A Survey on the method of moments

Werner Kirsch
Fakultät für Mathematik und Informatik
FernUniversität in Hagen
Germany

October 1, 2015
# Contents

1 Introduction .................................................. 5

2 Weak Convergence ............................................. 7
   2.1 Measures, Integrals and Function Spaces .................. 7
   2.2 Convergence of measures ................................... 9
   2.3 Tightness and Prohorov's Theorem ......................... 13
   2.4 Separating classes of functions ............................ 18
   2.5 Moments and weak convergence ............................. 22
   2.6 Random Variables and Their Distribution ................. 29

3 Independent Random Variables ................................ 33
   3.1 Warm-up: A Law of large numbers ......................... 33
   3.2 Some combinatorics ....................................... 34
   3.3 The central limit theorem .................................. 41
   3.4 More on the law of large numbers ......................... 44

4 Exchangeable Random Variables ............................... 51
   4.1 Basics .................................................... 51
   4.2 A Law of Large Numbers ................................ 54
   4.3 Random variables of de-Finetti-type ....................... 56
   4.4 Sequences of de-Finetti-type .............................. 59
   4.5 The Theorem of de Finetti ................................ 60

5 The Curie-Weiss Model ......................................... 63
   5.1 Introduction .............................................. 63
   5.2 Curie-Weiss as a model of de Finetti type ................ 67
   5.3 Laplace's method ........................................ 71
   5.4 Correlations for the Curie-Weiss model .................... 78
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5</td>
<td>The law of large numbers for the Curie-Weiss model</td>
<td>82</td>
</tr>
<tr>
<td>5.6</td>
<td>Central limit theorems for the Curie-Weiss model</td>
<td>84</td>
</tr>
<tr>
<td>5.7</td>
<td>Fluctuation of the magnetization for large $\beta$</td>
<td>91</td>
</tr>
</tbody>
</table>
1 Introduction

In this paper we give a survey on a variety of aspects of probability theory emphasizing the method of moments. If \((\Omega, \mathcal{F}, \mathbb{P})\) is a probability space and \(\mathbb{E}(X)\) denotes the expectation value of the random variable \(X\) then the \(k^{th}\) moment of \(X\) is defined by

\[ m_k = m_k(X) := \int X^k \, d\mathbb{P} = \mathbb{E}(X^k). \]

Many results in probability concern the convergence of random variables \(X_N\) as \(N \to \infty\). For example the classical Central Limit Theorem says that the random variables

\[ \Sigma_N := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \]

converge in distribution to a standard normal distribution, provided the \(X_i\) are independent, identically distributed and have expectation 0 and variance 1.

In the context of the Central Limit Theorem convergence in distribution can be rephrased as

\[ \mathbb{P}(\Sigma_N \leq x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \, dt \quad \text{for all } x \in \mathbb{R} \]

The method of moments is a way to prove convergence in distribution by showing that the corresponding moments converge. In this work we investigate under which conditions such a conclusion is correct and give a number of examples for which this method can be applied successfully.

Chapter 2 is an introduction to the theory of weak convergence of measures and, thus, to convergence in distribution. In this chapter we introduce the method of moments and derive (weak) conditions under which it can be applied.

Chapter 3 is devoted to classical limit theorems for independent, identically distributed (i.i.d.) random variables, such as the Law of Large Numbers and the Central Limit Theorem. In connection with the latter theorem we also introduce and discuss some combinatorial results which will play a role in the rest of the paper.

In Chapter 4 we discuss a generalization of i.i.d. random variables, namely exchangeable random variables. A sequence of random variables is called exchangeable if their finite dimensional distributions are invariant under permutation of the variables. Among others, we prove a Law of Large Numbers for such random
variables, and prove de Finetti’s theorem which gives an interesting and somewhat surprising representation theorem for a certain class of exchangeable sequences.

The last chapter, Chapter 5 deals with a model which comes from statistical physics: The Curie-Weiss Model. This model was introduced to describe magnetism and in fact is able to predict a phase transition (from paramagnetism to ferromagnetism). The same system can also be used to describe the behavior of voters.

We prove various limit theorems for the Curie-Weiss Model. The model depends on a parameter $\beta$ which can be interpreted in statistical mechanics as an ‘inverse temperature’. Our results show that the limit behavior of the Curie-Weiss Model depends qualitatively on this parameter $\beta$. In fact, it changes drastically at the value $\beta = 1$. Such a behavior is known as a ‘phase transition’ in statistical physics. In terms of voting models this can be interpreted as a sudden change of voting behavior if the interaction among voters exceeds a certain value.
2 Weak Convergence

2.1 Measures, Integrals and Function Spaces

By a bounded measure on $\mathbb{R}^d$ we always mean a positive measure $\mu$ on the Borel-$\sigma$-Algebra $\mathcal{B}(\mathbb{R}^d)$ with $\mu(\mathbb{R}^d) < \infty$. A probability measure $\mu$ on $\mathbb{R}^d$ is a bounded measure with $\mu(\mathbb{R}^d) = 1$.

**Examples 2.1.**

1. For any $x_0 \in \mathbb{R}^d$ (and $A \in \mathcal{B}(\mathbb{R}^d)$)
   \[\delta_{x_0}(A) = \begin{cases} 1, \text{ if } x_0 \in A; \\ 0, \text{ otherwise.} \end{cases}\]
   
defines a probability measure on $\mathbb{R}^d$, the Dirac measure in $x_0$.

2. If $\lambda$ denotes the Lebesgue measure on $\mathbb{R}^d$, $f \in L^1(\mathbb{R}^d, \lambda)$ and $f \geq 0$, then
   \[\mu = f \, d\lambda(A) := \int_A f(x) \, d\lambda(x)\]
   defines a bounded measure $\mu$ on $\mathcal{B}(\mathbb{R}^d)$. The function $f$ is called the Lebesgue density of $\mu$.

If $\mu$ is a bounded measure on $\mathbb{R}^d$ and $f$ a bounded continuous function on $\mathbb{R}^d$, the integral
\[I_\mu(f) := \int_{\mathbb{R}^d} f(x) \, d\mu(x)\]
(1)
is well defined.

**Definition 2.2.** We denote the set of bounded continuous functions on $\mathbb{R}^d$ by $C_b(\mathbb{R}^d)$.

Equipped with the sup-norm
\[\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)|\]
(2)
$C_b(\mathbb{R}^d)$ is a Banach space.
**Proposition 2.3.** For any bounded measure $\mu$ on $\mathbb{R}^d$ the mapping 
$I_\mu : C_b(\mathbb{R}^d) \rightarrow \mathbb{R}$, defined in (1), is a bounded, positive linear functional on $C_b(\mathbb{R}^d)$, i.e.

1. $I_\mu : C_b(\mathbb{R}^d) \rightarrow \mathbb{R}$ is a linear mapping:
   \[ I_\mu(\alpha f + \beta g) = \alpha I_\mu(f) + \beta I_\mu(g) \]  
   for all $\alpha, \beta \in \mathbb{R}$, and $f, g \in C_b(\mathbb{R}^d)$.
2. $I_\mu$ is positive: If $f \geq 0$, then $I_\mu(f) \geq 0$
3. $I_\mu$ is bounded: There is a constant $C < \infty$ such that for all $f \in C_b(\mathbb{R}^d)$
   \[ |I_\mu(f)| \leq C \|f\|_{\infty}. \]  
   In fact, we may choose $C = \mu(\mathbb{R}^d)$.

If $I_\mu(f) = I_\nu(f)$ for all $f \in C_b(\mathbb{R}^d)$, then $\mu = \nu$.

It is an interesting and useful fact that a converse to Proposition 2.3 is true.

**Theorem 2.4.** (Riesz representation theorem) If $I : C_b(\mathbb{R}^d) \rightarrow \mathbb{R}$ is a bounded, positive linear functional then there exists a bounded measure $\mu$ on $\mathbb{R}^d$, such that

$I(f) = I_\mu(f) = \int f(x) \, d\mu(x)$  
for all $f \in C_b(\mathbb{R}^d)$  
(5)

For a proof of this theorem see e.g. [8].

In the sense of Theorem 2.4 we may (and will) identity measures $\mu$ and the corresponding (bounded, positive) linear functionals $I_\mu$.

The Banach space $C_b(\mathbb{R}^d)$ has a rather unpleasant feature: It is not separable, i.e. it contains no countable dense subset. To prove this and for further use below we introduce a useful class of functions in $C_b(\mathbb{R}^d)$.

**Definition 2.5.** For $L > R \geq 0$ we define the function $\phi_{R,L} : \mathbb{R} \rightarrow [0, 1]$ by:

$\phi_{R,L}(x) = \begin{cases} 
1, & \text{for } |x| \leq R; \\
\frac{L-|x|}{L-R}, & \text{for } R < |x| < L; \\
0, & \text{for } |x| \geq L.
\end{cases}$  
(6)

We also define (for any dimension $d$) a function $\Phi_{R,L} : \mathbb{R}^d \rightarrow [0, 1]$ by

$\Phi_{R,L}(x) = \phi_{R,L}(\|x\|)$  
(7)
Remark 2.6. \( \Phi_{R, L} \) is a continuous function which is equal to 1 on the closed ball \( \overline{B_R} = \{ x \mid \| x \| \leq R \} \) and has support in \( \overline{B_L} \).

We remind the reader that the support \( \text{supp} f \) of a function \( f \) is the closure of the set \( \{ x \mid f(x) \neq 0 \} \).

Proposition 2.7. There is no countable dense subset in \( C_b(\mathbb{R}^d) \).

Remark 2.8. We recall that a metric space is called separable if it contains a countable dense subset. Proposition 2.7 states that \( C_b(\mathbb{R}^d) \) is not separable.

Proof: We construct an uncountable family \( \mathcal{F} \) of functions in \( C_b(\mathbb{R}^d) \) with the property that \( f, g \in \mathcal{F}, f \neq g \) implies \( \| f - g \|_\infty \geq 1 \). Set \( f_y(x) = \Phi_{\frac{1}{10^d}, \frac{1}{10^d}}(x - y) \) then \( \text{supp} f_i \cap \text{supp} f_j = \emptyset \) if \( i, j \in \mathbb{Z}^d \) and \( i \neq j \).

We set \( M = \{-1, 1\}^{\mathbb{Z}^d} \) and for \( m \in M \)

\[
F_m(x) = \sum_{i \in \mathbb{Z}^d} m_if_i(x)
\]

Equation (8)

\( F_m \) is a bounded continuous function and \( F_m(i) = m_i \), hence for \( m \neq m' \)

\[
\|F_m - F_{m'}\|_\infty \geq \sup_{i \in \mathbb{Z}^d} \left| m_i - m'_i \right| \geq 2
\]

Equation (9)

Since \( M \) is uncountable this proves the claim. \( \square \)

2.2 Convergence of measures

We are now going to define a suitable notion of convergence for bounded measures.

The first idea one may come up with is what we could call “pointwise” convergence: \( \mu_n \to \mu \) if \( \mu_n(A) \to \mu(A) \) for all \( A \in \mathcal{B}(\mathbb{R}^d) \). This notion of convergence is known as “strong convergence”.

It turns out that demanding \( \mu_n(A) \to \mu(A) \) for all Borel sets is a too strong requirement for most purposes. For example, it is quite desirable that \( \delta_{\frac{1}{n}}[-1, 0] \) should converge to \( \delta_0 \) as \( n \to \infty \). However, \( \delta_{\frac{1}{n}}([-1, 0]) = 0 \) but \( \delta_0([-1, 0]) = 1 \).
So, instead of asking that the measures \( \mu_n \) converge if applied to a Borel set, we require that the linear functionals \( I_{\mu_n} \) converge, if applied to a bounded continuous function, more precisely:

**Definition 2.9.** (weak convergence): Let \( \mu_n \) and \( \mu \) be bounded measures on \( \mathbb{R}^d \). We say that \( \mu_n \) converges weakly to \( \mu \) (denoted \( \mu_n \rightharpoonup \mu \)) if

\[
\int f(x) \, d\mu_n(x) \to \int f(x) \, d\mu(x)
\]

for all \( f \in C_b(\mathbb{R}^d) \).

**Remarks 2.10.**
1. If \( \mu_n \rightharpoonup \mu \) then \( \mu(\mathbb{R}^d) = \lim_{n \to \infty} \mu_n(\mathbb{R}^d) \)
2. \( \delta_\frac{1}{n} \rightharpoonup \delta_0 \)

As the names suggest, weak convergence is "weaker" than strong convergence, in fact one can prove:

**Theorem 2.11** (Portemanteau). If \( \mu_n \) and \( \mu \) are bounded measures on \( \mathbb{R}^d \), then \( \mu_n \rightharpoonup \mu \) if and only if \( \mu_n(A) \to \mu(A) \) for all Borel sets \( A \) with the property \( \mu(\partial A) = 0 \).

**Notation 2.12.** If \( A \subset \mathbb{R}^d \) we denote by \( \overline{A} \) the closure, by \( A^0 \) the inferior and by \( \partial A = \overline{A} \setminus A^0 \) the boundary of \( A \).

There is an extension to this result:

**Theorem 2.13.** If \( \mu_n \) and \( \mu \) are bounded measures on \( \mathbb{R}^d \), then \( \mu_n \rightharpoonup \mu \) if and only if

\[
\int f(x) \, d\mu_n(x) \to \int f(x) \, d\mu
\]

for all bounded measurable function \( f \) with the property that there is a set \( A \) with \( \mu(A) = 0 \) such that \( f \) is continuous for all \( x \in \mathbb{R}^d \setminus A \).

We will not use Theorems 2.11 and 2.13 in the following. For proofs we refer to [7], Theorem 13.16.

**Proposition 2.14.** Suppose \( \mu_n \) is a sequence of bounded measures on \( \mathbb{R}^d \) and that

\[
\lim_{n \to \infty} \int f(x) \, d\mu_n(x) = I(f)
\]

exists for all \( f \in C_b(\mathbb{R}^d) \), then there is a bounded measure \( \mu \) such that

\[
\mu_n \rightharpoonup \mu
\]
Proof: It is easy to see that the limit $I(f)$ defines a positive bounded linear functional on $C_b(\mathbb{R}^d)$. Thus, by Theorem 2.4, there is a measure $\mu$ such that

$$I(f) = \int f(x) \, d\mu(x)$$  \hspace{1cm} (13)

and $\mu_n \Rightarrow \mu$. 

For technical reasons we introduce yet another notion of convergence. To do so we introduce a subspace of $C_b(\mathbb{R}^d)$.

