the subregions K_1, K_2, K_3, an exemplary index k is chosen, and the pairs that give a non-trivial contribution are highlighted in red for $k \in K_1$, in blue for $k \in K_2$, and in green for $k \in K_3$.

Lemma 5.2.1. (see Section 5.4.1 for the proof)
In the short case,

$$\operatorname{Var}\left(\hat{X}_{\text{inv}}^{(d)}\right) = \frac{1}{72}\left(4d^3 + 6d^2 + 2d\right).$$

So, if $d \leq n/2$ and $d = \omega(n^{1/2})$, then $\operatorname{Var}\left(\hat{X}_{\text{inv}}^{(d)}\right) \sim \operatorname{Var}\left(X_{\text{inv}}^{(d)}\right)$.

Now, we consider the long case. Due to $d > n/2$, we now have $d > n-d$, and the subregions K_1, K_2, K_3 are redefined according to Remark 5.1.3. For the non-trivial contributions, we note that:

- If $k \in K_1$ or $k \in K_3$, then the remainder is the same as in the short case.
- If $k \in K_2$, then the two-sided overlap gives the same contribution as stated in the proof of Lemma 4.1.3, namely, $(n - 2k + 1)Z_k + (k - 1)$.

So, we again obtain a representation in the way of

$$\hat{X}_{\text{inv}}^{(d)} = \sum_{k=1}^{n} \omega_d(k) Z_k + \text{const}, \qquad \text{with} \quad \omega_d(k) := \begin{cases} d - k + 1, & k \in K_1 \\ n - 2k + 1, & k \in K_2 \\ n - d - k, & k \in K_3 \end{cases}. \tag{5.6}$$

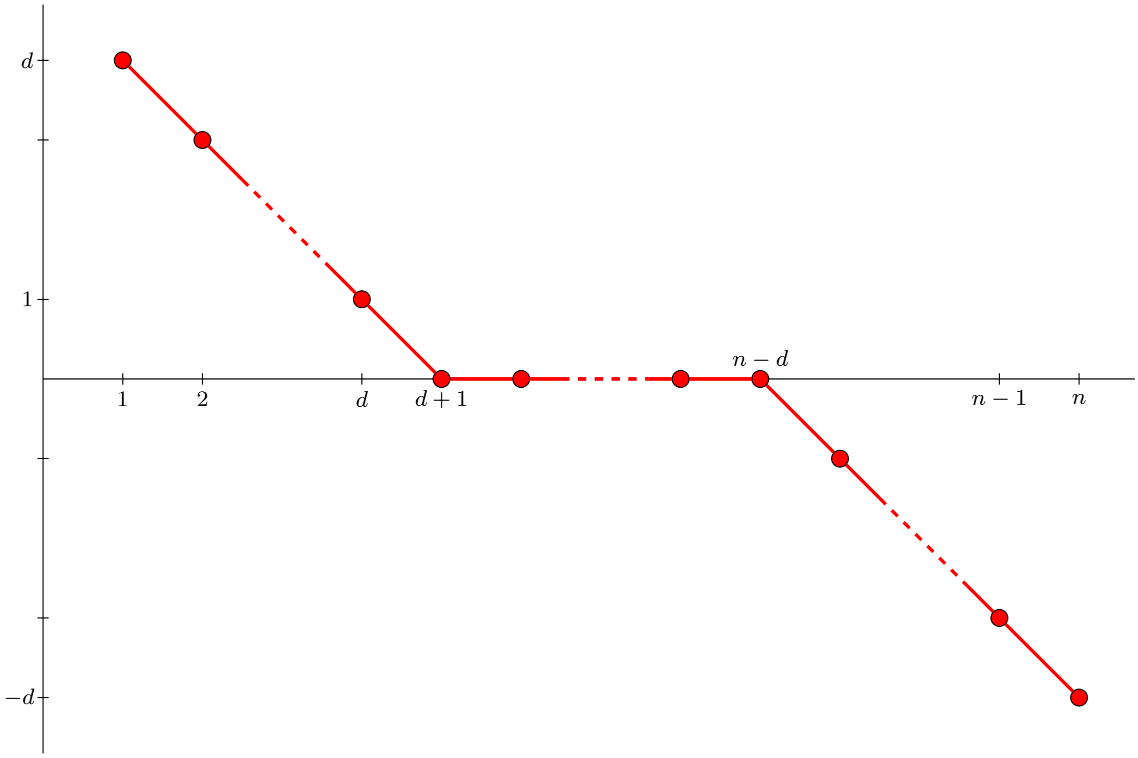


Figure 5.7: Plot of $\omega_d(k)$ in the short case ($d \leq n/2$).

Figure 5.8 illustrates the coefficients $\omega_d(k)$ in the long case. It is seen that the plot of Figure 5.7 smoothly transitions into that of Figure 5.8 as d transitions from the short case into the long case. Furthermore, in the case of common inversions, i.e., $d = n - 1$, Figure 5.8 coincides with Figure 4.2.

In the following lemma, we show that in the long case, the leading terms of $\mathrm{Var}\left(X_{\mathrm{inv}}^{(d)}\right)$ and $\mathrm{Var}\left(\hat{X}_{\mathrm{inv}}^{(d)}\right)$ are always matching.

Lemma 5.2.2. (see Section 5.4.2 for the proof)
In the long case,

$$\mathrm{Var}\left(\hat{X}_{\mathrm{inv}}^{(d)}\right) = -\frac{1}{6}d^3 + \left(\frac{1}{3}n - \frac{1}{4}\right)d^2 - \left(\frac{1}{6}n^2 - \frac{1}{3}n + \frac{1}{12}\right)d + \frac{1}{36}n^3 - \frac{1}{12}n^2 + \frac{1}{18}n\,.$$

Therefore, if $d > n/2$, we always have $\mathrm{Var}\left(\hat{X}_{\mathrm{inv}}^{(d)}\right) \sim \mathrm{Var}\left(X_{\mathrm{inv}}^{(d)}\right)$.

Combining these two observations, we can state:

Corollary 5.2.3. *Consider the generalized inversion statistic $X_{\mathrm{inv}}^{(d)}$ on the symmetric groups $(S_n)_{n\in\mathbb{N}}$, with $d = d_n$ being a sequence satisfying $1 \leq d \leq n-1$ $\forall n \in \mathbb{N}$. Then,* $\mathrm{Var}\left(\hat{X}_{\mathrm{inv}}^{(d)}\right) \sim \mathrm{Var}\left(X_{\mathrm{inv}}^{(d)}\right)$ *holds if and only if $d = \omega(n^{1/2})$.*

We now derive an analogous result for the other classical Weyl groups B_n and D_n. Recall the root poset structure explained in Section 5.1, the representations in Definition 5.1.6, and the asymptotic quantification of $\mathrm{Var}(X_{\mathrm{inv}}^{(d)})$ given in Lemma 5.1.7.

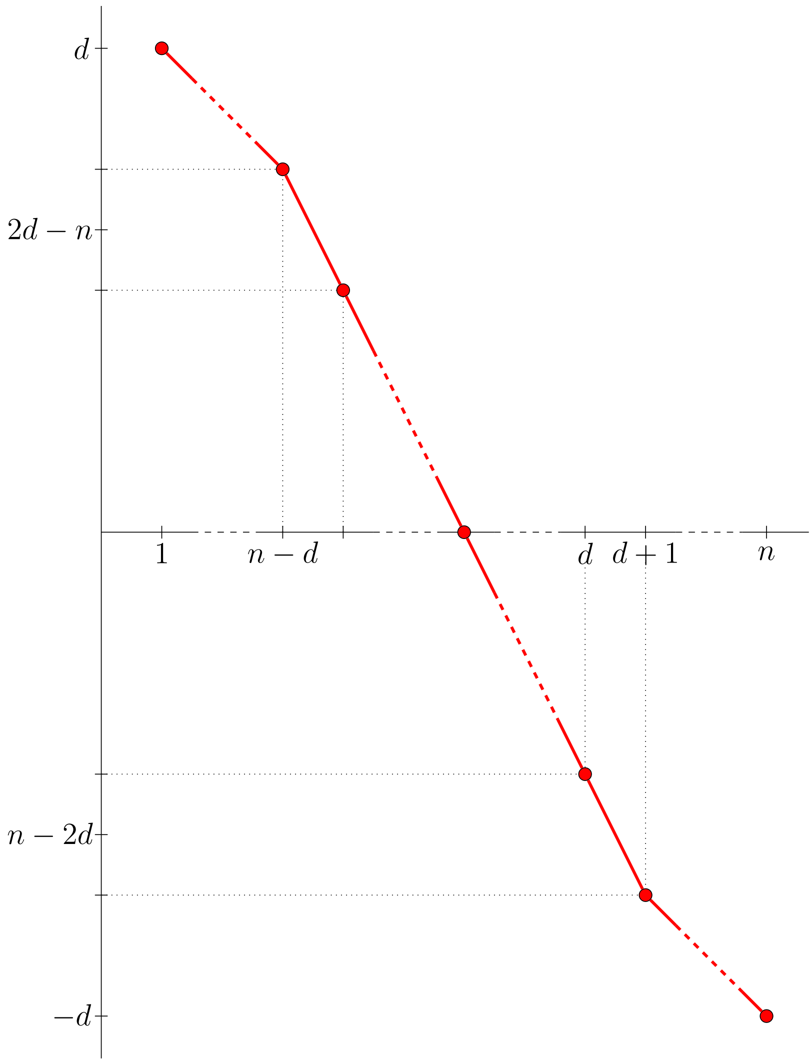


Figure 5.8: Plot of $\omega_d(k)$ in the long case ($d > n/2$).