Definition 2.15. We set

$$C_0(\mathbb{R}^d) = \{ f | f \text{ is a continuous function on } \mathbb{R}^d \text{ with compact support} \}$$  \hspace{1cm} (14)

$C_0(\mathbb{R}^d)$ is a subspace of $C_b(\mathbb{R}^d)$. Its closure in $C_b(\mathbb{R}^d)$ (with respect to the $\| \cdot \|_\infty$-norm) is

$$C_\infty(\mathbb{R}^d) = \{ f | f \text{ continuous and } \lim_{|x| \to \infty} f(x) = 0 \}. \hspace{1cm} (15)$$

Remark 2.16. We haven’t specified whether the function in $C_b(\mathbb{R}^d)$ etc. are real or complex valued. It should be clear from the respective context which alternative is meant.

We mention that for a complex valued function $f$ with real part $f_1$ and imaginary part $f_2$ we set (as usual)

$$\int f(x) \, d\mu(x) = \int (f_1(x) + if_2(x)) \, d\mu(x)$$

$$= \int f_1(x) \, d\mu(x) + i \int f_2(x) \, d\mu(x).$$

In contrast to $C_b(\mathbb{R}^d)$ the space $C_0(\mathbb{R}^d)$ contains a countable dense subset (with respect to the $\| \cdot \|_\infty$-norm).

Definition 2.17 (vague convergence). Let $\mu_n$ and $\mu$ be bounded measures on $\mathbb{R}^d$. We say that the sequence $\mu_n$ converges vaguely to $\mu$ (in symbols $\mu_n \vDash \mu$) if

$$\int f(x) \, d\mu_n(x) \to \int f(x) \, d\mu$$  \hspace{1cm} (16)

for all $f \in C_0(\mathbb{R}^d)$. 

11
Remark 2.18.  1. Vague convergence can be defined for all Borel measures \( \mu \) with \( \mu(K) < \infty \) for all compact sets \( K \subset \mathbb{R}^d \). We will not need this extension in the following.

2. It is clear that weak convergence implies vague convergence. The converse is not true: The sequence \( \mu_n = \delta_n \) converges vaguely to the "zero-measure" \( \mu(A) = 0 \) for all \( A \in \mathcal{B}(\mathbb{R}^d) \), but it doesn’t converge weakly. In a sense the measures \( \mu_n = \delta_n \) "run off to infinity". To avoid this one may require we will introduce the notion of ‘tightness’ in the following Section 2.3.

In Proposition 2.7, we learned that the space \( C_b(\mathbb{R}^d) \) is not separable, a property that makes it somewhat hard to deal with that space. One reason to introduce \( C_0(\mathbb{R}^d) \) is that this space is separable as the following proposition shows.

**Proposition 2.19.** There is a countable set \( D_0 \) in \( C_0(\mathbb{R}^d) \) such that for any \( f \in C_0(\mathbb{R}^d) \) there is a sequence \( f_n \in D_0 \) such that

\[
\|f - f_n\|_{\infty} \to 0 \quad \text{as} \quad n \to \infty
\]  

(17)

**Proof:** We define \( \mathcal{P}_Q \) to be the set of all polynomials on \( \mathbb{R}^d \) with rational coefficients. This set is countable.

We define

\[
D_0 = \left\{ \Psi : \mathbb{R}^d \to \mathbb{R} \mid \Psi(x) = P(x) \Phi_{L,2L}(x) \right\}
\]  

for some \( P \in \mathcal{P}_Q \) and some \( L \in \mathbb{N} \).  

(18)

Now, take \( f \in C_0(\mathbb{R}^d) \) and \( L \in \mathbb{N} \) so large that \( \text{supp } f \subset \overline{B}_L \), hence \( f = f \cdot \Phi_{L,2L} \). By the Weierstrass approximation theorem there is a sequence \( P_n \in \mathcal{P}_Q \) such that

\[
\sup_{x \in \overline{B}_{2L}} |f(x) - P_n(x)| \to 0 \quad \text{as} \quad n \to \infty
\]  

(19)
Set \( f_n(x) = P_n(x)\Phi_{L,2L} \) then
\[
\|f - f_n\|_{\infty} = \sup_{x \in \mathbb{R}^d} |f(x) - f_n(x)| \\
= \sup_{x \in \overline{B}_{2L}} |f(x) - P_n(x)\Phi_{L,2L}(x)| \\
= \sup_{x \in \overline{B}_{2L}} |f(x) - P_n(x)| \to 0
\]

\[\square\]

### 2.3 Tightness and Prohorov’s Theorem

If \( \mu \) is a bounded measure on \( \mathbb{R}^d \) then the following is true:

For any \( \varepsilon > 0 \) there is a compact set \( K_\varepsilon \subset \mathbb{R}^d \) such that
\[
\mu(\mathbb{C}K_\varepsilon) < \varepsilon .
\]  
\[\text{(20)}\]

Indeed, if \( \overline{B}_N \) denotes the closed ball of radius \( N \) around the origin, then

\[
\overline{B}_N \subset \overline{B}_{N+1} \quad \text{and} \quad \bigcup \overline{B}_N = \mathbb{R}^d , \\
\text{thus} \quad \mu(\overline{B}_N) \to \mu(\mathbb{R}^d) < \infty .
\]

Hence
\[
\mu(\mathbb{C}\overline{B}_N) = \mu(\mathbb{R}^d) - \mu(\overline{B}_N) < \varepsilon \quad \text{provided} \ N \ \text{is large enough.}
\]

**Definition 2.20.** A sequence \( \mu_n \) of bounded measures is called **tight** if for any \( \varepsilon > 0 \) there exists a compact set \( K_\varepsilon \subset \mathbb{R}^d \) such that for all \( n \)
\[
\mu_n(\mathbb{C}K_\varepsilon) < \varepsilon
\]  
\[\text{(21)}\]

**Remark 2.21.**  
1. We have seen above that \( K_\varepsilon = K_\varepsilon^n \) with \[\text{(21)}\] exists for each \( n \), tightness requires that \( K_\varepsilon \) can be chosen independent of \( n \).

2. To prove tightness it is sufficient to check \[\text{(21)}\] for \( n > N_0 \). Since there are compact set \( K_j \) with
\[
\mu_j(\mathbb{C}K_j) < \varepsilon \quad \text{for} \ j = 1, \ldots, N_0
\]
we have that \( \tilde{K} := K \cup \bigcup_{j=1}^{N_0} K_j \) is compact and
\[
\mu_n(\mathbb{C}\tilde{K}) < \varepsilon \quad \text{for all} \ n .
\]
**Examples 2.22.**

1. The sequence $\mu_n = \delta_n$ is not tight.

2. The sequence $\mu_n = (1 - \frac{1}{n})\delta_n + \frac{1}{n}\delta_n$ is tight.

There is a useful tightness criterion involving moments of measure:

**Proposition 2.23.** Suppose $\mu_n$ is a sequence of bounded measures on $\mathbb{R}^d$ such that

$$\int \|x\|^k \, d\mu_n(x) \leq C < \infty \quad \text{for some } k > 0. \quad (22)$$

Then the sequence $\mu_n$ is tight.

**Proof:** We estimate

$$\mu_n(\mathbb{C}_R) = \int \chi_{\|x\| > R}(x) \, d\mu_n(x) \leq \int \frac{\|x\|^k}{R^k} \, d\mu_n(x) = \frac{C}{R^k}.$$

This implies tightness of the sequence $\mu_n$. \qed

**Proposition 2.24.** Suppose the sequence $\mu_n$ of bounded measures on $\mathbb{R}^d$ converges vaguely to the bounded measure $\mu$, i.e. $\mu_n \overset{\text{v}}{\rightarrow} \mu$ then:

1. $\mu(\mathbb{R}^d) \leq \lim \mu_n(\mathbb{R}^d)$

2. If

$$\mu(\mathbb{R}^d) \geq \lim \mu_n(\mathbb{R}^d) \quad (23)$$

then $\mu_n$ is tight.

**Remark 2.25.** It follows immediately from Proposition 2.24 that a weakly convergent sequence of measures is tight.

**Proof:**
1. Choose $\varepsilon > 0$. Then there exists $R > 0$ such that $\mu(\overline{B_R}) < \frac{\varepsilon}{2}$. So

$$\mu(\mathbb{R}^d) \leq \mu(\overline{B_R}) + \frac{\varepsilon}{2}$$

$$\leq \int \Phi_{R,R+1}(x) \, d\mu(x) + \frac{\varepsilon}{2}$$

$$\leq \int \Phi_{R,R+1}(x) \, d\mu_n(x) + \varepsilon \quad \text{for } n \text{ large enough (by (16))}$$

$$\leq \mu_n(\mathbb{R}^d) + \varepsilon$$

The statement follows by taking $\varepsilon \downarrow 0$.

2. For $\varepsilon > 0$ choose again $R > 0$ such that $\mu(\overline{B_R}) < \frac{\varepsilon}{2}$ holds. Then

$$\mu_n(\overline{B_{R+1}}) = \mu_n(\mathbb{R}^d) - \mu_n(B_{R+1})$$

$$\leq \mu_n(\mathbb{R}^d) - \int \Phi_{R,R+1}(x) \, d\mu_n(x)$$

using assumption (23) and the vague convergence of the $\mu_n$ to $\mu$ we get

$$\leq \mu(\mathbb{R}^d) - \int \Phi_{R,R+1}(x) \, d\mu(x) + \frac{\varepsilon}{2} \quad \text{for } n \text{ large enough}$$

$$\leq \mu(\overline{B_R}) + \frac{\varepsilon}{2} < \varepsilon$$

So we proved $\mu_n(\overline{B_{R+1}}) < \varepsilon$ for $n$ large enough, hence the sequence $\mu_n$ is tight (see Remark 2.21).

\[\square\]

**Corollary 2.26.** If $\mu_n \overset{\text{v}}{\rightharpoonup} \mu$ for bounded measures $\mu_n, \mu$ on $\mathbb{R}^d$ and if $\mu(\mathbb{R}^d) = \lim \mu_n(\mathbb{R}^d)$ then $\mu_n \Rightarrow \mu$.

**Proof:** Take $f \in C_b(\mathbb{R}^d)$ and $\varepsilon > 0$. By Proposition 2.24 there exists $R$ such that

$$\mu(\overline{B_R}) < \frac{\varepsilon}{3 \|f\|_\infty} \quad \text{and} \quad \mu_n(\overline{B_R}) < \frac{\varepsilon}{3 \|f\|_\infty} \quad \text{for all } n$$

(24)
Then

\[ \left| \int f(x) \, d\mu(x) - \int f(x) \, d\mu_n(x) \right| \]

\[= \left| \int f(x) \left( \Phi_{R,R+1}(x) + (1 - \Phi_{R,R+1}(x)) \right) \, d\mu(x) \right. \]

\[ \left. - \int f(x) \left( \Phi_{R,R+1}(x) + (1 - \Phi_{R,R+1}(x)) \right) \, d\mu_n(x) \right| \]

\[\leq \left| \int f(x) \Phi_{R,R+1}(x) \, d\mu(x) - \int f(x) \Phi_{R,R+1}(x) \, d\mu_n(x) \right| \]

\[ + \|f\|_\infty \mu_n(\overline{C_{B_R}}) + \|f\|_\infty \mu(\overline{C_{B_R}}) \]

\[\leq \left| \int f(x) \Phi_{R,R+1}(x) \, d\mu_n(x) - \int f(x) \Phi_{R,R+1}(x) \, d\mu_n(x) \right| + \frac{2}{3} \varepsilon \]

\[< \varepsilon \]

if \( n \) is large enough, since \( f \Phi_{R,R+1} \in C_0(\mathbb{R}^d) \).

\[\square\]

**Theorem 2.27.** Suppose \( \mu_n \) is a tight sequence of bounded measures on \( \mathbb{R}^d \). If \( \mu_n \) converges vaguely to some measure \( \mu \) then \( \mu_n \) converges weakly to \( \mu \).

**Remarks 2.28.** We are going to prove a slightly stronger statement, namely: If for a tight sequence \( \mu_n \) the integrals \( \int f \, d\mu_n \) converge to some \( I(f) \) for all \( f \in C_0(\mathbb{R}^d) \), then \( \mu_n \) converges weakly to some bounded measure \( \mu \) and \( I(f) = \int f(x) \, d\mu(x) \).

**Proof** (of Theorem 2.27): Take some \( f \in C_b(\mathbb{R}^d) \). To show that \( \int f(x) \, d\mu_n(x) \) converges it is enough to prove that the (real valued) sequence \( I_n(f) \) is a Cauchy sequence.

Now

\[|I_n(f) - I_k(f)| = \left| \int f(x) \, d\mu_n(x) - \int f(x) \, d\mu_k(x) \right| \]

\[\leq \|f\|_\infty \mu_n(\overline{C_{B_L}}) + \|f\|_\infty \mu_k(\overline{C_{B_L}}) \]

\[+ \left| \int f(x) \Phi_{L,2L}(x) \, d\mu_n(x) - \int f(x) \Phi_{L,2L}(x) \, d\mu_k(x) \right| \]
Since $\{\mu_n\}$ is tight we may take $L$ so large, that $\mu_j(\overline{B_L}) < \frac{\varepsilon}{3\|f\|_\infty}$ for all $j$.

Since $\int g(x) \, d\mu_n(x)$ is a Cauchy sequence for each $g \in C_0(\mathbb{R}^d)$ we may take $n, k$ so large that

$$\left| \int f(x) \Phi_{L,2L}(x) \, d\mu_n(x) - \int f(x) \Phi_{L,2L}(x) \, d\mu_k(x) \right| < \frac{\varepsilon}{3}.$$ 

This proves that for each $f \in C_0(\mathbb{R}^d)$ there is a (real number) $I(f)$ such that $I_n(f) \to I(f)$.

Since the functions $I_n(f)$ are linear in $f$ the same is true for $I(f)$. In the same way we get that $I(f) \geq 0$ if $f \geq 0$. Moreover,

$$|I_n(f)| \leq \mu_n(\overline{B_L}) \|f\|_\infty + \left| \int f(x) \Phi_{L,L+1}(x) \, d\mu_n(x) \right|$$

$$\leq \varepsilon \|f\|_\infty + \int f(x) \Phi_{L,L+1}(x) \, d\mu_n(x) \|f\|_\infty$$

$$\leq \left( \int f(x) \Phi_{L,L+1}(x) \, d\mu_n(x) + \varepsilon \right) \|f\|_\infty$$

$$\leq C \|f\|_\infty$$

For $n$ large enough we have

$$|I(f)| \leq 2 |I_n(f)| \leq 2 C \|f\|_\infty.$$

Thus $f \mapsto I(f)$ is a bounded positive linear functional, hence by Theorem 2.4 there is a $\mu$ with $I(f) = \int f(x) \, d\mu(x)$ and

$$\int f(x) \, d\mu_n(x) = I_n(f) \to I(f) = \int f(x) \, d\mu(x).$$

The following theorem is of fundamental importance for the theory of weak convergence, it says that tight sets are sequentially compact.

**Theorem 2.29** (Prohorov). Any tight sequence of bounded measures $\mu_n$ with $\mu_n(\mathbb{R}^d) \leq C < \infty$ for all $n$ has a weakly convergent subsequence.

**Proof:** By Theorem 2.27 and the remark following it we only have to find a subsequence $\mu_{n_k}$ such that for any $f \in C_0(\mathbb{R}^d)$

$$I_k(f) = \int f(x) \, d\mu_{n_k}(x)$$

(26)
converges.

Let $D_0 = \{f_1, f_2, \cdots\}$ be a countable dense subset of $C_0(\mathbb{R}^d)$ (see Proposition 2.19). Then the sequence $I_n(f_1) = \int f_1(x) \, d\mu_n(x)$ is bounded. Consequently there is a subsequence $\mu_n^{(1)}$ of $\mu_n$ such that $\int f_1(x) \, d\mu_n^{(1)}(x)$ converges.

By the same argument there is a subsequence $\mu_n^{(2)}$ of $\mu_n^{(1)}$ such that $\int f_2(x) \, d\mu_n^{(2)}(x)$ converges. In general, if we have a subsequence of $\mu_n^{(k-1)}$ such that $\int f_j(x) \, d\mu_n^{(k-1)}(x)$ converges for all $j \leq k - 1$ we find a subsequence $\mu_n^{(k)}$ of $\mu_n^{(k-1)}$ such that (also) $\int f_n(x) \, d\mu_n^{(k)}(x)$ converges.

We build the diagonal sequence

$$\tilde{\mu}_n = \mu_n^{(n)}$$

(27)

It follows that $\int f_k(x) \, d\tilde{\mu}_n(x)$ converges for all $k \in \mathbb{N}$. We claim that $\int f(x) \, d\tilde{\mu}_n(x)$ converges for all $f \in C_0(\mathbb{R}^d)$. In deed, we have

$$\left| \int f(x) \, d\tilde{\mu}_n(x) - \int f(x) \, d\tilde{\mu}_m(x) \right|$$

$$\leq \left| \int f(x) \, d\tilde{\mu}_n(x) - \int f(x) \, d\tilde{\mu}_n(x) \right|$$

$$+ \left| \int f_k(x) \, d\tilde{\mu}_n(x) - \int f_k(x) \, d\tilde{\mu}_m(x) \right|$$

$$+ \left| \int f_k(x) \, d\tilde{\mu}_m(x) - \int f(x) \, d\tilde{\mu}_m(x) \right|$$

$$\leq \| f - f_k \|_{\infty} (\tilde{\mu}_n(\mathbb{R}^d) + \tilde{\mu}_m(\mathbb{R}^d))$$

$$+ \left| \int f_k(x) \, d\tilde{\mu}_n(x) - \int f_k(x) \, d\tilde{\mu}_m(x) \right|$$

$$\leq 2C \| f - f_k \|_{\infty} + \left| \int f_k(x) \, d\tilde{\mu}_n(x) - \int f_k(x) \, d\tilde{\mu}_m(x) \right|$$

for some $C$.

Now choose $f_k \in D_0$ such that $\| f - f_k \|_{\infty} \leq \frac{\varepsilon}{4C}$ and then $n, m$ so large that $\left| \int f_k \, d\tilde{\mu}_n - \int f_k \, d\tilde{\mu}_m \right| < \frac{\varepsilon}{2}$.  \hfill $\Box$

### 2.4 Separating classes of functions

From Theorem 2.27 we know that for tight sequences $\mu_n$ vague convergence and weak convergence are equivalent. In other words: If we know that $\mu_n$ is tight then
it is enough to check convergence of \( \int f(x) \, d\mu_n(x) \) for all \( f \in \mathcal{D} = C_0(\mathbb{R}^d) \) rather than to check it on \( C_b(\mathbb{R}^d) \).

In this section we will try to identify other - preferably "smaller" - sets \( \mathcal{D} \) for which this assertion is true.

**Definition 2.30.** A set of functions \( \mathcal{D} \subset C_b(\mathbb{R}^d) \) is called *separating* (or a *separating class*) if

\[
\int f(x) \, d\mu(x) = \int f(x) \, d\nu(x) \tag{28}
\]

for all \( f \in \mathcal{D} \) implies

\[
\mu = \nu \tag{29}
\]

**Examples 2.31.**

1. The set \( C_0(\mathbb{R}^d) \) is separating.

2. If \( \mathcal{D} \) is a separating class and \( \mathcal{D}_0 \) is dense in \( \mathcal{D} \) with respect to the sup-norm then \( \mathcal{D}_0 \) is separating as well.

3. The set \( C_\infty(\mathbb{R}^d) \) of infinitely differentiable functions with compact support is separating.

4. The set \( \mathcal{D}_0 \) defined in the proof of Proposition 2.19 is separating.

5. \( \mathcal{D}_\infty = \{ f \mid f(x) = \Phi_{R,L}(x-a), a \in \mathbb{R}^d, R < L \in \mathbb{R} \} \) is separating.

The proofs of these facts are left to the reader.

The importance of separating classes is based on the following theorem.

**Theorem 2.32.** Suppose \( \mu_n \) is a tight sequence of bounded measures with \( \sup \mu_n(\mathbb{R}^d) < \infty \) and let \( \mathcal{D} \) be a separating class of functions. If \( \int f(x) \, d\mu_n \) converges for all \( f \in \mathcal{D} \) then the sequence \( \mu_n \) converges weakly to some bounded measure \( \mu \).

**Proof:** By Prohorov’s Theorem 2.29 we know that any subsequence of \( \mu_n \) has a weakly convergent subsequence. Let \( \nu_k \) and \( \rho_k \) be two weakly convergent subsequences of \( \mu_n \) and suppose \( \nu_k \Rightarrow \nu \) and \( \rho_n \Rightarrow \rho \).

Take \( f \in \mathcal{D} \). By assumption

\[
\int f(x) \, d\mu_n(x) \to a \quad \text{for some } a \in \mathbb{R}
\]
and consequently for the subsequences $\nu_k$ and $\rho_k$ we also have
\[ \int f(x) \, d\nu_k(x) \to a \quad \text{and} \quad \int f(x) \, d\rho_k(x) \to a \]

On the other hand we know
\[ \int f(x) \, d\nu_k(x) \to \int f(x) \, d\nu(x) \quad \text{and} \quad \int f(x) \, d\rho_k(x) \to \int f(x) \, d\rho(x) \]

Thus
\[ \int f(x) \, d\nu(x) = \int f(x) \, d\rho(x). \]

for all $f \in D$. Thus $\nu = \rho$.

Using the elementary lemma below we conclude that $\mu_n$ converges weakly. \qed

**Lemma 2.33.** Let $a_n$ be a sequence of real numbers with the property that any subsequence has a convergent subsequence. Suppose that any convergent subsequence of $a_n$ converges to $a$ independent of the subsequence chosen. Then the sequence $a_n$ converges itself to $a$.

The elementary proof is left to the reader.

An important separating class of functions is given by the trigonometric polynomials.

**Proposition 2.34.** The set $D_{\text{trig}} = \{ f_p \in C_c(\mathbb{R}^d) \mid f_p(x) = e^{ip \cdot x}, p \in \mathbb{R}^d \}$ is separating.

**Notation 2.35.** Here and in the following $i = \sqrt{-1}$ and $x \cdot y = \sum_{j=1}^d x_j y_j$ for $x, y \in \mathbb{R}^d$.

**Definition 2.36.** For any bounded measure $\mu$ we define the characteristic function $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ (Fourier transform) of $\mu$ by
\[ \hat{\mu}(p) = \int e^{ip \cdot x} \, d\mu(x) \quad (30) \]

Proposition 2.34 says that $\hat{\mu}(p) = \hat{\nu}(p)$ for all $p$ implies $\mu = \nu$.

For the proof of Proposition 2.34 we need a well known result by Fejér.
Definition 2.37. We set \( \Lambda_L = \{ x \in \mathbb{R}^d \mid |x_i| \leq L \text{ for } i = 1, \ldots, d \} \). A trigonometric polynomial on \( \Lambda_L \) is a (finite) linear combination of functions \( e^{i \pi L n \cdot x} \) with \( n \in \mathbb{Z}^d \).

We call a continuous function \( f \in C(\Lambda_L) \) periodic if
\[
f(x_1, \ldots, x_r, L, x_{r+1}, \ldots, x_d) = f(x_1, \ldots, x_r, -L, x_{r+1}, \ldots, x_d)
\]
for \( r = 1, \ldots, d \), i.e. if \( f \) is the restriction of a \( 2L \mathbb{Z}^d \)-periodic continuous function on \( \mathbb{R}^d \).

In particular, any trigonometric polynomial is periodic.

Theorem 2.38. (Fejé) A continuous periodic function \( f \) on \( \Lambda_L \) can be uniformly approximated by trigonometric polynomials on \( \Lambda_L \), i.e. there exists a sequence \( f_n \) of trigonometric polynomials on \( \Lambda_L \), such that
\[
\sup_{x \in \Lambda_L} |f(x) - f_n(x)| \to 0 \quad \text{as } n \to \infty \quad (31)
\]

A proof of Fejér’s Theorem can be found for example in [4]. There are also various proofs available in the internet, for instance in [11].

Proof (Proposition 2.34): Suppose \( \mu \) is a bounded measure and \( f \in C_0(\mathbb{R}^d) \). Take \( \varepsilon > 0 \) and \( L \) so large that \( \mu(\mathbb{R}^d \setminus \Lambda_L) < \frac{\varepsilon}{4(\|f\|_\infty + 1)} \).

Without loss we may assume that
\[
\text{supp } f \subset \Lambda_L \quad (32)
\]

By Theorem 2.38 there is a trigonometric polynomial \( T(x) = \sum_{|n| \leq M} a_n e^{i \pi L n \cdot x} \) such that \( \sup_{x \in \Lambda_L} |f(x) - T(x)| < \frac{\varepsilon}{4\mu(\mathbb{R}^d)} \) and \( \sup_{x \in \Lambda_L} |T(x)| \leq \|f\|_\infty + 1 \).

Since \( T \) is periodic we also have \( \|T\|_\infty \leq \|f\|_\infty + 1 \). Hence
\[
\left| \int_{\Lambda_L} f(x) \, d\mu(x) - \int_{\Lambda_L} T(x) \, d\mu(x) \right| \leq \int_{\Lambda_L} |f(x) - T(x)| \, d\mu(x) + \int_{\Lambda_L} |T(x)| \, d\mu(x)
\]
\[
\leq \mu(\mathbb{R}^d) \frac{\varepsilon}{4\mu(\mathbb{R}^d)} + \|T\|_\infty \frac{\varepsilon}{4(\|f\|_\infty + 1)}
\]
\[
< \varepsilon
\]

We conclude: Knowing \( \int_{\Lambda_L} f(x) \, d\mu(x) \) for trigonometric polynomials \( f \), allows us to compute the integral for \( f \in C_0(\mathbb{R}^d) \) and this determines \( \mu \). \( \square \)
2.5 Moments and weak convergence

**Definition 2.39.** For a probability measure $\mu$ on $\mathbb{R}$ and any $k \in \mathbb{N}$ we define the $k^{th}$ absolute moment $m_k(\mu)$ by

$$m_k(\mu) = \int |x|^k \, d\mu(x)$$

(33)

If $m_k(\mu) < \infty$ we call

$$m_k(\mu) = \int x^k \, d\mu(x)$$

(34)

the $k^{th}$ moment of $\mu$.

**Remarks 2.40.**

1. Observe that $m_{2k}(\mu) = m_{2k}(\mu)$.

2. If $m_k(\mu) < \infty$ then $m_l(\mu) < \infty$ for all $l \leq k$. This can be seen by applying either Hölder’s inequality or Jensen’s inequality.

3. The first moment $E(\mu) := m_1(\mu)$ is usually called the mean or the expectation of $\mu$ and the quantity $\mathbb{V}(\mu) := m_2(\mu) - m_1(\mu)^2$ is called the variance of $\mu$.

**Example 2.41.** The Dirac measure $\mu = \delta_{x_0}$ has the moments $m_k(\delta_{x_0}) = x_0^k$.

**Example 2.42.** The normal distribution $\mathcal{N}(\mu, \sigma^2)$ on $\mathbb{R}$ is defined through its density

$$n_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(35)

It easy to see that $m_1(\mathcal{N}(\mu, \sigma^2)) = \mu$ and $\mathbb{V}(\mathcal{N}(\mu, \sigma^2)) = \sigma^2$.

To express the higher moments of the normal distribution in a closed form, we introduce the following notation:

**Definition 2.43.** For $n \in \mathbb{N}$ we define:

$$n!! := \begin{cases} n \cdot (n-2) \cdot (n-4) \cdot \ldots \cdot 1 & \text{for } n \text{ odd} \\ n \cdot (n-2) \cdot (n-4) \cdot \ldots \cdot 2 & \text{for } n \text{ even} \end{cases}$$

(36)

The moments $m_k$ of the normal distribution with mean zero are given in the following proposition:
**Proposition 2.44.** The moments of the normal distribution \( \mathcal{N}(0, \sigma^2) \) are given by:

\[
m_k = m_k(\mathcal{N}(0, \sigma^2)) = \begin{cases} 
(k - 1)!! \sigma^{k/2} & \text{for } k \text{ even} \\
0 & \text{for } k \text{ odd}
\end{cases}
\]  

(37)

**Proof:** For \( k \) odd, the function \( x^k e^{-x^2/2} \) is odd, hence \( m_k = 0 \).

We have \( m_0 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \, dx = 1 \).

For arbitrary \( k \) even we compute, using integration by parts

\[
m_{k+2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{k+2} e^{-x^2/2} \, dx
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{k+1} \left( x e^{-x^2/2} \right) \, dx
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{k+1} \left( e^{-x^2/2} \right)' \, dx
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (x^{k+1})' e^{-x^2/2} \, dx
= (k + 1) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-x^2/2} \, dx
= (k + 1) m_k
\]

The assertion follows by induction. \( \square \)

For later use we rewrite the moments of \( \mathcal{N}(0, \sigma^2) \).

**Corollary 2.45.** The (even) moments of the distribution \( \mathcal{N}(0, 1) \) can be written as

\[
m_{2k} = \frac{(2k)!}{2^k k!}
\]

(38)

This follows immediately from (37) and the following Lemma.

**Lemma 2.46.** For \( n \in \mathbb{N} \) we have

\[
(2n)!! = 2^n \, n!
\]

(39)

\[
(2n - 1)!! = \frac{(2n)!}{n! \, 2^n}
\]

(40)

**Proof:** Equation (39) is immediate.
We prove equation (40) by induction:

\[
(2(n + 1) - 1)!! = (2n + 1) \cdot (2n - 1)!! \\
= (2n + 1) \cdot \frac{(2n)!}{n! \cdot 2^n} \\
= \frac{(2n + 1)!}{n! \cdot 2^n} \\
= \frac{(2(n + 1))!}{(n + 1)! \cdot 2^{n+1}}
\]

In many cases expressions like \( \int x^k \, d\mu(x) \) are much easier to analyze than integrals \( \int f(x) \, d\mu(x) \) for arbitrary \( f \in C_b(\mathbb{R}) \). Thus, it would be very helpful to have a criterion under which convergence of moments \( m_k(\mu_n) \) implies convergence.