For simplicity reasons, we ignore any p-bias and consider only the uniform distribution on B_n and D_n. So, we recall that $X_{\text{inv}}^{B,(d)}$ is based on i.i.d. $Z_1, \dots, Z_n \sim U(-1,1)$. For each $k = 1, \dots, n$, we have

$$\mathbb{E}\left(X_{\text{inv}}^{B,(d)} \mid Z_k\right) = \sum_{(i,j)\in\mathfrak{N}_{n,d}} \mathbb{P}(Z_i > Z_j \mid Z_k) + \sum_{(i,j)\in\widetilde{\mathfrak{N}}_{n,d}} \mathbb{P}(-Z_i > Z_j \mid Z_k)$$
$$+ \sum_{i=1}^{n\wedge d} \mathbb{P}(Z_i < 0 \mid Z_k)\,.$$

Similar to the symmetric groups, we will compute coefficients $\omega_d(k)$ such that

$$\sum_{(i,j)\in\mathfrak{N}_{n,d}} \mathbb{P}(Z_i > Z_j \mid Z_k) + \sum_{(i,j)\in\widetilde{\mathfrak{N}}_{n,d}} \mathbb{P}(-Z_i > Z_j \mid Z_k) = \omega_d(k) Z_k + \text{const}\,.$$

If $d < k$, then the third sum $\sum_{i=1}^{n\wedge d} \mathbb{P}(Z_i < 0 \mid Z_k) = d/2$ is constant. Otherwise, we have

$$\begin{aligned}\operatorname{Var}\left(\mathbb{E}\left(X_{\text{inv}}^{B,(d)} \mid Z_k\right)\right) &= \operatorname{Var}\Big(\omega_d(k)Z_k + \mathbf{1}\{Z_k<0\} + \text{const}\Big)\\ &= \operatorname{Var}(\omega_d(k)Z_k) + \operatorname{Var}(\mathbf{1}\{Z_k<0\}) + 2\operatorname{Cov}(\omega_d(k)Z_k, \mathbf{1}\{Z_k<0\})\\ &= \frac{\omega_d(k)^2}{3} + \frac{1}{4} + 2\omega_d(k)\underbrace{\operatorname{Cov}(Z_k,\mathbf{1}\{Z_k<0\})}_{=\,-1/4}, \end{aligned} \tag{5.7}$$

from which we see that even if $d<k$, the leading terms of $\operatorname{Var}\left(\hat{X}_{\text{inv}}^{B,(d)}\right)$ are not influenced by the third sum. So, we have to determine the coefficients $\omega_d(k)$, from which we successfully obtain a statement analogous to Corollary 5.2.3:

Lemma 5.2.4. **(see Section 5.4.3 for the proof)**
On the signed permutation groups B_n, let $d=d_n$ be a sequence with $1\le d\le 2n-1$ $\forall n\in\mathbb{N}$. Then, $\operatorname{Var}\left(\hat{X}_{\text{inv}}^{(d)}\right)\sim\operatorname{Var}\left(X_{\text{inv}}^{(d)}\right)$ holds if and only if $d=\omega(n^{1/2})$.

This statement also extends to the even-signed permutation groups D_n, since the difference between $X_{\text{inv}}^{B,(d)}$ and $X_{\text{inv}}^{D,(d)}$ is asymptotically negligible (cf. (5.3a) and (5.3b)).

5.3 Asymptotic results

We now consider the joint statistic $\left(X_{\text{inv}}^{(d_1)}, X_{\text{des}}^{(d_2)}\right)^\top$ for two sequences $d_1=d_1(n)$, $d_2=d_2(n)$ with $d_1(n), d_2(n)\le d_{\max}-1$ $\forall n\in\mathbb{N}$. We first address the asymptotic normality of $\left(X_{\text{inv}}^{(d_1)}, X_{\text{des}}^{(d_2)}\right)^\top$, and then state the extreme value behavior of $X_{\text{inv}}^{(d_1)}$, $X_{\text{des}}^{(d_2)}$ and $\left(X_{\text{inv}}^{(d_1)}, X_{\text{des}}^{(d_2)}\right)^\top$. In this process, we keep using d as an umbrella notation for d_1 or d_2, depending on the context.

If d_1 and d_2 both remain fixed, then both $X_{\text{inv}}^{(d)}$ and $X_{\text{des}}^{(d)}$ are m-dependent ($m=d$) and it is not necessary to apply the Hájek approximation. From the CLT for m-dependent random vectors, it follows that:

Theorem 5.3.1. *For any two fixed numbers d_1, d_2, the joint distribution $\left(X_{\text{inv}}^{(d_1)}, X_{\text{des}}^{(d_2)}\right)^\top$ satisfies the CLT.*

In what follows, we assume that both $d_1=d_1(n)$ and $d_2=d_2(n)$ diverge. Recall the notations $N_{n,d}^{\le}, N_{n,d}^{=}$ from Theorem 5.1.8. To obtain bivariate asymptotic normality, it is obviously necessary that both $N_{n,d}^{\le}\longrightarrow\infty$ and $N_{n,d}^{=}\longrightarrow\infty$, in analogy to Theorem 5.1.8. These conditions are also necessary to allow for non-degeneracy of extreme values. To state a Gumbel EVLT for $X_{\text{inv}}^{(d_1)}$, we only require $N_{n,d_1}^{\le}\longrightarrow\infty$, which is already ensured as d_1 is divergent. In contrast, the divergence of $N_{n,d_2}^{=}$ is ensured if and only if d_2 is not too large, or precisely, if $d_2=d_{\max}-\omega(1)$.

Since $d_2=d_2(n)$ diverges, there exists no constant $m\in\mathbb{N}$ for which all $X_{\text{des}}^{(d_2)}$ are m-dependent. However, the dependency structure of $X_{\text{des}}^{(d_2)}$ is still sparse. According to (5.2), (5.3c), (5.3d), we can represent $X_{\text{des}}^{(d_2)}$ as a sum of indicator variables, each of which depends on at most three others. Therefore, the maximum degrees Δ_n, Δ_n^* of the corresponding dependency graphs are bounded in the way of $\Delta_n\le 3$ and $\Delta_n^*\le 9$. We can apply Theorem 4.3.1 again and proceed the same way as in Theorem 4.3.4.

If $d_1 = d_1(n)$ diverges as well, then the maximum degrees Δ_n, Δ_n^* of the dependency graphs of $\left(X_{\text{inv}}^{(d_1)}, X_{\text{des}}^{(d_2)}\right)^\top$ are bounded in the way of $\Delta_n \leq 4d_1$ and $\Delta_n^* \leq 8d_1$. Moreover, we have to take into account that by (5.2), (5.3a), (5.3b), $X_{\text{inv}}^{(d_1)}$ is based on $\Theta(nd_1)$ summands, so we have to replace n with nd in Theorem 4.3.1. Therefore, Theorem 4.3.1 gives an $o(1)$ bound of $r_n(\mathcal{A}^{\text{re}})$ only if $d_1^{2/3} \log(nd_1)^{7/6} = o((nd_1)^{1/6})$, which leads to the condition $d_1 = o\left(n^{1/3} \log(n^{4/3})^{-7/3}\right)$. For any faster growth rate of d_1, we have to replace $X_{\text{inv}}^{(d_1)}$ with $\hat{X}_{\text{inv}}^{(d_1)}$ and we still obtain the CLT on all classical Weyl groups from [17, Theorem 2], provided the respective conditions in Corollary 5.2.3 and Lemma 5.2.4.

Corollary 5.3.2. *If $d_1 = o\left(n^{1/3} \log(n^{4/3})^{-7/3}\right)$ or $d_1 = \omega(n^{1/2})$, then $\left(X_{\text{inv}}^{(d_1)}, X_{\text{des}}^{(d_2)}\right)^\top$ satisfies the CLT for any $d_2 = d_2(n) \in \{1, \ldots, d_{\max} - 1\}$ with $d_2 = d_{\max} - \omega(1)$.*

We now postulate the univariate and bivariate EVLTs for generalized inversions and descents. For a univariate triangular array consisting of generalized descents, the EVLT is straightforward since it is not necessary to use the Hájek projection. We already argued in Remark 4.4.2 that a subexponential bound on k_n can be obtained if the Hájek projection is not needed for the Gaussian approximation. So, in the univariate EVLT for generalized descents, we can use the bound on k_n stated in Theorem 4.4.3, but we have to take the number of summands of $X_{\text{des}}^{(d)}$ into account, which is $N_{n,d}^{=}$.

Theorem 5.3.3. *Let $(X_{nj})_{j=1,\ldots,k_n}$ be a row-wise i.i.d. triangular array with $X_{n1} \stackrel{\mathcal{D}}{=} X_{\text{des}}^{(d)}$ for a sequence $d = d_n$ with $1 \leq d \leq d_{\max} - 1$ and $d = d_{\max} - \omega(1)$, and let M_n, a_n, b_n be as in Theorem 3.3.2. If $k_n = \exp\left(o((N_{n,d}^{=})^{1/7})\right)$, then*

$$\forall x \in \mathbb{R}: \quad \mathbb{P}(M_n \leq a_n x + b_n) \longrightarrow \exp(-\exp(-x)) .$$

An analogous EVLT for $X_{\text{inv}}^{(d)}$ applies if d grows slow enough to permit the application of Theorem 4.3.1.