The first and obvious problem is the fact that the functions \( p_k(x) = x^k \) are not in \( C_b(\mathbb{R}) \). In fact, \( \int p_k(x) \, d\mu(x) \) is well defined only under the additional assumption that \( m_k(\mu) < \infty \).

So, let us suppose for what follows, that \( \overline{m}_k(\mu) < \infty \) for all \( k \in \mathbb{N} \).

**Definition 2.47.** We say that a bounded measure \( \mu \) is a measure with existing moments if \( \overline{m}_k(\mu) = \int |x|^k \, d\mu < \infty \) for all \( k \in \mathbb{N} \).

**Notation 2.48.** Unless stated otherwise we assume from now on that all measures mentioned have existing moments!

**Proposition 2.49.** If \( m_k(\mu) < \infty \) for all even \( k \), then \( \mu \) is a measure with existing moments.

**Proof:** For even \( k \) we have \( \overline{m}_k(\mu) = m_k(\mu) \).

For the odd numbered moments we estimate using the Cauchy-Schwarz inequality

\[
\overline{m}_{2k-1}(\mu) = \int |x|^{2k-1} \, d\mu \\
\leq \left( \int |x|^{2k} \, d\mu \right)^{1/2} \left( \int |x|^{2k-2} \, d\mu \right)^{1/2} \\
= (\overline{m}_{2k})^{1/2} (\overline{m}_{2k-2})^{1/2}
\]

\[\square\]
Proposition 2.50. Suppose $\mu_n$ is a sequence of bounded measures with existing moments and
\[ \sup_n m_k(\mu_n) \leq C_k < \infty. \tag{41} \]
If $\mu_n \rightharpoonup \mu$ then all moments of $\mu$ exist and
\[ m_k(\mu_n) \to m_k(\mu) \quad \text{for all } k. \]

Proof: We estimate:
\[
\int |x|^k \phi_{L,2L}(x) \, d\mu(x) \leq \sup_n \int |x|^k \phi_{L,2L}(x) \, d\mu_n(x) \\
\leq \sup_n m_k(\mu_n) \\
\leq C_k.
\]
Thus
\[
m_k(\mu) = \sup_L \int |x|^k \phi_{L,2L}(x) \, d\mu(x) \leq C_k < \infty.
\]
So, all moments of $\mu$ exist.

To prove the second assertion of the proposition we estimate
\[
\left| \int |x|^k \, d\mu - \int |x|^k \, d\mu_n \right| \\
\leq \left| \int |x|^k \phi_{L,2L}(x) \, d\mu - \int |x|^k \phi_{L,2L}(x) \, d\mu_n \right| \\
+ \int_{|x| \geq L} |x|^k \, d\mu_n + \int_{|x| \geq L} |x|^k \, d\mu.
\]
The first expression in the above sum goes to zero due to weak convergence. The second summand can be estimated by
\[
\int_{|x| \geq L} |x|^k \, d\mu_n \leq \left( \int |x|^{2k} \, d\mu_n \right)^{1/2} \left( \mu_n(|x| \geq L) \right)^{1/2} \\
\leq C_k^{1/2} \mu_n(|x| \geq L)^{1/2}
\]
Due to the tightness of the sequence $\mu_n$, the last expression can be made small by choosing $L$ large enough. A similar argument works for $\int_{|x| \geq L} |x|^k \, d\mu$. □

Under the assumption of existing moments, the reasoning of the last chapters could make us optimistic that convergence of moments, in deed, implies weak convergence of measures.
We have already proved in Proposition 2.23 that the convergence (hence boundedness) of the second moments implies tightness of the sequence. Thus the only ‘thing’ we have to ‘check’ is whether the polynomials form a separating class of functions, in other words we have to ‘prove’ that \( \int x^k \, d\mu(x) = \int x^k \, d\nu(x) \) for all \( k \in \mathbb{N} \) implies \( \mu = \nu \). Unfortunately, this assertion is wrong unless we impose some further condition on the probability measure \( \mu \).

**Definition 2.51.** We say that a bounded measure \( \mu \) on \( \mathbb{R} \) has **moderately growing moments** if all moments exist and

\[
\bar{m}_k(\mu) \leq A C^k k!
\]

for some constant \( A, C \) and all \( k \in \mathbb{N} \).

**Lemma 2.52.** It is sufficient to postulate (42) only for the even moments.

**Proof:**

\[
\begin{align*}
\bar{\mu}_{2k-1} &= \int |x|^{2k-1} \, d\mu \\
&\leq \left( \int |x|^{2k} \, d\mu \right)^{1/2} \left( \int |x|^{2k-2} \, d\mu \right)^{1/2} \quad \text{(by Cauchy-Schwarz)} \\
&\leq A C^{2k-1} (2k)!^{1/2} (2k-2)!^{1/2} \\
&\leq A C^{2k-1} (2k-2)! (2k-1)^{1/2} (2k)^{1/2} \\
&\leq 2 A C^{2k-1} (2k-1)!
\end{align*}
\]

\[\square\]

**Examples 2.53.**

1. If \( \mu \) has compact support, then \( \mu \) is a measure with moderately growing moments.

2. The normal distribution is a measure with moderately growing moments.

**Theorem 2.54.** Suppose \( \mu \) is a bounded measure with moderately growing moments (satisfying (42), then:
1. The characteristic function

\[ \hat{\mu}(z) = \int e^{izx} \, d\mu(x) \]  

(43)

is well defined on the strip

\[ S = \{ z = z_1 + iz_2 \in \mathbb{C} \mid |z_2| < \frac{1}{C} \} \]  

(44)

in the complex plane \( \mathbb{C} \). Here \( C \) is the constant from equation (42).

2. The function \( \hat{\mu} : S \to \mathbb{C} \) is analytic.

Proof:

1. For \( z = z_1 + i z_2 \) we estimate

\[ |\hat{\mu}(z)| \leq \int |e^{izx}| \, d\mu(x) \]

\[ \leq \int e^{\|z\|_2 x} \, d\mu(x) \]

\[ = \int \sum_{k=0}^{\infty} \frac{|z_2|^k}{k!} |x|^k \, d\mu(x) \]

\[ = \sum_{k=0}^{\infty} \frac{|z_2|^k}{k!} m_k(\mu) \]

\[ \leq A \sum_{k=0}^{\infty} |z_2|^k C^k \]

\[ < \infty \quad \text{for } |z_2| < C^{-1} \]

Thus \( \hat{\mu} \) is well defined on \( S \).

2. To prove that \( \hat{\mu} \) is analytic on \( S \) we develop it in a power series around an arbitrary \( z_0 \in \mathbb{R} \). Suppose \( |\zeta| < C^{-1} \), then

\[ \hat{\mu}(z_0 + \zeta) = \int e^{iz_0x} \sum_{k=0}^{\infty} \frac{\zeta^k}{k!} x^k \, d\mu(x) \]

\[ = \sum_{k=0}^{\infty} \left( \int \frac{e^{iz_0x} x^k \, d\mu}{k!} \right) \zeta^k \]
Interchanging sum and integral is justified by the estimate
\[ \left| \int e^{i\alpha x} \frac{x^k}{k!} d\mu(x) \right| \leq \int |x|^k \frac{d\mu(x)}{k!} \leq A C^k. \]

Above \( C \) denotes the constant from \((42)\). \qed

**Theorem 2.55.** Suppose \( \mu \) is a bounded measure with moderately growing moments. If for some bounded measure \( \nu \)
\[ m_k(\mu) = m_k(\nu) \quad \text{for all } k \in \mathbb{N} \]
then
\[ \mu = \nu \]

**Proof:** Since for all even \( k \) we have \( m_k(\nu) = m_k(\mu) < \infty \) we have that all moments of \( \nu \) exist as well and are moderately growing.
Therefore both \( \hat{\mu} \) and \( \hat{\nu} \) are analytic functions in a strip
\[ S = \{ z \in \mathbb{C} \mid \text{Im} z < C^{-1} \}. \]
For \(|\zeta| < C^{-1}\) we have as in the proof of Theorem 2.54
\[ \hat{\mu}(\zeta) = \sum_{k=0}^{\infty} m_k(\mu) \zeta^k = \sum_{k=0}^{\infty} m_k(\nu) \zeta^k = \hat{\nu}(\zeta) \]
Thus, \( \hat{\mu} \) and \( \hat{\nu} \) agree on a neighborhood of \( z = 0 \) and hence in the whole strip \( S \). \qed

**Theorem 2.56.** Suppose \( \mu_n \) is a sequence of bounded measures and \( m_k \in \mathbb{R} \) is a sequence with
\[ m_k \leq A C^k k! \quad \text{for all } k \in \mathbb{N}. \]
If
\[ m_k(\mu_n) \to m_k \quad \text{for all } k \in \mathbb{N} \]
then there is a bounded measure \( \mu \) with \( m_k(\mu) = m_k \) and such that
\[ \mu_n \to \mu \]

28
Proof: By Proposition 2.23 we know that $\mu_n$ is tight. So any subsequence has a convergent subsequence. Suppose $\tilde{\mu}_n$ is a convergent subsequence. Call its limit $\mu$. According to Proposition 2.50 the measure $\mu$ has moments $m_k$. Due to assumption (48), $\mu$ has moderately growing moments, thus it is uniquely determined by the numbers $m_k$. It follows that any subsequence of the $\mu_n$ has a convergent subsequence with limit $\mu$, thus $\mu_n$ converges weakly to $\mu$. \qed

2.6 Random Variables and Their Distribution

Definition 2.57. Suppose $(\Omega, A, \mathbb{P})$ is a probability space and $X : \Omega \rightarrow \mathbb{R}^d$ a random variable (i.e. an $(A - B(\mathbb{R}^d))$-measurable function), then the probability measure $P_X$ on $B(\mathbb{R}^d)$ defined by

$$P_X(A) = \mathbb{P}(X \in A) := \mathbb{P}\{\omega \in \Omega \mid X(\omega) \in A\}$$

is called the distribution of $X$.

We say that the random variables $\{X_i\}_{i \in I}$ are identically distributed if the distributions $P_{X_i}$ are the same for all $i \in I$.

Notation 2.58. In (51) we used the short hand notation $\mathbb{P}(X \in A)$ for $\mathbb{P}\{\omega \in \Omega \mid X(\omega) \in A\}$, a convention we will follow throughout this paper.

Definition 2.59. If $X_1, \ldots, X_N$ are realvalued random variables the distribution $P_{X_1, \ldots, X_N}$ of the $\mathbb{R}^N$-valued random variable $(X_1, X_2, \ldots, X_N)$ is called the joint distribution of $X_1, \ldots, X_N$, i.e.

$$P_{X_1, \ldots, X_N}(A) = \mathbb{P}\left((X_1, X_2, \ldots, X_N) \in A\right)$$

for any $A \in B(\mathbb{R}^N)$.

The random variables $X_1, \ldots, X_N$ are called independent if $P_{X_1, \ldots, X_N}$ is the product measure of the $P_{X_i}$, i.e. if

$$P_{X_1, X_2, \ldots, X_N}(A_1 \times A_2 \times \ldots \times A_N) = P_{X_1}(A_1) \cdot P_{X_2}(A_2) \cdot \ldots \cdot P_{X_N}(A_N)$$

A sequence $\{X_i\}_{i \in \mathbb{N}}$ of random variables is called independent if the random variables $X_1, X_2, \ldots, X_N$ are independent for all $N \in \mathbb{N}$.

Definition 2.60. If $X_n$ and $X$ are random variables (with values in $\mathbb{R}^d$) we say that $X_n$ converges in distribution to $X$, denoted by $X_n \overset{D}{\rightarrow} X$, if the corresponding distributions converge, i.e. if $P_{X_n} \overset{D}{\rightarrow} P_X$. 

29
Remark 2.61. Sometimes we also say $X_n$ converges in distribution to a probability measure $\mu$ or $X_n \xrightarrow{D} \mu$, if $P_{X_n} \Rightarrow \mu$.

Notation 2.62. If $(\Omega, \mathcal{A}, P)$ is a probability space we will usually call the expectation with respect to $P$ by $E$. So convergence in distribution of $\mathbb{R}^d$-valued random variables $X_n$ to a random variable $X$ means:

$$E\left(f(X_n)\right) \rightarrow E\left(f(X)\right) \quad \text{for all } f \in C_b(\mathbb{R}^d) \quad (54)$$

Definition 2.63. Suppose $X_n$ are random variables on a probability space $(\Omega, \mathcal{F}, P)$. We say $X_n$ converges $P-$stochastically to $X$ (notation: $X_n \xrightarrow{P} X$), if for any $\varepsilon > 0$

$$P(|X_n - X| > \varepsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

Proposition 2.64. Suppose $X_n$ are random variables on a probability space $(\Omega, \mathcal{F}, P)$. If $X_n$ converges in distribution to a constant $c$, then $X_n$ converges $P-$stochastically to $c$.

Proof:

$$P(|X_n - c| > \varepsilon) = 1 - P(|X_n - c| \leq \varepsilon) \leq 1 - E\left(\Phi_{\frac{\varepsilon}{2}}(X_n - c)\right) \rightarrow 0$$

Finally, we add an observation about the expectation value of products of random variables which will be useful later.

Lemma 2.65. If $X_1, X_2, \ldots, X_N$ are random variables with $E(|X_i|^N) < \infty$ for all $i = 1, \ldots, N$ then

$$E(|X_1 \cdot X_2 \cdot \ldots \cdot X_N|) \leq \max_{1 \leq i \leq N} E(|X_i|^N) < \infty \quad (55)$$
Proof: By a repeated application of the Hölder inequality we get

\[
m\mu \mathbb{E}(|X_1 X_2 \cdots X_N|) \\
\leq \mathbb{E}(|X_1|^N)^{1/N} \mathbb{E}(|X_2 \cdots X_N|^{N/(N-1)})(N-1)/N \\
\leq \mathbb{E}(|X_1|^N)^{1/N} \mathbb{E}(|X_2|^N)^{1/N} \mathbb{E}(|X_3 \cdots X_N|^{N/(N-2)})(N-2)/N \\
\cdots \\
\leq \max_{1 \leq i \leq N} \mathbb{E}(|X_i|^N) < \infty
\]

\[\Box\]
3 Independent Random Variables

3.1 Warm-up: A Law of large numbers

In this section we start with a very easy example: A (weak) version of the law of large numbers. This serves merely as a warm-up for the more complicated and more interesting things to come. Assume $X_n$ are independent identically distributed (i.i.d.) random variables on a probability space $(\Omega, F, P)$. As usual in this paper we assume that all moments of $X_n$ exist.