Theorem 5.3.4. *Let $(X_{n1}, \ldots, X_{nk_n})$ be a row-wise i.i.d. triangular array with $X_{n1} \stackrel{\mathcal{D}}{=} X_{\text{inv}}^{(d)}$ and $d = d_n$ as in Theorem 5.3.3. Let M_n, a_n, b_n be as in Theorem 3.3.2. If $d = o(n^{1/3})$ and $k_n = \exp(o(n^{1/7} d^{-3/7}))$, then*

$$\forall x \in \mathbb{R}: \quad \mathbb{P}(M_n \leq a_n x + b_n) \longrightarrow \Lambda(x) .$$

Proof. According to the above considerations, the maximum degrees Δ_n, Δ_n^* of the dependency graphs of the representations (5.2), (5.3a), and (5.3b) are bounded in the way of $\Delta_n \leq 4d$ and $\Delta_n^* \leq 8d$. Since these representations are based on $\Theta(nd)$ summands, we need to replace n with nd when applying Theorem 4.3.1. An application of Theorem 4.3.1 with $n \vee k_n$ i.i.d. iterations of $X_{\text{inv}}^{(d)}$ yields

$$|\mathbb{P}(M_n \leq a_n x + b_n) - \mathbb{P}(\mathcal{M}_n \leq \alpha_n x + \beta_n)| = O\left(n^{-1/6} d^{1/2} \log(k_n)^{7/6}\right) = o(1) ,$$

where $\mathcal{M}_n$ is the maximum of n i.i.d. copies of the standard normal distribution. The claim follows. □

For any other growth rate of d, we can state an EVLT only in cases where the Hájek approximation of $X_{\text{inv}}^{(d)}$ is successful. These cases are characterized by Corollary 5.2.3 and

Lemma 5.2.4. Again, the Hájek projection causes a strongly reduced asymptotic bound on k_n.

Recall that the proof of Theorem 4.4.1 consists of two parts: The first part is to show that the error resulting from replacing X_{inv} with $\hat{X}_{\text{inv}}$ vanishes in probability (cf. (4.8)). This part is pivotal for the upper bound on the number of samples k_n from a finite Coxeter group of rank n. The second part is to apply Theorem 4.3.1 for $(\hat{X}_{\text{inv}}, X_{\text{des}})$ (cf. (4.7)). The number of descents can be ignored within these considerations, so we can first use this method to state the univariate EVLT for $X^{(d)}_{\text{inv}}$. Depending on the magnitude of d, the upper bound on k_n can now be even stricter than the bound given in Theorem 4.4.1.

Theorem 5.3.5. *Let $(X_{nj})_{j=1,\dots,k_n}$ be a row-wise i.i.d. triangular array with $X_{n1} \overset{\mathcal{D}}{=} X^{(d)}_{\text{inv}}$ on a classical Weyl group of rank n, and with $d = d_n$ such that $d = \omega(n^{1/2})$. Assume $k_n \log(k_n) = o(d^2/n)$ and let M_n, a_n, b_n be as in Theorem 3.3.2. Then,*

$$\forall x \in \mathbb{R}\colon \quad \mathbb{P}(M_n \leq a_n x + b_n) \longrightarrow \exp(-\exp(-x))\,.$$

Proof. The conditions in Corollary 5.2.3 (for symmetric groups) and Lemma 5.2.4 (for other classical Weyl groups) ensure that

$$1 - \frac{\operatorname{Var}(X^{(d)}_{\text{inv}})}{\operatorname{Var}(\hat{X}^{(d)}_{\text{inv}})} = o(1)\,. \tag{5.8}$$

As seen in the proof of Theorem 4.4.1, the rate of convergence in (5.8) determines the bound on k_n by means of (4.9) and (4.10). We compute this rate for symmetric groups, since the same conclusions can be obtained on the other classical Weyl groups. In the short case, we have

$$\begin{aligned} 1 - \frac{\operatorname{Var}(X^{(d)}_{\text{inv}})}{\operatorname{Var}(\hat{X}^{(d)}_{\text{inv}})} &= 1 - \frac{4d^3 + 6nd + 3d^2 - d}{4d^3 + 6d^2 + 2d} = \frac{4d^2 + 6n + 3d - 1}{4d^2 + 6d + 2} \\ &= 1 - \frac{4d^2}{4d^2 + 6d + 2} - \frac{3d + 1}{4d^2 + 6d + 2} - \frac{6n}{4d^2 + 6d + 2} \\ &= \Theta\left(\frac{1}{d}\right) - \frac{6n}{4d^2 + 6d + 2}\,. \end{aligned}$$

Apparently, $6n/(4d^2 + 6d + 2)$ always dominates $1/d$. In conclusion,

$$1 - \frac{\operatorname{Var}(X^{(d)}_{\text{inv}})}{\operatorname{Var}(\hat{X}^{(d)}_{\text{inv}})} = \Theta\left(\frac{n}{d^2}\right),$$

giving the condition of $k_n \log(k_n) = o(d^2/n)$ according to the arguments in the proof of Theorem 4.4.1. From here, we proceed as in the proof of Theorem 4.4.1. In the long case, we always have $\operatorname{Var}(X^{(d)}_{\text{inv}}) \sim \operatorname{Var}(\hat{X}^{(d)}_{\text{inv}})$ and therefore,

$$1 - \frac{\operatorname{Var}(X^{(d)}_{\text{inv}})}{\operatorname{Var}(\hat{X}^{(d)}_{\text{inv}})} = \Theta\left(\frac{1}{d}\right) = \Theta\left(\frac{1}{n}\right) = \Theta\left(\frac{n}{d^2}\right).$$

Again, the proof now follows the same steps as in Theorem 4.4.1. □

The bivariate EVLT for $\left(X_{\text{inv}}^{(d_1)}, X_{\text{des}}^{(d_2)}\right)^\top$ can now be stated analogously to the previous three EVLTs. Note that the descent component does not interfere with the arguments in the proofs of Theorems 5.3.4 and 5.3.5.

Corollary 5.3.6. *Let $(X_{nj})_{j=1,\dots,k_n}$ be a row-wise i.i.d. triangular array with $X_{n1} \stackrel{\mathcal{D}}{=} (X_{\text{inv}}^{(d_1)}, X_{\text{des}}^{(d_2)})$ for two sequences $d_1 = d_1(n), d_2 = d_2(n)$ as above.*

(a) *If $d_1 = o(n^{1/3})$, then assume $k_n = \exp\left(o\left((N_{n,d_2}^{=})^{1/7} \wedge n^{1/7} d_1^{-3/7}\right)\right)$.*

(b) *If $d_1 = \omega(n^{1/2})$, then we assume $k_n \log(k_n) = o(d^2/n)$ in analogy to Theorem 5.3.5.*

Let M_n, a_n, b_n be as in Theorem 4.4.1. Then,

$$\forall \mathbf{x} \in \mathbb{R}^2: \quad \mathbb{P}(M_n \leq a_n * \mathbf{x} + b_n) \longrightarrow \Lambda_2(\mathbf{x}) .$$

5.4 Technical proofs

5.4.1 Proof of Lemma 5.2.1

We recall the representation of $\hat{X}_{\text{inv}}^{(d)}$ in the short case provided in (5.5). Its variance results only from the sum of all $\omega_d(k) Z_k$. Therefore,

$$\begin{aligned}
\operatorname{Var}\left(\hat{X}_{\text{inv}}^{(d)}\right) &= \operatorname{Var}\left(\sum_{k=n-d+1}^{n} (n-k-d) Z_k + \sum_{k=1}^{d} (d+1-k) Z_k\right) \\
&= \sum_{k=n-d+1}^{n} \frac{1}{12} (n-k-d)^2 + \sum_{k=1}^{d} \frac{1}{12} (d+1-k)^2 \\
&= \frac{1}{12} \sum_{k=1}^{d} k^2 + \frac{1}{12} \sum_{k=1}^{d} (-k)^2 = \frac{1}{6} \frac{d(d+1)(2d+1)}{6} \\
&= \frac{1}{72} \left(4d^3 + 6d^2 + 2d\right) .
\end{aligned}$$

So, the leading term is always $4d^3/72$. According to Theorem 5.1.4, we have

$$\operatorname{Var}(X_{\text{inv}}^{(d)}) = \frac{1}{72} \left(4d^3 + 6nd + 3d^2 - d\right) ,$$

therefore it must be ensured that $d^3 \gg nd \iff d \gg \sqrt{n}$, proving Lemma 5.2.1. $\square$

5.4.2 Proof of Lemma 5.2.2

In the long case, the calculations are slightly more complex due to the indices in K_2 giving a two-sided overlap. Proceeding from (5.6), we state that

$$\operatorname{Var}\left(\hat{X}_{\text{inv}}^{(d)}\right) = \frac{1}{12} \sum_{k=1}^{n-d} (d+1-k)^2 \tag{5.9a}$$

$$+ \frac{1}{12} \sum_{k=n-d+1}^{d} (n-2k+1)^2 \tag{5.9b}$$

$$+\frac{1}{12}\sum_{k=d+1}^{n}(n-k-d)^2\,. \tag{5.9c}$$

By appropriate index shifting, we calculate

$$\begin{aligned}
(5.9a), (5.9c) &= \frac{1}{12}\left(\sum_{k=1}^{d}k^2-\sum_{k=1}^{2d-n}k^2\right)\\
\implies (5.9a)+(5.9c) &= \frac{1}{6}\left(\sum_{k=1}^{d}k^2-\sum_{k=1}^{2d-n}k^2\right)\\
&= \frac{1}{36}(n-d)(14d^2+d(9-10n)+2n^2-3n+1)\,,\\
(5.9b) &= \frac{1}{36}(2d-n)(4d^2-4dn+n^2-1)\,.
\end{aligned}$$

This gives the overall result

$$\operatorname{Var}\left(\hat{X}_{\text{inv}}^{(d)}\right)=\frac{1}{36}\left(-6d^3+12d^2n-9d^2-6dn^2+12dn-3d+n^3-3n^2+2n\right).$$