**Theorem 3.1 (Law of Large Numbers).** If $X_n$ are independent, identically distributed random variables with moderately growing moments, then the random variables

$$S_N = \frac{1}{N} \sum_{i=1}^{N} X_i$$

converge $\mathbb{P}$-stochastically to $m := \mathbb{E}(X_1)$.

**Proof:** From Theorem 2.56 and Proposition 2.64 we learn that it is enough to prove that the moments of $S_n$ converge in distribution to (the Dirac measure in) $m$. So, let us compute the moments of the random variable $S_N$:

$$\mathbb{E}(S_N^L) = \frac{1}{N^L} \sum_{i_1, i_2, \ldots, i_L} \mathbb{E}(X_{i_1}X_{i_2} \cdots X_{i_L})$$  \hspace{1cm} (56)

Observe that the sum has $N^L$ summands, since the indices can run from 1 to $N$ independent of each other.

Moreover, by Lemma 2.65 each summand is bounded by $\mathbb{E}(X_1^L)$. We split the sum in a term for which all the indices are distinct and the rest:

$$\mathbb{E}(S_N^L) = \frac{1}{N^L} \sum_{\substack{i_1, i_2, \ldots, i_L \\
all \text{indices distinct}}} \mathbb{E}(X_{i_1}X_{i_2} \cdots X_{i_L}) + \frac{1}{N^L} \sum_{\substack{i_1, i_2, \ldots, i_L \\
not \text{all distinct}}} \mathbb{E}(X_{i_1}X_{i_2} \cdots X_{i_L})$$  \hspace{1cm} (57) \hspace{1cm} (58)

In the sum (58) at least two of the indices coincide, thus the number of summands is bounded by $N^{L-1}$. Hence, the whole term (58) is bounded by $\frac{1}{N} \mathbb{E}(X_1^L)$. Consequently, it converges to zero as $N \to \infty$. 

33
We turn to the expression (57). Due to independence, we have:

\[ E(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_L}) = \prod_{\ell=1}^{L} E(X_{i_\ell}) = m^L \quad \text{for distinct } i_1, i_2, \ldots, i_L \]  \hspace{1cm} (59)

Since the indices are all distinct, there are \( \frac{N!}{(N-L)!} \) summands, so

\[ (57) = \frac{1}{N^L} \cdot \frac{N!}{(N-L)!} m^L \xrightarrow{N \to \infty} m^L \]  \hspace{1cm} (60)

From Example 2.41 we know that \( m^L \) are the moments of the Dirac measure \( \delta_m \), so \( S_N \xrightarrow{D} \delta_m \) in distribution, thus by Proposition 2.64 the theorem follows.

\[ \square \]

Remark 3.2. The assumption that the random variables are independent and identically distributed was used in the proof above only to compute the sum (57). To estimate the sum (58) only a rough estimate of the moments as in Lemma 2.65 was needed.

Remark 3.3. The reader experienced in probability may have noticed that the above proof is much more complicated than necessary. With a simple use of the Chebyshev inequality an estimate of the second moment would have been sufficient. In fact, we will pursue this idea in Section 3.4.

The more complicated argument above was presented to introduce the general idea which will be used in more sophisticated form again and again in this text.

3.2 Some combinatorics

In section 3.1 and, in fact, throughout this paper we got to estimate sums over expressions of the form

\[ E(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_L}) \]  \hspace{1cm} (61)

If the random variables \( X_i \) are independent and identically distributed, then the value of expression (61) depends only on the number of times each index \( i \) occurs among the indices in (61).

To make this statement precise we introduce:

Definition 3.4. We define the multiindex \( \bar{i} = (i_1, i_2, \ldots, i_L) \in \{1, 2, \ldots, N\}^L \).
1. For \( j \in \{1, 2, \ldots, N\} \) we set
\[
\nu_j(i) = |\{ k \in \{1, 2, \ldots, L\} \mid i_k = j \}|. \tag{62}
\]
Here \(|M|\) denotes the number of elements in the set \( M \).

2. For \( \ell = 0, 1, \ldots, L \) we define
\[
\rho_\ell(i) = |\{ j \mid \nu_j(i) = \ell \}| \tag{63}
\]
and
\[
\rho(i) = (\rho_1(i), \rho_2(i), \ldots, \rho_L(i)) \tag{64}
\]
In words: \( \nu_j(i) \) tells us how frequently the index \( j \) occurs among the indices \((i_1, i_2, \ldots, i_L)\) and \( \rho_\ell(i) \) gives the number of indices that occur exactly \( k \) times in \((i_1, i_2, \ldots, i_L)\).

**Remark 3.5.** Setting \( r = (r_1, r_2, \ldots, r_L) = \rho(i) \) we have
\[
\mathbb{E}(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_L}) = \mathbb{E}\left(\prod_{j=1}^{N} X_j^{\nu_j(i)}\right) \\
= \mathbb{E}\left(\prod_{j_1=1}^{r_1} X_{j_1} \cdot \prod_{j_2=1}^{r_2} X_{j_1+j_2}^2 \cdot \prod_{j_{r_1}+j_{r_2}+\ldots+j_{r_L}=1}^{r_L} X_{\sum_{\ell=1}^{L-1} r_\ell+j_L}^L\right) \\
= \prod_{\ell=1}^{L} \mathbb{E}(X_\ell^{r_\ell})^r
\]
Thus, we showed

**Proposition 3.6.** If \( X_n, n \in \mathbb{N} \) are independent and identically distributed random variables, then
\[
\mathbb{E}(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_L}) = \mathbb{E}(X_{j_1} \cdot X_{j_2} \cdot \ldots \cdot X_{j_L}) \tag{65}
\]
whenever \( \rho(i_1, i_2, \ldots, i_L) = \rho(j_1, j_2, \ldots, j_L) \).

Therefore the following abbreviation is well defined.

**Notation 3.7.** If \( r = (r_1, r_2, \ldots, r_L) = \rho(i) \) we set
\[
\mathbb{E}(X(r)) = \mathbb{E}(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_L}) \tag{66}
\]
To evaluate sums of expressions as in (61) we investigate the combinatorics of \(L\)-tuples of indices.

**Lemma 3.8.** For each \(i \in \{1, 2, \ldots, N\}^L\) we have:

1. \(0 \leq \nu_j(i) \leq L\) and \(\sum_{j=1}^N \nu_j(i) = L\)

2. \(0 \leq \rho_\ell(i) \leq L\), for \(\ell \neq 0\) and \(\rho_\ell(i) = 0\) for \(\ell > L\),

3. \(\sum_{\ell=0}^L \rho_\ell(i) = N\)

4. \(\sum_{\ell=1}^L \rho_\ell(i) = |\{i_1, i_2, \ldots, i_L\}|\)

   i.e. the sum over the \(\rho_\ell\) is the number of distinct indices in \(i\).

5. \(\sum_{\ell=1}^L \ell \rho_\ell(i) = L\)

The proof of the Lemma is left to the reader.

**Definition 3.9.** Let us set \(\mathcal{N} = \{1, 2, \ldots, N\}\) and \(\mathcal{L} = \{0, 1, 2, \ldots, L\}\). We call an \(L\)-tuple \(r = (r_1, r_2, \ldots, r_L) \in \{0, 1, \ldots, L\}^L\) a profile if

\[
\sum_{\ell=1}^L \ell r_\ell = L.
\]

Given a profile \(r\) we define

\[
W(r) = W_L(r) = \{i \in \mathcal{N}^k \mid \rho(i) = r\}
\]

and

\[
w(r) = w_L(r) = |\{i \in \mathcal{N}^k \mid \rho(i) = r\}|.
\]

From Proposition (3.6) we learn that the expectation value \(E(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_L})\) for i.i.d. random variables depends only on the profile of \((i_1, i_2, \ldots, i_L)\). More generally this is true for exchangeable random variables. We’ll have more to say about this in Chapter 4.
Definition 3.10. We denote the set of all profiles in $L^L$ by $\Pi^{(L)}$ or simply by $\Pi$, if $L$ is clear from the context, thus
\[
\Pi := \Pi^{(L)} := \left\{ (r_1, r_2, \ldots, r_L) \in \{0, 1, \ldots, L\}^L \mid \sum_{\ell=1}^L \ell r_\ell = L \right\} \quad (70)
\]
For later use, we define some subsets of $\Pi$:
\[
\Pi_k := \left\{ r \in \Pi^{(L)} \mid r_1 = k \right\} \quad (71)
\]
\[
\Pi_+ := \left\{ r \in \Pi^{(L)} \mid r_1 > 0 \right\} \quad (72)
\]
\[
\Pi^0 := \left\{ r \in \Pi^{(L)} \mid r_\ell = 0 \text{ for all } \ell \geq 3 \right\} \quad (73)
\]
\[
\Pi^+ := \left\{ r \in \Pi^{(L)} \mid r_\ell > 0 \text{ for some } \ell \geq 3 \right\} \quad (74)
\]
Remark 3.11. We may combine the sub- and superscripts in the previous definition, for example
\[
\Pi_k^+ = \left\{ r \in \Pi^{(L)} \mid r_1 = k \text{ and } r_\ell > 0 \text{ for some } \ell \geq 3 \right\} \quad (75)
\]
In particular the set $\Pi^0_0$ consists of those $r$ with $r_\ell = 0$ for all $\ell \neq 2$. Note that $\Pi^0_0 = \emptyset$ if $L$ is odd.

The above defined sets decompose $\Pi$ into disjoint subsets in various ways, for example
\[
\Pi = \bigcup_{k=0}^L \Pi_k
\]
\[
= \Pi_+ \cup \Pi^0_0 \cup \Pi^+_0
\]
and so on.

We emphasize that the number $|\Pi|$ does not depend on $N$, in fact, we have the following very rough estimate:

Lemma 3.12. $|\Pi^{(L)}| \leq (L + 1)^L$

Proof: Since $0 \leq r_\ell$ and $\sum_{\ell=1}^L \ell r_\ell = L$ we have $0 \leq r_\ell \leq L$. Hence there are at most $L + 1$ possible values for each of the $L$ components of the $L$-tuple $(r_1, r_2, \ldots, r_L)$. □

We note the following important observation
Theorem 3.13. For independent identically distributed random variables $X_i$ we have

$$\sum_{i_1,i_2,\ldots,i_L=1}^N \mathbb{E}(X_{i_1}X_{i_2}\cdots X_{i_L}) = \sum_{\mathbf{r} \in \Pi} w_L(\mathbf{r}) \mathbb{E}(X(\mathbf{r})) = \sum_{\mathbf{r} \in \Pi} w_L(\mathbf{r}) \prod_{\ell=1}^L \mathbb{E}(X^\ell_i)^{r_\ell}$$  \hspace{1cm} (76)

It is clear now that a good knowledge about $w_L(\mathbf{r})$ will be rather useful. In fact, we have

Theorem 3.14. For $\mathbf{r} \in \Pi^{(L)}$ set $r_0 = N - \sum_{\ell=1}^L r_\ell$, then

$$w_L(\mathbf{r}) = \frac{N!}{r_1! r_2! \cdots r_L! r_0!} \prod_{\ell=1}^L \frac{L!}{1! r_\ell^1 \cdots 2! r_\ell^2 \cdots 3! r_\ell^3 \cdots L! r_\ell^L}$$  \hspace{1cm} (77)

Much of the rest of this section is devoted to the proof of Theorem 3.14. We start with a little digression about multinomial coefficients.

About multinomial coefficients

Definition 3.15. For $m_1, m_2, \ldots, m_K \in \mathbb{N}$ with $\sum_{k=1}^K m_k = M$ we define the multinomial coefficient

$$\binom{M}{m_1 m_2 \ldots m_K} = \frac{M!}{m_1! m_2! \ldots m_K!}$$.  \hspace{1cm} (78)

The multinomial coefficients are generalizations of the binomial coefficients $\binom{M}{m}$, more precisely: $\binom{M}{m} = \binom{M}{m M - m}$.

The binomial coefficient $\binom{M}{m}$ gives the number of ways to choose $m$ objects from $M$ possible elements, in other words: to partition a set of $M$ elements into a set of $m$ objects (‘the chosen ones’) and a set of $M - m$ objects (‘the rest’).

The multinomial coefficient counts the partitions into (in general) more than two sets.

Proposition 3.16. Let $m_1, m_2, \ldots, m_K$ be positive integers with $\sum_{k=1}^K m_k = M$. The multinomial coefficient

$$\binom{M}{m_1 m_2 \ldots m_K}$$  \hspace{1cm} (79)
counts the number of ways to distribute the $M$ elements of a set $C_0$ into $K$ classes $C_1, C_2, \ldots, C_K$ in such a way that the class $C_1$ contains exactly $m_1$ elements, $C_2$ $m_2$ elements and so on.

**Proof:** There are $\binom{M}{m_1}$ ways to choose the elements of class $C_1$.

Then there are $\binom{M-m_1}{m_2}$ ways to choose the elements of $C_2$ from the remaining $M - m_1$ objects of $C_0$.

Then there are $\binom{M-m_1-m_2}{m_3}$ possibilities to choose the elements of $C_3$ from the remaining $M - m_1 - m_2$ objects of $C_0$.

And so on . . .

This gives a total of

$$\binom{M}{m_1} \cdot \binom{M-m_1}{m_2} \cdot \binom{M-m_1-m_2}{m_3} \cdot \cdots \binom{M-\sum_{k=1}^{K-1} m_k}{m_K}$$  \hspace{1cm} (80)

possible ways to partition $M$ objects in the way prescribed. This expression equals

$$\binom{M}{m_1 \, m_2 \, \ldots \, m_K}$$  \hspace{1cm} (81)

The multinomial coefficient can also be used to compute the number of $M$-tuples (‘words of length $M$’) which can be built from a finite set $A$ (‘alphabet’) with prescribed number of occurrences of the various elements (‘characters’) of $A$, more precisely:

**Proposition 3.17.** Suppose $A = \{a_1, a_2, \ldots, a_K\}$ is a set with $K$ elements, let $M \geq K$ and $m_1, m_2, \ldots, m_K$ be (strictly) positive integers with $\sum_{k=1}^{K} m_k = M$, then the number of distinct $M$-tuples for each $k$ containing $a_k$ exactly $m_k$ times is given by the multinomial coefficient

$$\binom{M}{m_1 \, m_2 \, \ldots \, m_K}$$  \hspace{1cm} (82)

**Proof:** The idea of the proof is to use the numbers from $M = \{1, 2, \ldots, M\}$ to mark the positions in the $M$-tuple $(w_1, w_2, \ldots, w_M)$.