In contrast, by Theorem 5.1.4,

$$\operatorname{Var}\left(X_{\text{inv}}^{(d)}\right)=\frac{1}{36}\left(-6d^3+\left(12n-\frac{21}{2}\right)d^2-\left(6n^2-15n+\frac{9}{2}\right)d+n^3-3n^2+2n\right).$$

Since the long case implies $n/2<d<n$, all monomials of order 3 are leading terms. It is easily seen that these leading terms are matching, i.e.,

$$\operatorname{Var}\left(X_{\text{inv}}^{(d)}\right),\operatorname{Var}\left(\hat{X}_{\text{inv}}^{(d)}\right)=\frac{1}{36}\left(-6d^3+12nd^2-6n^2d+n^3+O(n^2)\right).$$

This proves Lemma 5.2.2. □

5.4.3 Proof of Lemma 5.2.4

In light of (5.7), we need to determine the linear coefficients $\omega_d(k)$ stemming from

$$\sum_{(i,j)\in\mathfrak{N}_{n,d}}\mathbb{P}(Z_i>Z_j\mid Z_k)+\sum_{(i,j)\in\widetilde{\mathfrak{N}}_{n,d}}\mathbb{P}(-Z_i>Z_j\mid Z_k)\,.$$

We write $\omega_d(k)=\omega_d(k)^+ + \omega_d(k)^-$, with $\omega_d(k)^+$ stemming from the first sum and $\omega_d(k)^-$ stemming from the second. For $\omega_d(k)^+$, we can use the observations from Lemmas 5.2.1 and 5.2.2. However, we need to take into account that due to $Z_1,\dots,Z_n\sim U(-1,1)$, we now have

$$\begin{aligned}
\mathbb{P}(Z_k>Z_j\mid Z_k)&=\frac{Z_k+1}{2}\,, & \mathbb{P}(Z_i>Z_k\mid Z_k)&=\frac{1-Z_k}{2}\,,\\
\mathbb{P}(-Z_k>Z_j\mid Z_k)&=\frac{1-Z_k}{2}\,, & \mathbb{P}(-Z_i>Z_k\mid Z_k)&=\frac{1-Z_k}{2}\,.
\end{aligned}$$

In conclusion, the coefficients $\omega_d(k)^+$ on B_n are half of the coefficients $\omega_d(k)$ on S_n if $d<n$. Otherwise, for $d\geq n$ we always have $\omega_d(k)^+=(n-2k+1)/2$. Moreover, $\omega_d(k)^-=-\widetilde{N}_{n,d}^{(k)}/2$, where

$$\widetilde{N}_{n,d}^{(k)}:=|\{(i,j)\in\widetilde{\mathfrak{N}}_{n,d}\mid i=k\text{ or }j=k\}|\,.$$

By analogy with the proofs of [75, Theorems A.4 and A.13], we need to distinguish the four cases $d \le n/2$, $n/2 \le d \le 2n/3$, $2n/3 \le d < n$, and $d \ge n$. If $d \le n/2$, then all pairs in $\widetilde{\mathfrak{N}}_{n,d}$ are located within K_1, yielding

$$\omega_d(k) = \begin{cases} (d-k+1)/2 - (d-k-1)/2, & k \le d/2 \\ (d-k+1)/2 - (d-k)/2, & d/2 < k \le d \\ 0, & d < k \le n-d \\ (n-d-k)/2, & n-d < k \le n \end{cases}$$

$$= \begin{cases} (n-d-k)/2, & n-d < k \le n \\ O(1), & \text{otherwise} \end{cases} .$$

In conclusion, if $d \le n/2$, then

$$\begin{aligned} \operatorname{Var}\left(\hat{X}_{\text{inv}}^{B,(d)}\right) &= \frac{1}{3}\sum_{k=1}^{n} \omega_d(k)^2 + O(d^2) = \frac{1}{12}\sum_{k=n-d+1}^{n} (n-d-k)^2 + O(d^2) \\ &= \frac{1}{12}\frac{d(d+1)(2d+1)}{6} + O(d^2) = \frac{1}{36}d^3 + O(d^2) . \end{aligned}$$

Due to $\operatorname{Var}\left(X_{\text{inv}}^{B,(d)}\right) = d^3/36 + nd/12 + O(d^2)$ according to Lemma 5.1.7, we obtain the same condition as in Lemma 5.2.1, namely, $d = \omega(n^{1/2})$.

If $n/2 < d < n$, then the pairs in $\widetilde{\mathfrak{N}}_{n,d}$ also cover K_2. For $k \in K_2$, there cannot be any pairs (k, j) if $n/2 < d \le 2n/3$, while this is possible if $d > 2n/3$. However, the difference between these two subcases is only marginal. If $n/2 < d \le 2n/3$, we obtain

$$\omega_d(k) = \begin{cases} O(1), & k \in K_1 \\ (n-2k+1-(d-k))/2, & k \in K_2 \\ (n-d-k)/2, & k \in K_3 \end{cases} .$$

If $d > 2n/3$, then

$$\omega_d(k) = \begin{cases} O(1), & k \in K_1 \\ (n-2k+1-(d-k-1))/2, & k \in K_2, k \le d/2 \\ (n-2k+1-(d-k))/2, & k \in K_2, k > d/2 \\ (n-d-k)/2, & k \in K_3 \end{cases} .$$

In conclusion, if $n/2 < d < n$, then

$$\begin{aligned} \operatorname{Var}\left(\hat{X}_{\text{inv}}^{B,(d)}\right) &= \frac{1}{3}\sum_{k=1}^{n} \omega_d(k)^2 + O(d^2) \\ &= \frac{1}{12}\sum_{k=n-d+1}^{d} (n-d-k+1)^2 + \frac{1}{12}\sum_{k=d+1}^{n} (n-d-k)^2 + O(d^2) \\ &= \frac{1}{12}\sum_{k=n-d+1}^{d} (n-d-k+1)^2 + \frac{1}{12}\sum_{k=1}^{2d-n} k^2 + O(d^2) \end{aligned}$$

$$= \frac{1}{12} \sum_{k=1}^{2d-n-1} k^2 + \frac{1}{12} \left(\sum_{k=1}^{d} k^2 - \sum_{k=1}^{2d-n} k^2 \right) + O(d^2)$$
$$= \frac{1}{36} d^3 + O(d^2) \, .$$

In the remaining case of $d \geq n$, the main focus is on counting $\widetilde{N}_{n,d}^{(k)}$. Figure 5.9 illustrates the positions of pairs $(i, j) \in \widetilde{\mathfrak{N}}_{n,d}$ for the exemplary choice of $n = 12, d = 16$.

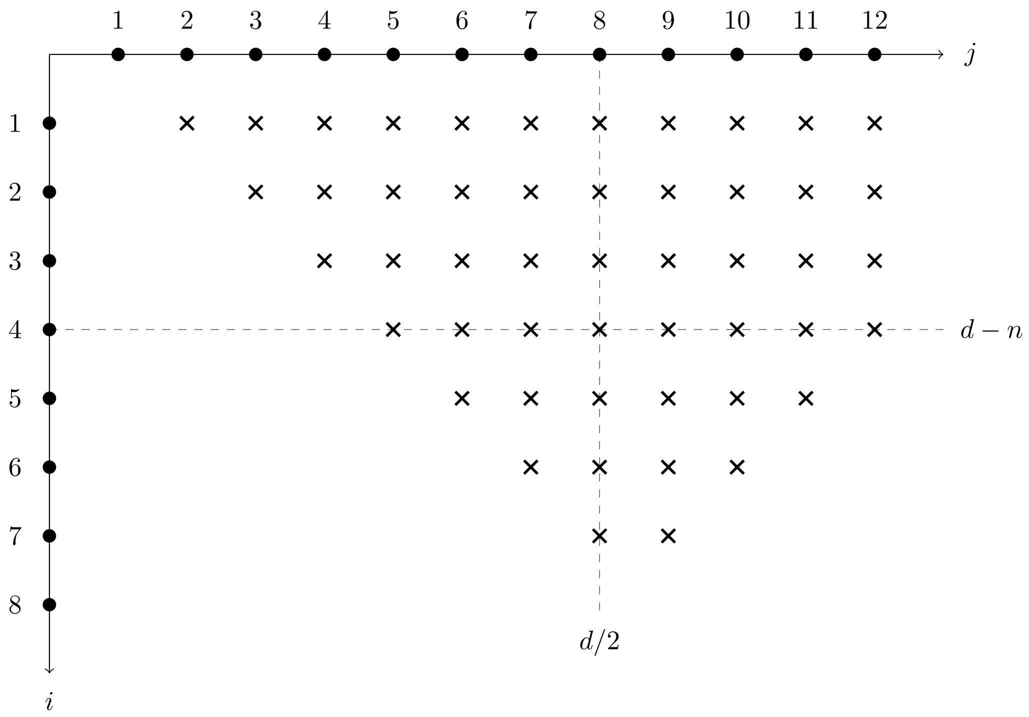


Figure 5.9: Visualization of pairs (i, j) in $\widetilde{\mathfrak{N}}_{n,d}$ for $n = 12$ and $d = 16$. The numbers $d - n$ and $d/2$ are important case distinction thresholds for counting the pairs (k, j) and (i, k), respectively.

With help of Figure 5.9, it is straightforward to count

$$\widetilde{N}_{n,d}^{(k)} = \begin{cases} n - 1, & 1 \leq k \leq d - n \\ d - k - 1, & d - n < k \leq d/2 \\ d - k, & d/2 < k \leq n \end{cases} .$$

This result is also illustrated in Figure 5.10, which displays the number of pairs (i, k) and the number of pairs (k, j).
Therefore, if $d \geq n$, we have

$$\omega_d(k) = \begin{cases} 1 - k, & 1 \leq k \leq d - n \\ (n + 2 - d - k)/2, & d - n < k \leq d/2 \\ (n + 1 - d - k)/2, & d/2 < k \leq n \end{cases} .$$

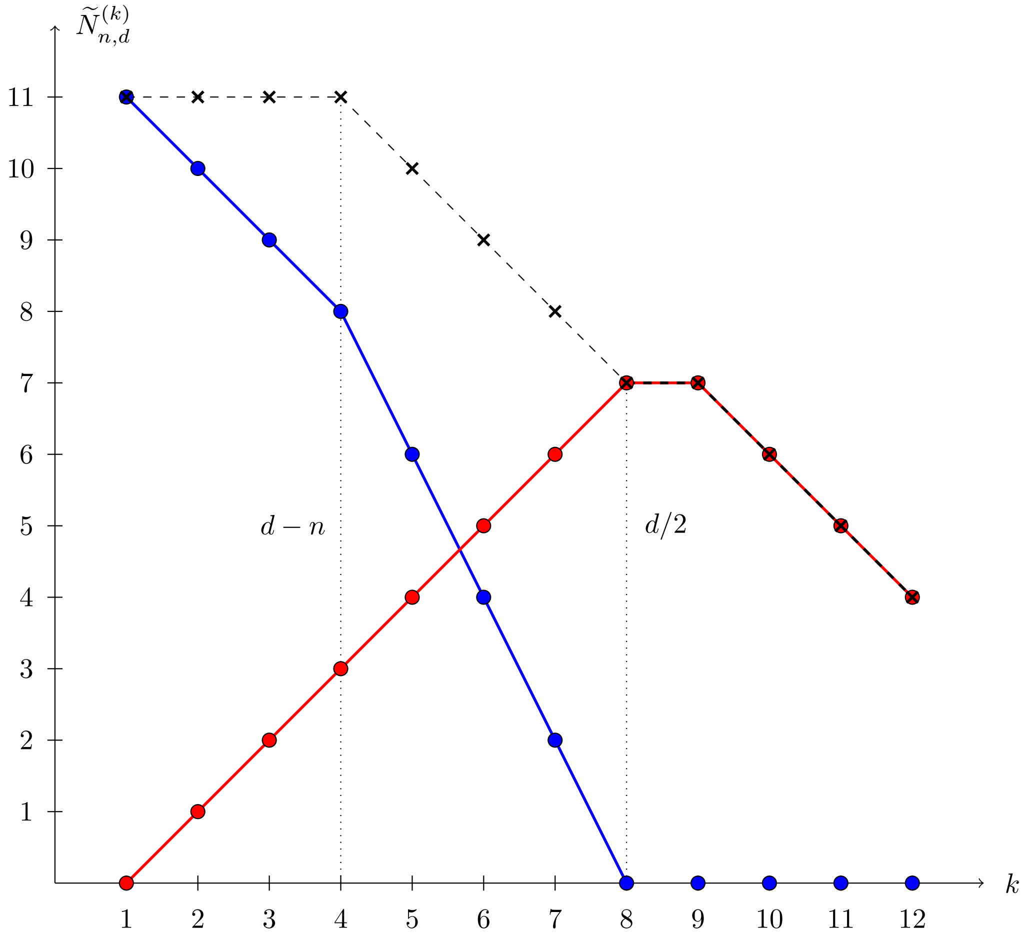


Figure 5.10: Plots of the numbers of pairs (i,k) (red) and (k,j) (blue) in $\widetilde{\mathfrak{N}}_{n,d}$. The sum of these two numbers is $\widetilde{N}_{n,d}^{(k)}$, which is displayed by the black crosses and the dashed line.

We compute

$$\begin{aligned}\sum_{k=1}^{n}\omega_d(k)^2 &= \sum_{k=1}^{d-n}(k-1)^2 + \frac{1}{4}\sum_{k=d-n+1}^{n}(n-d+1-k)^2 + O(n^2)\\ &= \frac{1}{6}\left(2d^3 - 3d^2(2n+1) + d(6n^2+6n+1) - n(2n^2+3n+1)\right)\\ &\quad + \frac{1}{24}(2n-d)(14d^2 - 20dn - 9d + 8n^2 + 6n + 1) + O(n^2)\,.\end{aligned}$$

Due to $\operatorname{Var}(Z_k) = 1/3$, we obtain

$$\begin{aligned}\operatorname{Var}(\hat{X}_{\mathrm{inv}}^{B,(d)}) &= \frac{1}{72}\Big(-6d^3 + 3d^2(8n-1) - 3d(8n^2-1) + 2n(4n^2-1)\Big) + O(n^2)\\ &= -\frac{1}{12}d^3 + \frac{1}{3}nd^2 - \frac{1}{3}n^2 d + \frac{1}{9}n^3 + O(n^2)\,.\end{aligned}$$

By Lemma 5.1.7, this also applies for $\operatorname{Var}(X_{\mathrm{inv}}^{B,(d)})$, completing the proof. □

List of Symbols

Chapter 1

a_n, b_n	affine-linear normalization sequences for extreme values
α_n, β_n	normalization sequences for the standard normal distribution
C	copula of a multivariate distribution
$\mathrm{CDA}(C)$	copula domain of attraction
$\mathfrak{D}, \mathfrak{D}(\cdot), \mathfrak{D}'(\cdot)$	symbols for mixing conditions
F	(joint) distribution function of random variables
F_n	sequence of distribution functions
$F_1, \dots, F_d$	marginals of a multivariate distribution function
$\overline{F}$	tail of a distribution function
$F^{\leftarrow}$	quantile function of a distribution
F^*	auxiliary function $F(x^* - 1/x)$
F_*	transformation of a multivariate distribution function to Fréchet-1 marginals
G	an arbitrary extreme value distribution
G_*	transformation of a multivariate extreme value distribution to Fréchet-1 marginals
k_n	number of (i.i.d.) random variables in the n-th row of a triangular array
L	a slowly varying function
Λ	Gumbel distribution (univariate, standard)
λ_u	upper tail dependence coefficient
$\mathrm{MDA}(G)$	max-domain of attraction
M_n	maximum of a sequence of random variables, or of a row of a triangular array
M_n^*	maximum of independent copies of random variables
N_n	ranks of binomial variables in a triangular array
Φ	CDF of the standard normal distribution
Φ_α	Fréchet distribution function with shape parameter $\alpha > 0$
Ψ_α	Weibull distribution function with shape parameter $\alpha > 0$
RV_α	regular variation with index α

$\mathcal{S}_n$	partial sum $X_1 + \ldots + X_n$ or $X_{n1} + \ldots + X_{n,k_n}$
u_n	sequence of thresholds
x^*, x_n^*	right endpoint of a distribution function
$X_1, X_2, \ldots$	sequence of random variables
$X_1^*, X_2^*, \ldots$	sequence of independent copies of $X_1, X_2, \ldots$
$X_{n1}, \ldots, X_{n,k_n}$	triangular array of random variables

Chapter 2

A_n	symmetric group as a family of finite Coxeter groups
B_n	signed permutation group
$d_1, \ldots, d_n$	degrees of a finite Coxeter group
$d_{\max}$	maximum degree of a finite Coxeter group
$\mathrm{des}(\cdot)$	number of descents on a finite Coxeter group
$\mathrm{Des}(\cdot)$	set of descents on a finite Coxeter group
D_n	even-signed permutation group
e	neutral group element
$\mathcal{G}_{\mathrm{inv}}, \mathcal{G}_{\mathrm{des}}$	generating function of inversions or descents
$I_2(m)$	dihedral group of order $2m$
$\mathrm{inv}(\cdot)$	number of inversions on a finite Coxeter group
$\mathrm{Inv}(\cdot)$	set of inversions on a finite Coxeter group
$l(w)$	word length within a Coxeter group
M	Coxeter matrix
$m_{n,i}$	sizes of dihedral components in a sequence of Coxeter groups
q_i	negatives of roots of $\mathcal{G}_{\mathrm{des}}$
$\mathrm{rk}(\cdot)$	rank of a finite Coxeter group
S	generating set of a Coxeter group
$s, s_1, s_2, \ldots$	generators of a Coxeter group
s_n^2	variance of inversions or descents in a sequence of Coxeter groups
S_n	symmetric group
T	reflection set of a Coxeter group (conjugates of generators)
$\tau_1, \ldots, \tau_{n-1}$	generators (neighboring transpositions) of the symmetric group S_n
W	Coxeter group
w	element of a Coxeter group
(W, S)	Coxeter system

W_n	sequence of Coxeter groups
$W_{n,i}$	components in a sequence of composed Coxeter groups
$\mathcal{W}, \mathcal{W}_1, \ldots, \mathcal{W}_d$	doubly-indexed permutation statistics based on antisymmetric matrices
$X_{\mathrm{inv}}, X_{\mathrm{des}}$	inversions and descents as random variables
$X^B_{\mathrm{inv}}, X^B_{\mathrm{des}}$	random number of inversions and descents on the groups B_n
$X^D_{\mathrm{inv}}, X^D_{\mathrm{des}}$	random number of inversions and descents on the groups D_n
X_{ides}	random number of inverse descents
X_T	two-sided Eulerian statistic $X_{\mathrm{des}} + X_{\mathrm{ides}}$