To implement this idea, we partition the set $M$ into $K$ classes $A_1, \ldots, A_K$ such that $A_k$ contains exactly $m_k$ elements.
Then we construct the $M$-tuple $(w_1, w_2, \ldots, w_M)$ by setting
$$w_\ell = a_k \quad \text{if and only if} \quad a_k \in A_\ell$$ (83)
This gives an $M$-tuple with the prescribed distribution of elements of $A$ and each $M$-tuple with this property is obtained in this way.

The number of such partitions of $M$ is given by
$$\binom{M}{m_1 m_2 \ldots m_K}$$ (84)
as we saw in Proposition 3.16.

This ends our detour about multinomial indices. We turn to the proof of Theorem 3.14.

**Proof** (Theorem 3.14): Let us construct an $L$-tuple $(i_1, i_2, \ldots, i_L)$ of elements of $N$ with profile $(r_1, r_2, \ldots, r_L)$. To do so we first partition the set $N$ into classes $R_1, R_2, \ldots, R_L$ and $R_0$. The class $R_\ell$ will be interpreted as those numbers in $N$ which are going to occur exactly $\ell$ times in the tuple, in particular the elements in $R_0$ will not occur at all in $(i_1, i_2, \ldots, i_L)$. We want exactly $r_\ell$ elements in the class $R_\ell$, including $r_0 = N - \sum_{\ell=1}^{L} r_\ell$ elements in $R_0$.

According to Proposition 3.16, the number of such partitions is
$$\binom{N}{r_1 r_2 \ldots r_L r_0} = \frac{N!}{r_1! r_2! \ldots r_L! r_0!}$$ (85)
which is the first fraction in (77).

Next, we have to construct an $L$-tuple $i = (i_1, i_2, \ldots, i_L)$ which contains each of the $r_1$ elements in $R_1$ exactly once, each of the $r_2$ elements of $R_2$ twice and so on. By Proposition 3.17 the number of ways this can be done is given by:
$$\binom{L}{r_1 \times 1 r_2 \times 2 r_3 \times 3 \ldots r_L \times L} = \frac{L!}{1! r_1 2! r_2 3! r_3 \ldots L! r_L}$$ (86)
Observe that the right hand side of (86) is indeed a multinomial coefficient since
$$\sum_{\ell=1}^{L} \ell r_\ell = L$$
This proves the theorem.

The following Corollary to Theorem 3.14 will be useful in the following sections.

**Corollary 3.18.** For $r \in \Pi^{(L)}$ we have

$$w_L(r) \approx C(r) N \sum_{\ell=1}^{L} r_{\ell} \quad \text{as} \ N \to \infty$$

(87)

where the constant $C(r)$ is given by

$$C(r) = \frac{1}{r_1! r_2! \ldots r_L!} \frac{L!}{\prod_{\ell=1}^{L} \ell!^{r_{\ell}}} \leq L!$$

(88)

**Notation 3.19.** The expression

$$a_N \approx b_N \quad \text{as} \ N \to \infty$$

is an abbreviation for

$$\lim_{N \to \infty} \frac{a_N}{b_N} = 1$$

We used this Corollary tacitly and in a weak form already in the proof of Theorem 3.1, when we concluded that the contribution of those $\hat{j}$ with $\sum_{\ell=2}^{L} \rho_{\ell}(\hat{j}) > 0$ is negligible compared to the normalization $\frac{1}{N}$.

### 3.3 The central limit theorem

We are now prepared to prove the following (version of the) central limit theorem

**Theorem 3.20** (Central Limit Theorem). Let $X_n$ be a sequence of independent identically distributed random variables with $\mathbb{E}(|X_n|^k) < \infty$ for all $k$ and denote by $m = \mathbb{E}(X_1) = \mathbb{E}(X_n)$, $\sigma^2 = \mathbb{V}(X_1) = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2$ their common mean and variance.

Then the random variable

$$\Sigma_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( X_i - \mathbb{E}(X_i) \right)$$

(89)

converges in distribution to a normal distribution $\mathcal{N}(0, \sigma^2)$ with mean zero and variance $\sigma^2$. 41
Proof: By theorem 2.56 it is sufficient to prove that the moments of $\Sigma_N$ converge to those of $\mathcal{N}(0, \sigma^2)$.

Without loss of generality we may assume that $m_1 = E(X_i) = 0$.

We compute:

$$
E\left((\Sigma_N)^L\right) = \frac{1}{N^{L/2}} \sum_{i_1, i_2, \ldots, i_L} E(X_{i_1} X_{i_2} \cdots X_{i_L})
$$

$$
= \frac{1}{N^{L/2}} \sum_{\mathbf{r} \in \Pi} w_L(\mathbf{r}) \ E\left(X(\mathbf{r})\right)
$$

$$
= \frac{1}{N^{L/2}} \sum_{\mathbf{r} \in \Pi_+} w_L(\mathbf{r}) \ E\left(X(\mathbf{r})\right)
$$

$$
+ \frac{1}{N^{L/2}} \sum_{\mathbf{r} \in \Pi_0^\perp} w_L(\mathbf{r}) \ E\left(X(\mathbf{r})\right)
$$

$$
+ \frac{1}{N^{L/2}} \sum_{\mathbf{r} \in \Pi_0^\perp} w_L(\mathbf{r}) \ E\left(X(\mathbf{r})\right)
$$

(90)

(91)

(92)

where the sets $\Pi_+$, $\Pi_0^\perp$ and $\Pi_0^\perp$ are defined in Definition 3.10.

We handle the summands in (90)-(92) separately.

1. We start with (90). So, take $\mathbf{i} \in \Pi_+$. Then $r_1(\mathbf{i}) > 0$ and there is an index, say $i_j$, which occurs only once. Since the $X_i$ are independent and $E(X_i) = 0$ we have

$$
E(X_{i_1} X_{i_2} \cdots X_{i_L}) = E(X_{i_j}) E\left(\prod_{\ell \neq j} X_{i_\ell}\right) = 0
$$

Therefore, the summand (90) vanishes.

2. We turn to (92), the third expression in the above sum. For $\mathbf{r} \in \Pi_0^\perp$ we have $r_1 = 0$ and $r_\ell \geq 1$ for some $\ell \geq 3$.

Consequently, since

$$
\sum_{\ell=1}^L \ell r_\ell = L
$$

42
we get

\[\sum_{\ell=1}^{L} r_\ell = \frac{1}{2} \left( 2r_2 + 2 \sum_{\ell=3}^{L} r_\ell \right) \leq \frac{1}{2} \left( 2r_2 + \frac{2}{3} \sum_{\ell=3}^{L} \ell r_\ell \right) \leq \frac{1}{2} \left( 2r_2 + \sum_{\ell=3}^{L} \ell r_\ell - \frac{1}{3} \sum_{\ell=3}^{L} \ell r_\ell \right) \leq \frac{1}{2} \left( 2r_2 + \sum_{\ell=3}^{L} \ell r_\ell - 1 \right) = \frac{1}{2} \sum_{\ell=1}^{L} \ell r_\ell - \frac{1}{2} = L/2 - 1/2 \] (93)

Using Corollary 3.18, Lemma 3.12 and Lemma 2.65 we estimate:

\[
\left| \frac{1}{NL^{L/2}} \sum_{r \in \Pi_0^0} w_L(r) \mathbb{E}(X(r)) \right| \leq \frac{N^{L/2-1/2}}{NL^{L/2}} L! (L + 1)^L \mathbb{E}(|X_1|^L)
\]

For \(N \to \infty\) the last expression goes to zero. (Remember: \(L\) is fixed!)

It follows that both (90) and (92) don’t contribute to the limit of the moments.

3. Finally we consider the second term, i.e. (91).

The class \(\Pi_0^0\) consist of those \(r\) for which \(r_\ell = 0\) for \(\ell \neq 2\). Since \(\sum \ell r_\ell = L \Pi_0^0 = \emptyset\) for odd \(L\) and consist of exactly one element, namely \((0, L/2, 0, \ldots, 0)\) for even \(L\).

Thus, for odd \(L\)

\[
\frac{1}{NL^{L/2}} \sum_{r \in \Pi_0^0} w_L(r) \mathbb{E}(X(r)) = 0
\] (94)
For even \( L = 2K \) we have

\[
\frac{1}{N^{L/2}} \sum_{r \in \Pi_0} w_L(r) \mathbb{E}(X(r)) = \frac{1}{N^K} \left( \frac{N!}{(K)! (N-K)!} \right) \frac{(2K)!}{2^K} (\sigma^2)^K \rightarrow (2K-1)!! (\sigma^2)^K \quad \text{as } N \rightarrow \infty
\]

(95)

Summarizing, we have shown that \( \mathbb{E}(\Sigma_N^L) \) converges to \((L-1)!! (\sigma^2)^{L/2}\) if \( L \) is even and to 0 if \( L \) is odd. Thus \( \mathbb{E}(\Sigma_N^L) \) converges to the \( L^{th} \) moment of the normal distribution \( \mathcal{N}(0,\sigma^2) \) (see 2.42).

This finishes the prove of Theorem 3.20. \( \square \)

**Remark 3.21.** The argument in estimate[2] of summand (92) didn’t use the independence of the random variables, in fact, all that was needed of the \( X_i \) was that \( \sup_i \mathbb{E}\left( \|X_i\|^L \right) < \infty \). We will use this observation in later chapters.

### 3.4 More on the law of large numbers

In this section we use the insight from the previous section for a closer look at the law of large numbers. In particular, we will say something about the rate of convergence in the weak form of this theorem. As a "byproduct" this implies the strong law of large numbers.

The following inequality (or rather collection of inequalities) is a central tool in our proof, and, in fact, in probability theory.

**Theorem 3.22.** *(Chebyshev-Markov inequality):*

Suppose \( X \) is a random variable on the probability space \((\Omega, \mathcal{F}, \mathbb{P})\). If \( f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0} \) is a non decreasing function then

\[
\mathbb{P}(|X| > a) \leq \frac{\mathbb{E}(f(X))}{f(a)}
\]

(96)
Proof:
\[
P(|X| > a) = \int \chi_{|X|>a}(w) \, d\mathbb{P}(w)
\]  
\[
\leq \int \chi_{|X|>a}(w) \frac{f(|X|)(w)}{f(a)} \, d\mathbb{P}(w)
\]  
\[
\leq \int \frac{f(|X|)}{f(a)} \, d\mathbb{P} = \frac{\mathbb{E}(f(X))}{f(a)}
\]

We used that \(\frac{f(X)}{f(a)} \leq 1\) whenever \(X \geq a\).

\[\square\]

**Corollary 3.23.** For all \(p > 0\) :
\[
P(|X| > a) \leq \frac{1}{a^p} \mathbb{E}(|X|^p)
\]

Chebyshev’s inequality
\[
P(|X - E(X)| > a) \leq \frac{1}{a^2} E(|X - E(X)|^2)
\]  
\[
= \frac{\mathbb{V}(X)}{a^2}
\]

follows immediately from Corollary 3.23 with \(p = 2\).

The importance of the Chebyshev-Markov inequality comes from the fact that it is usually easier to deal with expectation (e.g. moments) than to compute probabilities. The previous sections of this chapter are a - we hope convincing - example for this observation.

Now we use the Chebyshev-Markov inequality together with the techniques from the previous sections to estimate the probability \(\mathbb{P}(\bigg|\frac{1}{N} \sum_{i=1}^{N} (X_i - E(X_i)) \bigg| > a)\).

**Theorem 3.24.** Suppose \(X_i\) are independent random variables on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\). Suppose furthermore that \(\sup_{i \in \mathbb{N}} \mathbb{E}(X_i^K) \leq M_K < \infty\) for each \(K \in \mathbb{N}\).

Then for any \(L \in \mathbb{N}\) there is a constant \(C_L\) such that for any \(a > 0\)
\[
\mathbb{P}\left(\bigg|\frac{1}{N} \sum_{i=1}^{N} (X_i - E(X_i)) \bigg| > a\right) \leq \frac{C_L}{a^{2L}N^L}
\]

**Remark 3.25.** The constant \(C_L\) can be chosen to be
\[
C_L = \frac{(2L+1)^{2L}(2L)!}{2^{2L}} \sup_{i \in \mathbb{N}} \mathbb{E}((X_i - E(X_i))^{2L})
\]
Remark 3.26. Theorem 3.24 is a kind of a quantitative version of the weak law of large numbers. It asserts that $\mathbb{P}(\left\lvert \frac{1}{N} \sum_{i=1}^{N} (X_i - E(X_i)) \right\rvert > a)$ not only converges to zero, but it converges faster than any power of $N$.

Proof: Using the Chebyshev-Markov inequality we obtain
\begin{align*}
\mathbb{P}\left( \left\lvert \frac{1}{N} \sum_{i=1}^{N} (X_i - E(X_i)) \right\rvert > a \right) &= \mathbb{P}\left( \left\lvert \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i - E(X_i)) \right\rvert > a \sqrt{N} \right) \tag{105} \\
&\leq \frac{1}{a^{2L}N^L} \mathbb{E}\left( \left\lvert \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i - E(X_i)) \right\rvert^{2L} \right) \tag{106}
\end{align*}

In the proof of Theorem 3.20 we have shown that the above expectation converge (as $N \to \infty$), in particular, they are bounded. This is the assertion of the theorem.

To give an explicit estimate of the constant we redo the estimate of the expectation. We set $Y_i = X_i - E(X_i)$.

\begin{align*}
\mathbb{E}\left( \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Y_i \right)^{2L} \right) &\leq \frac{1}{N^L} \sum_{r \in \Phi(2L)} w(r) \left\lvert \mathbb{E}(Y(r)) \right\rvert \tag{108} \\
&= \frac{1}{N^L} \sum_{r \in \Phi_0} w(r) \left\lvert \mathbb{E}(Y(r)) \right\rvert \quad \text{(since $\mathbb{E}(Y_i) = 0$)} \tag{109} \\
&\leq \sum_{r \in \Phi_0} \frac{(2L)!}{2^{2L}} \sup_{i \in \mathbb{N}} \mathbb{E}(Y_i^{2L}) \quad \text{(due to Corollary 3.18 and Lemma 2.65)} \tag{110} \\
&\leq \frac{(2L + 1)2^{2L}(2L)!}{2^{2L}} \sup_{i \in \mathbb{N}} \mathbb{E}(Y_i^{2L}) \quad \text{(by Lemma 3.12)} \tag{111}
\end{align*}

We single out the following result:

Corollary 3.27. Under the assumptions of Theorem 3.24 we have
\begin{align*}
\mathbb{E}\left( \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i - E(X_i)) \right)^{2L} \right) &\leq \frac{(2L + 1)2^{2L}(2L)!}{2^{2L}} \sup_{i \in \mathbb{N}} \mathbb{E}\left( (X_i - E(X_i))^{2L} \right) \tag{112}
\end{align*}
Theorem 3.28 (Strong law of large numbers).
Suppose $X_i$ are independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose furthermore that $\sup_{i \in \mathbb{N}} \mathbb{E}(X_i^K) \leq M_K < \infty$ for all $K$.