Chapter 3

$a_{n,k}$	k-th coefficient of the power series $\mathcal{L}_n$
F_n	distribution function of a (standardized) partial sum of n summands
G_n	normal part of a composed finite Coxeter group
γ_{kj}	k-th cumulant of X_j
Γ_{kn}	arithmetic mean of cumulants $\gamma_{k1}, \ldots, \gamma_{kn}$
h_n	number of dihedral components
I_n	dihedral part of a composed finite Coxeter group
l_n	number of classical Weyl group components
L_j	cumulant-generating function of the random variable X_j
$\mathcal{L}$, $\mathcal{L}_n$	cumulant-based power series in Cramér's and Petrov-Robinsons' theorem
$m_{\max}$	maximum size of dihedral components
Φ, Φ_d	CDF of the (d-variate) standard normal distribution
Q_n	moment-based power series in Feller's theorem
$q_{n,\nu}$	ν-th coefficient of the power series Q
r_n	total rank of a group with both classical and dihedral components
s_n^2	(total) variance of a partial sum $\mathcal{S}_n$
$\sigma_1^2, \ldots, \sigma_n^2$	individual variances of the n summands of $\mathcal{S}_n$
$X^G_{\mathrm{inv}}, X^G_{\mathrm{des}}$	inversions and descents on the classical parts of a composed Coxeter group
$X^I_{\mathrm{inv}}, X^I_{\mathrm{des}}$	inversions and descents on the dihedral parts of a composed Coxeter group
$X_{n1}, \ldots, X_{nk_n}$	triangular array of either random inversions or random descents

Chapter 4

$*$	component-wise multiplication
$\alpha_n(k), \alpha(k)$	α-mixing coefficients at lag k

$\mathcal{A}, \mathcal{A}^{\mathrm{re}}$	system of hyperrectangles
$\mathcal{A}^{\mathrm{ext}}$	system of positive orthants $[\mathbf{x}, \infty)^d$
Δ_n, Δ_n^*	first and second order maximum degrees of the n-th dependency graph
$\mathrm{GR}(p)$	generalized Rademacher distribution with parameter p
$\mathrm{Inv}^+(\cdot)$	inversions on B_n, D_n analogous to those on S_n
$\mathrm{Inv}^-(\cdot)$	negative sum pair inversions on B_n, D_n
$\mathrm{Inv}^\circ(\cdot)$	inversions induced by negative entries on B_n
l	number of components in direct products of classical Weyl groups
Λ_2	bivariate Gumbel distribution with independent marginals
$n_1, \ldots, n_l$	decreasingly sorted ranks of components in direct products of classical Weyl groups
$\mathrm{N}_2(\cdot, \cdot)$	bivariate normal distribution with specified mean and covariance matrix
$\omega_d(k)$	linear coefficients of Z_k in the Hájek projection of $X_{\mathrm{inv}}^{(d)}$
p	probability of choosing a negative sign on B_n, D_n (sign bias)
$\mathfrak{p}$	high dimension
ρ_n	correlation between two sequences of random variables
r_n	Gaussian approximation error
$\Sigma^{(n)}$	covariance matrix of the high-dimensional sum $X^{(n)}$
$U_1, \ldots, U_n$	absolute values of $Z_1, \ldots, Z_n$ (important on B_n, D_n)
$X^{(n)}$	partial sum $X_1^{(n)} + \ldots + X_n^{(n)}$ in the high-dimensional framework
$\hat{X},\ \hat{X}_n$	Hájek projection of a random variable
$\hat{X}_{\mathrm{inv}}, \hat{X}_{\mathrm{des}}$	Hájek projection of inversions, or respectively, descents
$Y_{\mathrm{inv}}, Y_{\mathrm{des}}$	standardizations of random inversions and descents
$Z_1, \ldots, Z_n$	i.i.d. variables $U(0,1)$, $U(-1,1)$, or $\mathrm{GR}(p)$, depending on the context

Chapter 5

d, d_1, d_2	range of generalized inversions or descents
$[i], [ij], [\widetilde{ij}]$	positive roots of the classical Weyl groups S_n, B_n, D_n
K_1, K_2, K_3	splitting of $\{1, \ldots, n\}$ into subregions to analyze d-inversions
$\mathfrak{N}_{n,d}$	pairs of indices (i,j) with $i, j \in \{1, \ldots, n\}$ and $1 \leq j - i \leq d$
$\widetilde{\mathfrak{N}}_{n,d}$	pairs of indices (i,j) with $i, j \in \{1, \ldots, n\}$ and $j + i \leq d$
$N_{n,d}^{\leq}$	number of positive roots with height at most d
$N_{n,d}^{=}$	number of positive roots with height exactly d
$X_{\mathrm{inv}}^{(d)}, X_{\mathrm{des}}^{(d)}$	generalized inversions and descents as random variables

List of Figures

1.1 A plot of the (standard) Gumbel distribution. 5
1.2 A plot of three exemplary Fréchet distributions. Large choices of the shape parameter α produce more curved distributions. 5
1.3 A plot of three exemplary Weibull distributions, each cut off at the right endpoint $x^* = 0$. Large choices of the shape parameter α produce more curved distributions. 6
1.4 Display of a distribution function F together with its tail function $\overline{F}$. . . . 7
1.5 Plot of the normalization constants α_n, β_n for $n = 2, \ldots, 10$. 9
1.6 Plot of the two-dimensional Gumbel distribution $\Lambda_2(x, y)$ with independent marginals, for $-4 \leq x, y \leq 4$. 18

2.1 Example of a permutation $\pi = (4, 5, 1, 2, 3)$ on five numbers, highlighting an inverted pair: $2 < 4$, but $\pi(i) = 5 > 2 = \pi(j)$. 20
2.2 Graphical display of descents (highlighted in blue) of the permutation $\pi = (4, 6, 3, 2, 1, 5)$. The arrows emphasize the convention that the points at the beginning of descending segments are referred to as descents. 21
2.3 The Coxeter graph of $A_n = S_{n+1}$. 22
2.4 Example of a signed permutation derived from the permutation illustrated in Figure 2.1. This signed permutation can be written as $(4, -5, 1, -2, 3)$. The negative signs are marked in red. 22
2.5 The Coxeter graph of the signed permutation group B_n. The leftmost node represents the additional generator τ_0. 23
2.6 The Coxeter graph of the even-signed permutation group D_n. The top node represents the additional generator $\tilde{\tau}_0$. 23
2.7 True to scale probability mass function of Eulerian numbers on S_n for $n = 31$. 29

3.1 Beginning of a triangular array of permutation statistics on all symmetric groups. The row lengths k_n have no specific regularity. 36
3.2 Beginning of a uniform triangular array of permutation statistics with an exemplary choice of the first five symmetric groups. The choices $r(n)$ of the ranks have no specific regularity. 36

4.1 Display of the non-trivial contributions to $\mathbb{E}(X_{\text{inv}} \mid Z_k)$ stemming from (i, k), $i = 1, \ldots, k-1$ (red), and (k, j), $j = k+1, \ldots, n$ (blue). 53
4.2 Display of the coefficients of Z_k appearing in $\hat{X}_{\text{inv}}$. They add up in pairs to some $U(-j, j)$ with j even (left) or j odd (right). These pairs are highlighted in red and blue, respectively. 54
4.3 Probability distribution functions of generalized Rademacher distributions for $p = 1/4$ (blue) and $p = 3/4$ (red). 58

4.4 Plots of empirical distribution functions of $(M_n - b_n)/a_n$ for the symmetric groups S_n, with M_n, a_n, b_n as given in Theorem 4.4.1 and $n \in \{20, 50, 100, 200, 500, 1000\}$. These empirical distribution functions converge to Λ_2 as $n \to \infty$. 71
4.5 Plots of the standard Gumbel distribution function Λ and the simulated empirical distribution function of $\mathbb{P}(M_n \leq a_n x + b_n)$ for $x \in [-4, 4]$ and M_n, a_n, b_n stemming from X_T, as given in Conjecture 4.5.5. 73
4.6 Canceling pairs of positive and negative covariances between $\mathbf{1}\{Z_i > Z_j\}$ and $\mathbf{1}\{Z_k > Z_{k+1}\}$ as k passes over $1, \ldots, n-1$ in the non-exceptional case of $1 < i < i+1 < j < n$. The covariance is zero for all other values of k. . . 77
4.7 Display of covariances between $\mathbf{1}\{Z_i > Z_j\}$ and $\mathbf{1}\{Z_k > Z_{k+1}\}$ in the case of i and $j = i+1$ being subsequent. 78

5.1 A central index $k \in K_2$ for which the whole segment $\{k-d, \ldots, k+d\}$ is contained in $\{1, \ldots, n\}$. 94
5.2 A left-positioned index $k \in K_2$ with left-sided overlap, due to $k - d \notin \{1, \ldots, d\}$. The overlap is indicated by the red segment. 94
5.3 A right-positioned index $k \in K_3$ with right-sided overlap, due to $k + d \notin \{1, \ldots, n\}$. The overlap is indicated by the red segment. 95
5.4 Hasse diagrams of the root posets of the signed permutation group B_5 (left) and the even-signed permutation group D_6 (right). The roots $[ij]$ are marked blue, the roots $[\widetilde{ij}]$ are marked red and the roots $[i]$ are marked green. The vertical axis indicates the heights of all positive roots. 96
5.5 Overview of cancellation and remainders of Z_k and $1 - Z_k$ in the three subregions K_1, K_2, K_3. 98
5.6 Overview of relevant pairs (i, j) for computing the variance of $\mathbb{E}\left(X_{\text{inv}}^{(d)} \mid Z_k\right)$, where $n = 15$ and $d = 4$. For each of the subregions K_1, K_2, K_3, an exemplary index k is chosen, and the pairs that give a non-trivial contribution are highlighted in red for $k \in K_1$, in blue for $k \in K_2$, and in green for $k \in K_3$. 99
5.7 Plot of $\omega_d(k)$ in the short case ($d \leq n/2$). 100
5.8 Plot of $\omega_d(k)$ in the long case ($d > n/2$). 101
5.9 Visualization of pairs (i, j) in $\widetilde{\mathfrak{N}}_{n,d}$ for $n = 12$ and $d = 16$. The numbers $d - n$ and $d/2$ are important case distinction thresholds for counting the pairs (k, j) and (i, k), respectively. 108
5.10 Plots of the numbers of pairs (i, k) (red) and (k, j) (blue) in $\widetilde{\mathfrak{N}}_{n,d}$. The sum of these two numbers is $\widetilde{N}_{n,d}^{(k)}$, which is displayed by the black crosses and the dashed line. 109

Bibliography

[1] Clive W. Anderson, *Extreme value theory for a class of discrete distributions with applications to some stochastic processes*, Journal of Applied Probability **7** (1970), no. 1, 99–113.

[2] Clive W. Anderson, Stuart G. Coles, and Jürg Hüsler, *Maxima of Poisson-like variables and related triangular arrays*, Annals of Applied Probability **7** (1997), no. 4, 953–971.

[3] Richard Arratia, Andrew D. Barbour, and Simon Tavaré, *Logarithmic combinatorial structures: a probabilistic approach*, vol. 1, European Mathematical Society, 2003.

[4] İlker Arslan, Ümit Işlak, and Cihan Pehlivan, *On unfair permutations*, Statistics & Probability Letters **141** (2018), 31–40.

[5] Edward A. Bender, *Central and local limit theorems applied to asymptotic enumeration*, Journal of Combinatorial Theory. Series A **15** (1973), 91–111.

[6] Rabi N. Bhattacharya and R. Ranga Rao, *Normal approximation and asymptotic expansions*, corrected ed., Classics in Applied Mathematics, vol. 64, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010.

[7] Patrick Billingsley, *Probability and measure*, Wiley series in probability and mathematical statistics (1995).

[8] Anders Björner and Francesco Brenti, *Combinatorics of Coxeter groups*, vol. 231, Springer Science & Business Media, 2006.

[9] Miklós Bóna, *Generalized descents and normality*, Electronic Journal of Combinatorics **15** (2008), no. 1, Note 21, 8.

[10] Miklós Bóna, *On three different notions of monotone subsequences*, Permutation patterns, London Mathematical Society Lecture Note Series, vol. 376, Cambridge University Press, 2010, pp. 89–114.

[11] Heejong Bong, Arun K. Kuchibhotla, and Alessandro Rinaldo, *Dual induction CLT for high-dimensional m-dependent data*, The Annals of Statistics **41** (2024), no. 6, 2786–2819.

[12] Arup Bose, Amites Dasgupta, and Krishanu Maulik, *Maxima of Dirichlet and triangular arrays of gamma variables*, Statistics & Probability Letters **78** (2008), no. 16, 2811–2820.

[13] Petter Brändén, *Unimodality, log-concavity, real-rootedness and beyond*, Handbook of enumerative combinatorics, Discrete Mathematics and its Applications, CRC Press, Boca Raton, FL, 2015, pp. 437–483.

[14] Francesco Brenti, *q-Eulerian polynomials arising from Coxeter groups*, European Journal of Combinatorics **15** (1994), no. 5, 417–441.

[15] Benjamin Brück and Frank Röttger, *A central limit theorem for the two-sided descent statistic on Coxeter groups*, Electronic Journal of Combinatorics **29** (2022), no. 1, Paper No. 1.1, 25.

[16] Enrique Castillo, *Extreme value theory in engineering*, Elsevier, 2012.

[17] Jinyuan Chang, Xiaohui Chen, and Mingcong Wu, *Central limit theorems for high dimensional dependent data*, Bernoulli **30** (2024), no. 1, 712–742.

[18] Jinyuan Chang, Qiwei Yao, and Wen Zhou, *Testing for high-dimensional white noise using maximum cross-correlations*, Biometrika **104** (2017), no. 1, 111–127.

[19] Sourav Chatterjee, *A new method of normal approximation*, The Annals of Probability **36** (2008), no. 4, 1584–1610.

[20] Sourav Chatterjee and Persi Diaconis, *A central limit theorem for a new statistic on permutations*, Indian Journal of Pure and Applied Mathematics **48** (2017), no. 4, 561–573.

[21] Louis H. Y. Chen and Qi-Man Shao, *Normal approximation under local dependence*, The Annals of Probability **32** (2004), no. 3A, 1985–2028.

[22] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato, *Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors*, The Annals of Statistics **41** (2013), no. 6, 2786–2819.

[23] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato, *Central limit theorems and bootstrap in high dimensions*, The Annals of Probability **45** (2017), no. 4, 2309–2352.

[24] Victor Chernozhukov, Denis Chetverikov, and Yuta Koike, *Nearly optimal central limit theorem and bootstrap approximations in high dimensions*, The Annals of Applied Probability **33** (2023), no. 3, 2374–2425.

[25] Victor Chernozhuokov, Denis Chetverikov, Kengo Kato, and Yuta Koike, *Improved central limit theorem and bootstrap approximations in high dimensions*, The Annals of Statistics **50** (2022), no. 5, 2562–2586.

[26] Mark A. Conger, *A refinement of the Eulerian numbers, and the joint distribution of* $\pi(1)$ *and* $\mathrm{Des}(\pi)$ *in* S_n, Ars Combinatoria. A Canadian Journal of Combinatorics **95** (2010), 445–472.

[27] Mark A. Conger and Divakar Viswanath, *Normal approximations for descents and inversions of permutations of multisets*, Journal of Theoretical Probability **20** (2007), no. 2, 309–325.

[28] Harold S. M. Coxeter, *The Complete Enumeration of Finite Groups of the Form* $r_i^2 = (r_i r_j)^{k_{ij}} = 1$, Journal of the London Mathematical Society **1** (1935), no. 1, 21–25.

[29] Harald Cramér, *Sur un nouveau theoreme-limite de la theorie des probabilities*, Scientifiques et Industrielles **736** (1938), 5–23.

[30] Debraj Das and Soumendra Lahiri, *Central Limit Theorem in high dimensions: The optimal bound on dimension growth rate*, Transactions of the American Mathematical Society **374** (2021), no. 10, 6991–7009.

[31] Laurens de Haan, *Sample extremes: an elementary introduction*, Statistica Neerlandica **30** (1976), no. 4, 161–172.

[32] Filippo De Mari and Mark A. Shayman, *Generalized Eulerian numbers and the topology of the Hessenberg variety of a matrix*, Acta Applicandae Mathematicae **12** (1988), no. 3, 213–235.

[33] Stefano Demarta, *The copula approach to modelling multivariate extreme values*, Ph.D. thesis, ETH Zürich, 2007.

[34] Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, vol. 38, Springer Science & Business Media, 2009.

[35] Hang Deng, *Slightly conservative bootstrap for maxima of sums*, arXiv preprint arXiv:2007.15877 (2020).

[36] Hang Deng and Cun-Hui Zhang, *Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors*, The Annals of Statistics **48** (2020), no. 6, 3643–3671.

[37] Pascal Sielenou Dkengne, Nicolas Eckert, and Philippe Naveau, *A limiting distribution for maxima of discrete stationary triangular arrays with an application to risk due to avalanches*, Extremes **19** (2016), no. 1, 25–40.

[38] Philip Dörr, *Extremes of generalized inversions on symmetric groups*, arXiv preprint arXiv:2404.06598 (2024), Submitted to: Séminaire Lotharingien de Combinatoire.

[39] Philip Dörr and Johannes Heiny, *Joint extremes of inversions and descents of random permutations*, Journal of Theoretical Probability **38** (2025), no. 2, 1–37.

[40] Philip Dörr and Thomas Kahle, *Extreme Values of Permutation Statistics*, The Electronic Journal of Combinatorics **31** (2024), no. 3, Paper No. 3.10, 1–18.

[41] Debbie J. Dupuis and Bruce L. Jones, *Multivariate extreme value theory and its usefulness in understanding risk*, North American Actuarial Journal **10** (2006), no. 4, 1–27.

[42] Fabrizio Durante, Juan Fernández-Sánchez, and Carlo Sempi, *A topological proof of Sklar's theorem*, Applied Mathematics Letters **26** (2013), no. 9, 945–948.

[43] Anne Dutfoy, Sylvie Parey, and Nicolas Roche, *Multivariate extreme value theory – a tutorial with applications to hydrology and meteorology*, Dependence Modeling **2** (2014), no. 1, 191–228.

[44] Xiao Fang and Yuta Koike, *High-dimensional central limit theorems by Stein's method*, The Annals of Applied Probability **31** (2021), no. 4, 1660–1686.

[45] Xiao Fang and Adrian Röllin, *Rates of convergence for multivariate normal approximation with applications to dense graphs and doubly indexed permutation statistics*, Bernoulli **21** (2015), no. 4, 2157–2189.

[46] William Feller, *Generalization of a probability limit theorem of Cramér*, Transactions of the American Mathematical Society **54** (1943), no. 3, 361–372.

[47] William Feller, *An introduction to probability theory and its applications*, vol. 2, John Wiley and Sons, 1968.

[48] Valentin Féray, *On the central limit theorem for the two-sided descent statistics in Coxeter groups*, Electronic Communications in Probability **25** (2020), Paper No. 28, 6.

[49] Anne-Laure Fougères, *Multivariate extremes*, Extreme values in finance, telecommunications, and the environment **373** (2004), 388.