Then

$$S_N := \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} (X_i - \mathbb{E}(X_i)) = 0 \quad \mathbb{P}\text{-almost surely},$$

i.e. $\mathbb{P}\left( \lim_{N \to \infty} S_N = 0 \right) = \mathbb{P}\left( \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} (X_i - \mathbb{E}(X_i)) = 0 \right) = 1$. (114)

Remark 3.29. If all $\mathbb{E}(X_i)$ are equal then $\frac{1}{N} \sum_{i=1}^{N} X_i \to \mathbb{E}(X_1)$ $\mathbb{P}$-almost surely.

We start the proof of the theorem with a discussion of the limes superior and limes inferior of a sequence of sets and the celebrated Borel-Cantelli Lemma.

Definition 3.30. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A_n \in \mathcal{F}$ a sequence of sets. We define

1. $\limsup_{n \to \infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$ is called the *lim-sup of the $A_n$*.
2. $\liminf_{n \to \infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m$ is called the *lim-inf of the $A_n$*.

Lemma 3.31. 1. $\limsup_{n \to \infty} A_n = \{ \omega \mid \omega \in A_n \text{ for infinitely many } n \}.$
2. $\liminf_{n \to \infty} A_n = \{ \omega \mid \omega \in A_n \text{ for all } n \text{ large enough} \}.$

Proof: Define $B_n = \bigcup_{m=n}^{\infty} A_m$. So, $\omega \in B_n$ iff (i.e. if and only if) there is an $m \geq n$ with $\omega \in A_m$. By definition

$$\limsup_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} B_n$$

(115)

Consequently, $\omega \in \limsup_{n \to \infty} A_n$ iff $\omega \in B_n$ for all $n$ which is the case iff $\omega \in A_n$ infinitely often. \qed

The following proposition is ‘one half’ of the famous Borel-Cantelli Lemma.
Proposition 3.32. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $A_n \in \mathcal{F}$.

1. If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(\limsup_{n \to \infty} A_n) = 0$.

2. If $\sum_{n=1}^{\infty} \mathbb{P}(\overline{A}_n) < \infty$, then $\mathbb{P}(\liminf_{n \to \infty} A_n) = 1$.

Proof: 1. Set (as in the proof above) $B_n = \bigcup_{m=n}^{\infty} A_m$.

Then $B_n \supset B_{n+1}$ and $\liminf_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} B_n$. Thus

$$\mathbb{P}(\liminf_{n \to \infty} A_n) = \lim_{n \to \infty} \mathbb{P}(B_n) \leq \lim_{n \to \infty} \sum_{m=n}^{\infty} \mathbb{P}(A_m)$$

Since $\sum_{m=1}^{\infty} \mathbb{P}(A_m) < \infty$ we know $\sum_{m=n}^{\infty} \mathbb{P}(A_m) \to 0$ as $n \to \infty$.

2. $\liminf_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \overline{A}_m$.

Thus 2. follows from 1. \qed

Proposition 3.33. Let $Y_n$ be a sequence of random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, set $A_{m,k} = \{\omega \mid |Y_m| \geq \frac{1}{k}\}$ and $A = \{\omega \mid \lim_{n \to \infty} Y_n = 0\}$. If $\sum_{m=1}^{\infty} \mathbb{P}(A_{m,k}) < \infty$ for all $k \in \mathbb{N}$ then $\mathbb{P}(A) = 1$.

Proof:

$$A = \{\omega \mid \forall k \in \mathbb{N} \exists n \in \mathbb{N} \forall m \geq n \ |Y_m| < \frac{1}{k}\}$$

$$= \bigcap_{k \in \mathbb{N}} \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} A_{m,k}$$

Defining $M_k = \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} A_{m,k}$ we have $M_{k+1} \subset M_k$ and consequently

$$\mathbb{P}(A) = \lim_{k \to \infty} \mathbb{P}(M_k)$$
\[ \mathbb{P}(M_k) = \mathbb{P}\left( \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} A_{m,k} \right) \quad (119) \]
\[ = \mathbb{P}\left( \lim \inf_{n \to \infty} A_{n,k} \right) \quad (120) \]
\[ = 1 \quad (121) \]

since \( \sum_{n=1}^{\infty} \mathbb{P}(\mathcal{C}_{A_{n,k}}) = \sum_{n=1}^{\infty} \mathbb{P}\left( |Y_n| > \frac{1}{k} \right) < \infty \) by assumption, hence \( \mathbb{P}(A) = 1 \).

\[ \square \]

**Proof (Theorem 3.28):** Defining

\[ A_{N,k} := \{|S_N| \geq \frac{1}{k}\} \]

and applying Proposition 3.33 we have to prove that

\[ \sum_{N=1}^{\infty} \mathbb{P}\left( A_{N,k} \right) < \infty. \]

By Theorem 3.24 we have

\[ \mathbb{P}\left( |S_N| > \frac{1}{k} \right) \leq C_4 k^4 \frac{1}{N^2}\]

which proves the assertion. \( \square \)
4 Exchangeable Random Variables

4.1 Basics

In this section we discuss the notion of exchangeability for sequences of random variables. Independent and identically distributed random variables constitute a particular case of exchangeable random variables, but the latter notion is considerably more general.

Definition 4.1. A sequence \( X_1, \ldots, X_N \) of random variables on a probability space \( (\Omega, \mathcal{A}, \mathbb{P}) \) is called exchangeable, if for any set \( A \in \mathcal{A} \) and any permutation \( \pi \) of \( \{1, \ldots, N\} \) we have

\[
\mathbb{P}\left( (X_1, \ldots, X_N) \in A \right) = \mathbb{P}\left( (X_{\pi(1)}, \ldots, X_{\pi(N)}) \in A \right)
\]  

An infinite sequence \( X_1, X_2, \ldots \) of random variables is called exchangeable, if the finite sequences \( X_1, \ldots, X_N \) are exchangeable for any \( N \).

Examples 4.2.

1. A (finite or infinite) sequence of independent, identically distributed random variables is exchangeable.

2. Suppose \( X_1, X_2, \ldots \) are independent, identically distributed random variables and the random variable \( X \) is independent of the \( X_i \), then the sequence \( Y_i = X + X_i \) is exchangeable. Note that the \( Y_i \) are not independent in general. The sequence \( Y_i = X \) is a special case (with \( X_i = 0 \)).

3. Prominent examples for sequences of exchangeable variables are given by various versions of Polya’s urn scheme. We start with an urn containing \( P \) balls with a label \( 1 \) and \( M \) balls labeled \(-1\). We draw balls (as long as the urn is not empty) and note the numbers on the balls. We denote these (random) numbers by \( X_1, X_2, \ldots \).

   (a) In the first version we return each drawn ball into the urn before we draw the next ball. This is known as ‘sampling with replacement’ and leads to independent identically distributed random variables.

   (b) If we don’t replace drawn balls we obtain ‘sampling without replacement’. This process comes to a halt after \( M + P \) steps since the urn is empty then.
(c) In the general scheme after drawing a ball (without replacement) we put \( \ell \) (new) balls with the same number as the drawn one into the urn. Thus \( \ell = 1 \) is equivalent to sampling with replacement, \( \ell = 0 \) to sampling without replacement.

For \( \ell = 1, 2, \ldots \) the random variables \( X_1, X_2, \ldots \) given by Poly’s urn scheme are an infinite sequence of exchangeable random variables.

For \( \ell = 0 \) (sampling without replacement) the random variables \( X_1, \ldots X_N \) are exchangeable for \( N \leq M + P \).

**Proposition 4.3.** If \( X_1, X_2, \ldots, X_N, \ N \geq 2 \) is a sequence of exchangeable random variables then the covariance

\[
\text{Cov} \left( X_i, X_j \right) := \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) \tag{123}
\]

of \( X_i \) and \( X_j \) is bounded from below by

\[
\text{Cov} \left( X_i, X_j \right) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) \geq - \frac{1}{N - 1} \mathbb{V}(X_i)
\]

for all \( i, j \in \mathcal{N} = \{1, 2, \ldots, N\} \).

In particular, if \( \{X_i\}_{i \in \mathbb{N}} \) is an infinite sequence of exchangeable random variables then

\[
\mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) \geq 0
\]

**Proof:** We set

\[
\tilde{X}_i = X_i - \mathbb{E}(X_i) = X_i - \mathbb{E}(X_1).
\]

then \( \tilde{X}_i \) are exchangeable and \( \mathbb{E}(\tilde{X}_i) = 0 \).

We will prove that \( \mathbb{E}(\tilde{X}_1 \tilde{X}_2) \geq - \frac{1}{N - 1}\mathbb{E}(X_1^2) \). From this the assertion of the proposition follows immediately.

Since the \( \tilde{X}_i \) are exchangeable we have

\[
\mathbb{E}(\tilde{X}_1 \tilde{X}_2) = \frac{1}{N(N - 1)} \sum_{i \neq j}^{N} \mathbb{E}(\tilde{X}_i \tilde{X}_j)
\]

\[
= \frac{1}{N(N - 1)} \sum_{i,j=1}^{N} \mathbb{E}(\tilde{X}_i \tilde{X}_j) - \frac{1}{N(N - 1)} \sum_{i=1}^{N} \mathbb{E}(\tilde{X}_i^2)
\]

\[
= \frac{1}{N(N - 1)} \mathbb{E}(\sum_{i=1}^{N} \tilde{X}_i)^2 - \frac{1}{N - 1} \mathbb{E}(\tilde{X}_1^2)
\]

\[
\geq - \frac{1}{N - 1} \mathbb{E}(\tilde{X}_1^2)
\]

52
Example 4.4. Consider Polya’s urn scheme with $\ell = 0$ (sampling without replacement) and with $P$ balls labeled ’+1’ and $M$ balls with ’-1’. We assume $M, P \geq 1$ and set $N = M + P$.

Then

$$\mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) = -\frac{1}{N-1} \mathbb{V}(X_1).$$

(124)

It follows that Polya’s urn scheme cannot be extended to an exchangeable sequence $X_1, \ldots, X_L$ beyond $L = N$.

Proof (of (124)): We have

$$\mathbb{E}(X_1) = \mathbb{E}(X_2) = \frac{P - M}{P + M}$$

and

$$\mathbb{V}(X_1) = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 = 1 - \left(\frac{P - M}{P + M}\right)^2.$$

We compute

$$\mathbb{E}(X_1X_2) = \frac{P}{P+M} \frac{P-1}{P+M-1} + \frac{M}{P+M} \frac{M-1}{P+M-1}$$

$$\quad - \frac{P}{P+M} \frac{M}{P+M-1} - \frac{M}{P+M} \frac{P}{P+M-1}$$

$$= \frac{(P - M)^2 - (P + M)}{(P + M)^2 - (P + M)}$$

53
So, we conclude
\[
\begin{align*}
\mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) \\
= \frac{(P - M)^2 - (P + M)}{(P + M)(P + M - 1)} - \frac{(P - M)^2}{(P + M)^2} \\
= - \frac{1}{N-1} \left( \frac{(P - M)^2(P + M - 1)}{(P + M)^2} - \frac{(P - M)^2 - (P + M)}{P + M} \right) \\
= - \frac{1}{N-1} \left( 1 - \frac{(P - M)^2}{(P + M)^2} \right) \\
= - \frac{1}{N-1} \mathbb{V}(X_1) 
\end{align*}
\]

\[\square\]

4.2 A Law of Large Numbers

In this section, we consider averaged partial sums \( S_N = \frac{1}{N} \sum_{i=1}^{N} X_i \) of exchangeable random variables. For independent, identically distributed random variables this \( S_N \) converges in distribution to the probability measure carried by the single point \( \mathbb{E}(X_1) \) (under very mild conditions, see Theorem 3.1). For exchangeable random variables \( S_N \) converges in distribution as well (again under mild conditions). However, generally speaking, it converges to a more complicated measure, which is not concentrated in a single point.

**Theorem 4.5.** If \( \{X_i\}_{i \in \mathbb{N}} \) is a sequence of exchangeable random variables with moderately growing moments, then the normalized sums
\[
S_N = \frac{1}{N} \sum_{i=1}^{N} X_i
\]
(converge in distribution to a probability measure \( \mu \).

The support \( \text{supp}(\mu) \) of \( \mu \) is contained in the (possibly infinite) interval
\[
[\inf(\text{supp}P_{X_1}), \sup(\text{supp}P_{X_1})]
\]
and the moments \( m_k(\mu) \) of \( \mu \) are given by:
\[
m_k(\mu) = \mathbb{E}(X_1 \cdots X_k)
\]
**Definition 4.6.** We call the measure $\mu$ of Theorem 4.5 the *de Finetti measure* associated with the sequence $\{X_i\}_{i \in \mathbb{N}}$.

**Proof (Theorem 4.5):** An application of the Hölder inequality gives (see Lemma 2.65)
\[
|\mathbb{E}(X_1, X_2, \ldots, X_K)| \leq \mathbb{E}(|X_1|^K)
\]  
(127)
thus, according to Theorem 2.55 there is at most one measure with
\[
m_K(\mu) = \mathbb{E}(X_1 \cdot \ldots \cdot X_K)
\]
We compute
\[
\mathbb{E}(S_N^K) = \frac{1}{N^K} \mathbb{E}\left(\left(\sum_{i=1}^{N} X_i\right)^K\right)
\]
\[
= \frac{1}{N^K} \sum_{i_1,\ldots,i_K=1}^{N} \mathbb{E}(X_{i_1}, X_{i_2}, \ldots, X_{i_K})
\]
\[
= \frac{1}{N^K} \sum_{i_1,\ldots,i_K \text{ all distinct}} \mathbb{E}(X_1, X_2, \ldots, X_K)
\]
\[
+ \frac{1}{N^K} \sum_{i_1,\ldots,i_K \text{ not all distinct}} \mathbb{E}(X_{i_1}, X_{i_2}, \ldots, X_{i_K})
\]  
(128)
The first term in (128) equals
\[
\frac{1}{N^K} \frac{N!}{(N-K)!} \mathbb{E}(X_1, X_2, \ldots, X_K)
\]
which converges to
\[
\mathbb{E}(X_1, X_2, \ldots, X_K) \quad \text{as } N \to \infty
\]
since $\frac{1}{N^K} \frac{N!}{(N-K)!} \to 1$ for $K$ fixed and $N \to \infty$.
The second term in (128) can be bounded in absolute value by:
\[
(1 - \frac{1}{N^K} \frac{N!}{(N-K)!})\mathbb{E}(|X_1|^K)
\]
Consequently it converges to zero (for $N \to \infty$).
4.3 Random variables of de-Finetti-type

In this section we introduce a class of examples of exchangeable random variables with values in \{+1, −1\} (or any other set with exactly two elements). These example are “mixtures” of i.i.d. random variables in a sense we will make precise below.