[50] Jason Fulman, *The distribution of descents in fixed conjugacy classes of the symmetric groups*, Journal of Combinatorial Theory. Series A **84** (1998), no. 2, 171–180.

[51] ———, *Stein's method and non-reversible Markov chains*, Stein's method: expository lectures and applications, IMS Lecture Notes Monogr. Ser., vol. 46, Inst. Math. Statist., Beachwood, OH, 2004, pp. 69–77.

[52] Janos Galambos, *Extreme value theory for applications*, Proceedings of the Conference on Extreme Value Theory and Applications, Volume 1 Gaithersburg Maryland 1993, Springer, 1994, pp. 1–14.

[53] Lenka Glavaš, Jelena Jocković, and Pavle Mladenović, *Maximum of the sum of consecutive terms in random permutations*, Journal of Statistical Planning and Inference **197** (2018), 15–20.

[54] Boris Gnedenko, *Sur la distribution limite du terme maximum d'une série aléatoire*, Ann. of Math. (2) **44** (1943), 423–453.

[55] Emil J. Gumbel, *Statistics of extremes*, Columbia university press, 1958.

[56] Peter Hall and Christopher C. Heyde, *Martingale limit theory and its application*, Academic press, 2014.

[57] L. H. Harper, *Stirling behavior is asymptotically normal*, Annals of Mathematical Statistics **38** (1967), 410–414.

[58] Jimmy He, *A central limit theorem for descents of a Mallows permutation and its inverse*, Annales de l'Institut Henri Poincaré Probabilités et Statistiques **58** (2022), no. 2, 667–694.

[59] Klaus Herrmann, Marius Hofert, and Johanna G. Neslehova, *Limiting behavior of maxima under dependence*, arXiv preprint arXiv:2405.02833 (2024).

[60] Lisa Hofer, *A central limit theorem for vincular permutation patterns*, Discrete Mathematics & Theoretical Computer Science **19** (2018), no. 2, 1–26.

[61] Rob A. Holman, *Extreme value statistics for wave run-up on a natural beach*, Coastal Engineering **9** (1986), no. 6, 527–544.

[62] Jürg Hüsler, *On applications of extreme value theory in optimization*, Experimental Methods for the Analysis of Optimization Algorithms, Springer, 2010, pp. 185–207.

[63] Mikhail Isaev, Igor Rodionov, Rui-Ray Zhang, and Maksim Zhukovskii, *Extremal independence in discrete random systems*, arXiv preprint arXiv:2105.04917 (2021).

[64] Svante Janson, *Normal convergence by higher semi-invariants with applications to sums of dependent random variables and random graphs*, The Annals of Probability **16** (1988), no. 1, 305–312.

[65] Svante Janson, Luca Pratelli, and Pietro Rigo, *Quantitative bounds in the central limit theorem for m-dependent random variables*, Latin American Journal of Probability and Mathematical Statistics **21** (2024), no. 1, 245–265.

[66] Thomas Kahle and Christian Stump, *Counting inversions and descents of random elements in finite Coxeter groups*, Mathematics of Computation **89** (2020), no. 321, 437–464.

[67] Gene B. Kim and Sangchul Lee, *Central limit theorem for descents in conjugacy classes of S_n*, Journal of Combinatorial Theory. Series A **169** (2020), 105–123.

[68] Yuta Koike, *Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles*, Japanese Journal of Statistics and Data Science **4** (2021), no. 1, 257–297.

[69] Samuel Kotz and Saralees Nadarajah, *Extreme value distributions: theory and applications*, World Scientific, 2000.

[70] Arun K. Kuchibhotla and Alessandro Rinaldo, *High-dimensional CLT for Sums of Non-degenerate Random Vectors: $n^{-1/2}$-rate*, arXiv preprint arXiv:2009.13673 (2020).

[71] Malcolm R. Leadbetter, Georg Lindgren, and Holger Rootzén, *Extremes and related properties of random sequences and processes*, Springer Science & Business Media, 1983.

[72] François Longin, *Extreme events in finance: A handbook of extreme value theory and its applications*, John Wiley & Sons, 2016.

[73] Miles E. Lopes, Zhenhua Lin, and Hans-Georg Müller, *Bootstrapping max statistics in high dimensions: near-parametric rates under weak variance decay and application to functional and multinomial data*, The Annals of Statistics **48** (2020), no. 2, 1214–1229.

[74] Colin L. Mallows, *Non-null ranking models. I*, Biometrika **44** (1957), 114–130.

[75] Kathrin Meier and Christian Stump, *Central limit theorems for generalized descents and generalized inversions in finite root systems*, Electronic Journal of Probability **28** (2023), Paper No. 1.

[76] John P. Mills, *Table of the ratio: area to bounding ordinate, for any portion of normal curve*, Biometrika (1926), 395–400.

[77] Kosto V. Mitov, *A limit theorem for the maximum of random variables with logarithmic distribution*, Mathematics and Education in Mathematics **50** (2021), 179–184.

[78] Pavle Mladenović, *Limit theorems for the maximum terms of a sequence of random variables with marginal geometric distributions*, Extremes **2** (1999), no. 4, 405–419.

[79] Pavle Mladenović, *A note on random permutations and extreme value distributions*, Proceedings Japan Academy Series A. Mathematical Sciences **78** (2002), no. 8, 157–160.

[80] Ekaterina Morozova and Vladimir Panov, *Extreme Value Analysis for Mixture Models with Heavy-Tailed Impurity*, Mathematics **9** (2021), no. 18, 2208.

[81] Saralees Nadarajah and Kosto Mitov, *Asymptotics of maxima of discrete random variables*, Extremes **5** (2002), no. 3, 287–294.

[82] Alperen Özdemir, *Martingales and descent statistics*, Advances in Applied Mathematics **140** (2022), 1–34.

[83] JE Paguyo, *Fixed points, descents, and inversions in parabolic double cosets of the symmetric group*, arXiv preprint arXiv:2112.07728 (2021).

[84] Kyle Petersen, *Eulerian numbers*, Eulerian Numbers, Springer, 2015, pp. 3–18.

[85] Valentin V. Petrov, *Asymptotic behavior of probabilities of large deviations*, Theory of Probability & Its Applications **13** (1968), no. 3, 408–420.

[86] Valentin V. Petrov and John Robinson, *Large deviations for sums of independent non identically distributed random variables*, Communications in Statistics—Theory and Methods **37** (2008), no. 18, 2984–2990.

[87] John Pike, *Convergence rates for generalized descents*, Electronic Journal of Combinatorics **18** (2011), no. 1, Paper 236, 1–14.

[88] Jim Pitman, *Probabilistic bounds on the coefficients of polynomials with only real zeros*, Journal of Combinatorial Theory. Series A **77** (1997), no. 2, 279–303.

[89] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2023.

[90] M. M. Rao and R. J. Swift, *Probability theory with applications*, second ed., Mathematics and Its Applications (Springer), vol. 582, Springer, New York, 2006.

[91] Sidney I. Resnick, *Extreme values, regular variation, and point processes*, vol. 4, Springer Science & Business Media, 2008.

[92] Paul Ressel, *Stable tail dependence functions – some basic properties*, Dependence Modeling **10** (2022), no. 1, 225–235.

[93] Marco Rocco, *Extreme value theory in finance: A survey*, Journal of Economic Surveys **28** (2014), no. 1, 82–108.

[94] Frank Röttger, *Asymptotics of a locally dependent statistic on finite reflection groups*, Electronic Journal of Combinatorics **27** (2020), no. 2, Paper No. 2.24, 1–11.

[95] Martin Rubey, Christian Stump, et al., *FindStat - The combinatorial statistics database*, `http://www.FindStat.org`, Accessed: June 5, 2025.

[96] Magid Sabbagh, *Copula models and pickands dependence functions*, McGill University (Canada), 2019.

[97] Carla D. Savage and Mirkó Visontai, *The* $\mathbf{s}$*-Eulerian polynomials have only real roots*, Transactions of the American Mathematical Society **367** (2015), no. 2, 1441–1466.

[98] Takaaki Shimura, *Discretization of distributions in the maximum domain of attraction*, Extremes **15** (2012), no. 3, 299–317.

[99] Masaaki Sibuya, *Bivariate extreme statistics. I*, Annals of the Institute of Statistical Mathematics **11** (1960), 195–210.

[100] Richard Stanley, *Eulerian partitions of a unit hypercube*, Higher Combinatorics **31** (1977), 49.

[101] Maxwell Sun, *A Central Limit Theorem on Two-Sided Descents of Mallows Distributed Elements of Finite Coxeter Groups*, arXiv preprint arXiv:2406.07597 (2024).

[102] Wenpin Tang, *Mallows ranking models: maximum likelihood estimate and regeneration*, International Conference on Machine Learning, PMLR, 2019, pp. 6125–6134.

[103] Maud Thomas, Magali Lemaitre, Mark L. Wilson, Cécile Viboud, Youri Yordanov, Hans Wackernagel, and Fabrice Carrat, *Applications of extreme value theory in public health*, PloS one **11** (2016), no. 7, 1–7.

[104] Aad W. van der Vaart, *Asymptotic statistics*, vol. 3, Cambridge university press, 2000.

[105] V. A. Vatutin, *The numbers of ascending segments in a random permutation and in one inverse to it are asymptotically independent*, Diskretnaya Matematika **8** (1996), no. 1, 41–51.

[106] Hermann Weyl, *The classical groups: their invariants and representations*, vol. 45, Princeton university press, 1946.

[107] Mengjia Yu and Xiaohui Chen, *A robust bootstrap change point test for high-dimensional location parameter*, Electronic Journal of Statistics **16** (2022), no. 1, 1096–1152.