**Definition 4.7.** For $-1 \leq r \leq 1$ let $P_r^{(1)}$ be the probability measure on $\{-1, 1\}$ with

$$P_r^{(1)}(1) = \frac{1}{2}(1 + r) \quad \text{and} \quad P_r^{(1)}(-1) = \frac{1}{2}(1 - r)$$  \hspace{1cm} (129)

such that the expectation of a $P_r^{(1)}$-distributed random variable $X_1$ is given by

$$E_r^{(1)}(X_1) = r$$

We also let $P_r^{(N)}$ denote the $N$-fold product measure

$$P_r^{(N)} = \bigotimes_{i=1}^{N} P_r^{(1)}$$  \hspace{1cm} (130)

on $\{-1, 1\}^N$, such that

$$P_r^{(N)}( (x_1, x_2, \ldots, x_N) ) = \prod_{i=1}^{N} P_r^{(1)}(x_i)$$

$$= \frac{1}{2^N} (1 + r)^{n_+(x_1, \ldots, x_N)} (1 - r)^{n_-(x_1, \ldots, x_N)}$$  \hspace{1cm} (131)

where $x_i \in \{-1, 1\}$ and $n_\pm(x_1, \ldots, x_N) = \#\{i \in \{1, \ldots, N\} \mid x_i = \pm1\}$. As usual we denote the expectation with respect to $P_r^{(N)}$ by $E_r^{(N)}$.

**Remark 4.8.** If $\xi_1, \xi_2, \ldots, \xi_N$ are independent random variables with common distribution $P_r^{(1)}$ then the $\xi_i$ can be realized on $\{-1, 1\}^N$ with measure $P_r^{(N)}$.

**Notation 4.9.** We write $P_r$ instead of $P_r^{(N)}$ if $N$ is clear from the context.

**Definition 4.10.** Let $\mu$ be a probability measure on $[-1, 1]$. We say that a sequence $X_1, \ldots, X_N$ of $\{-1, 1\}$-valued random variables is of de Finetti type with de Finetti measure $\mu$ if

$$\mathbb{P}(X_1 = a_1, X_2 = a_2, \ldots, X_N = a_N)$$

$$= \int_{-1}^{1} P_t^{(N)}( (a_1, a_2, \ldots, a_N) ) \, d\mu(t)$$

$$= \frac{1}{2^N} \int_{-1}^{1} (1 + t)^{n_+(a_1, \ldots, a_N)} (1 - t)^{n_-(a_1, \ldots, a_N)} \, d\mu(t)$$  \hspace{1cm} (132)
An infinite sequence $X_1, X_2, \ldots$ is called of de Finetti type if $X_1, \ldots, X_N$ is of de Finetti type for each $N \in \mathbb{N}$.

We define the probability measure $\mathbb{P}_\mu$ on $\{1,+1\}^N$ by

$$\mathbb{P}_\mu(x_1, x_2, \ldots, x_N) = \int_{-1}^{1} P_t^{(N)}((x_1, x_2, \ldots, x_N)) d\mu(t)$$

where

$$\mathbb{P}_\mu(x_1, x_2, \ldots, x_N)$$

is a short hand notation for

$$\mathbb{P}_\mu(\{\omega \in \{1,+1\}^N \mid \omega_1 = x_1, \omega_2 = x_2, \ldots, \omega_N = x_N\})$$

Remark 4.11. A sequence $X_1, X_2, \ldots$ of de Finetti type is obviously exchangeable. It is a remarkable result of Bruno de Finetti that ‘the converse’ is true, more precisely:

Any infinite $\{-1,1\}$-valued sequence of exchangeable random variables is actually of de Finetti type ([3], see also Theorem 4.16 below).

If we regard the random variables $X_1, \ldots, X_N$ as functions on the probability space $(\{-1,1\}^N, P(\{-1,1\}^N), P_t^{(N)})$ then the $X_1, \ldots, X_N$ are independent under the measure $P_t^{(N)}$ with common distribution $P_t^{(1)}$.

We denote the expectation with respect to $P_t = P_t^{(N)}$ by $E_t = E_t^{(N)}$ and the expectation with respect to $\mathbb{P}_\mu$ by $\mathbb{E}_\mu$.

Proposition 4.12. Suppose $X_1, \ldots, X_N$ is a sequence of random variables of de Finetti type with de Finetti measure $\mu$, then

1. For any function $F : \{-1,1\}^N \rightarrow \mathbb{R}$ we have

$$\mathbb{E}_\mu(F(X_1, X_2, \ldots, X_N)) = \int E_t(F(X_1, \ldots, X_N)) d\mu(t)$$

2. 

$$\mathbb{E}_\mu(X_1, X_2, \ldots, X_K) = \int_{-1}^{1} t^K d\mu(t) = m_K(\mu)$$
3. \( \text{Cov}_\mu(X_1, X_2) : = \mathbb{E}_\mu(X_1 X_2) - \mathbb{E}_\mu(X_1) \mathbb{E}_\mu(X_2) \)
\[ = \int t^2 d\mu(t) - \left( \int t \, d\mu(t) \right)^2 \geq 0 \quad (138) \]

4. \( X_1, \ldots, X_N \) are independent if and only if \( \mu \) is concentrated in a single point, i.e. \( \mu = \delta_m \), with \( m = \int t \, d\mu(t) \).

5. The random variables \( S_N = \frac{1}{N} \sum X_i \) converge in distribution to the measure \( \mu \).

\textbf{Proof:}

1) follows immediately from Definition 4.10.

2) From Definition 4.7 we get \( E_t(X_1) = t \) and consequently from the independence of the \( X_i \) (under \( P_\mu \))
\[ E_t(X_1, X_2, \ldots, X_K) = E_t(X_1)E_t(X_2) \ldots E_t(X_K) = t^K \quad (139) \]

3) With \( m := \int t \, d\mu(t) \) we compute
\[ \text{Cov}_\mu(X_1, X_2) = \int t^2 d\mu(t) - \left( \int t \, d\mu(t) \right)^2 \quad (140) \]
\[ = \int (t - m)^2 d\mu(t) \quad (141) \]
So, positivity follows.

4) The expression \( \int (t - m)^2 d\mu(t) \) can vanish only if \( \mu = \delta_m \). If the \( X_i \) are independent (with respect to \( P_\mu \)) the covariance \( \text{Cov}(X_1, X_2) \) must vanish, so \( \mu \) has to be \( \delta_m \).

On the other hand, if \( \mu = \delta_m \) then
\[ \mathbb{P}_\mu(X_1 = a_1, \ldots, X_N = a_N) = P_m(X_1 = a_1, \ldots, X_N = a_N) \quad (142) \]
hence the \( X_i \) are independent in this case.

5) We know already from Theorem 4.5 that \( S_N \) converges in distribution to a measure \( \nu \) with
\[ m_K(\nu) = \mathbb{E}_\nu(X_1 \cdot X_2 \cdot \ldots \cdot X_K) = m_K(\mu) \]
Hence \( \nu = \mu \). \qed

58
4.4 Sequences of de-Finetti-type

In this section we study sequences \( \{X^{(N)}_1, \ldots, X^{(N)}_N\}_{N \in \mathbb{N}} \) of de-Finetti-type with corresponding de-Finetti-measures \( \mu_N \) and their behavior for \( N \to \infty \).

**Definition 4.13.** For each \( N \in \mathbb{N} \) let \( X^{(N)}_1, \ldots, X^{(N)}_N \) be a sequence of de-Finetti-type with de-Finetti-measure \( \mu_N \).

Then the scheme
\[
\{ \{X^{(N)}_i\}_{i=1}^N \}_{N=1,2,\ldots}
\]
is called a *de-Finetti sequence*.

We say that a de Finetti sequence converges in distribution to a probability measure \( \mu \) on \( \{-1, 1\}^N \) if for each \( M \leq N \)
\[
P_{\mu_N}(x_1, \ldots, x_M) \to P_{\mu}(x_1, \ldots, x_M) \tag{143}
\]
In this case we write \( P_{\mu_N} \implies P_{\mu} \).

**Proposition 4.14.** If \( \mu_N \implies \mu \) then \( P_{\mu_N} \implies P_{\mu} \).

The proof follows immediately from the definition.

**Theorem 4.15.** If \( \mu_N \implies \mu \) then
\[
\frac{1}{N} \sum_{i=1}^{N} X_i^{(N)} \overset{D}{\implies} \mu \tag{144}
\]

**Proof:** We compute the moments of the random variables \( S_N = \frac{1}{N} \sum_{i=1}^{N} X_i^{(N)} \).

As in (128) we write
\[
E_{\mu_N}(S_N^K) = \frac{1}{N^K} \frac{N!}{(N-K)!} E_{\mu_N}(X_1 \cdot X_2 \cdot \ldots \cdot X_K) \\
+ \frac{1}{N^K} \sum_{i_1, \ldots, i_K \text{not all distinct}} E_{\mu_N}(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_K}) \tag{145}
\]

Since by assumption \( \mu_N \implies \mu \) we know that the first term of (145) converges to \( E_{\mu}(X_1 \cdot X_2 \cdot \ldots \cdot X_K) \). The second term can be bounded by \( \frac{1}{N^K} \left( 1 - \frac{N!}{(N-K)!} \right) \) and thus goes to zero as \( N \) goes to infinity. \( \square \)
4.5 The Theorem of de Finetti

Let \( X_1, X_2, \ldots \) be a sequence of \( \{-1, 1\}\)-valued random variables. Due to exchangeability the distribution of \( X_1, \cdots, X_N \) is determined by the quantities

\[
p_k^N = \mathbb{P}(X_1 = -1, X_2 = -1, \cdots, X_k = -1, X_{k+1} = 1, \cdots, X_N = 1) \quad (146)
\]

with \( k = 0, 1, \ldots, L \).

In particular,

\[
\mathbb{P}(\#\{i \leq L \mid X_i = -1\} = k) = \binom{N}{k} p_k^N \quad (147)
\]

and

\[
\mathbb{E}(X_1 \cdots X_N) = \sum_{k=0}^{N} (-1)^k \binom{N}{k} p_k^N \quad (148)
\]

**Theorem 4.16 (de Finetti).** If \( \{X_i\}_{i \in \mathbb{N}} \) is an infinite sequence of \( \{-1, 1\}\)-valued exchangeable random variables and \( \mu \) its associated de Finetti measure then

\[
\mathbb{P}(X_1 = a_1, X_2 = a_2, \ldots, X_N = a_N) = \int_{-1}^{1} p_t^{(N)}((a_1, a_2, \ldots, a_N)) \, d\mu(t) .
\]

In other words: Each infinite sequence of \( \{-1, 1\}\)-valued exchangeable random variables is of de Finetti type.

**Proof:** Given the measure \( \mu \) we know due to Theorem 4.5 that

\[
E_k := \mathbb{E}(X_1 \cdot X_2 \cdot \ldots \cdot X_k) = \int_{-1}^{1} t^k \, d\mu . \quad (149)
\]

Hence it is sufficient to show that the quantities \( p_k^N \) for \( k = 0, \ldots, N \) can be computed from the knowledge of the \( E_k \) (again for \( k = 0, \ldots, N \), where \( E_0 = 1 \)).

We prove this by induction over \( N \).

**Induction basis:** Let \( N = 1 \) then a short computation shows that

\[
p_0^1 = \frac{1}{2} (1 + \mathbb{E}(X_1))
\]

and

\[
p_1^1 = \frac{1}{2} (1 - \mathbb{E}(X_1))
\]
**Induction step:** Suppose, we can compute the numbers \( p_k^N \) from the knowledge of the \( E_k = \mathbb{E}(X_1 \cdot X_2 \cdot \ldots \cdot X_k) \) for \( k = 0, \ldots, N \).

We have:

\[
p_k^{N+1} = \mathbb{P}\left(X_1 = -1, \ldots, X_k = -1, X_{k+1} = 1, \ldots, X_N = 1, X_{N+1} = 1\right)
\]

and

\[
p_{k+1}^{N+1} = \mathbb{P}\left(X_1 = -1, \ldots, X_k = -1, X_{k+1} = -1, X_{k+2} = 1 \ldots, X_N = 1, X_{N+1} = 1\right)
= \mathbb{P}\left(X_1 = -1, \ldots, X_k = -1, X_{k+1} = 1, X_{k+2} = 1 \ldots, X_N = 1, X_{N+1} = -1\right)
\]

hence

\[
p_{k+1}^{N+1} + p_k^{N+1} = p_k^N \tag{150}
\]

It follows that the numbers \( p_{k+1}^{N+1} \) for \( k = 1, \ldots, N + 1 \) can be computed from the quantities \( p_k^N \) and the number \( p_0^{N+1} \).

More precisely, since \( p_{k+1}^{N+1} = p_{k-1}^N - p_{k-1}^{N+1} \) we conclude

\[
p_k^{N+1} = \sum_{l=1}^{k} (-1)^{l+1} p_{k-l}^N + (-1)^{k} p_0^{N+1} \tag{151}
\]

Setting

\[
A_k^N = \sum_{l=1}^{k} (-1)^{l+1} p_{k-l}^N \tag{152}
\]

and

\[
p_k^{L+1} = A_k^L + (-1)^k p_0^{L+1} \tag{153}
\]

we obtain

\[
\mathbb{E}(X_1 \cdot \ldots \cdot X_{N+1}) = \sum_{k=0}^{N+1} (-1)^k \binom{N+1}{k} p_k^{N+1}
= \sum_{k=0}^{N} A_k^N + \sum_{k=0}^{N+1} \binom{N+1}{k} p_0^{N+1}
= \sum_{k=0}^{N} A_k^N + 2^{N+1} p_0^{N+1} \tag{154}
\]
Note that $A_k^N$ depends only on the $p_k^N$, hence, by induction hypothesis, $A_k^N$ can be computed by $\mathbb{E}(X_1 \cdots X_L)$ with $L \leq N$. Hence, $p_0^{N+1}$ can be computed from the $\mathbb{E}(X_1, \cdots, X_L), L \leq N + 1$ by (154). Knowing $p_0^{N+1}$ we can compute the $p_k^{N+1}$ for $k = 1, \cdots, N + 1$ from (153). 

Remark 4.17 (Example 4.4 revisited). A finite sequence $X_1, X_2, \ldots, X_N$ of exchangeable random variables is not necessarily of de-Finetti-type. We have seen in Example 4.4 that we may have $\text{Cov}(X_i, X_j) < 0$. For de-Finetti-type random variables this is impossible due to Proposition 4.12.
The Curie-Weiss Model

5.1 Introduction

Probably the simplest nontrivial system of statistical physics is the Curie-Weiss model. It describes (a caricature of) a magnetic system where the $N$ elementary magnets $X_1, X_2, \ldots, X_N$ can take only two values: $+1$ (‘spin up’) and $-1$ (‘spin down’). The $X_i$ are $\{-1, +1\}$-valued random variables for which ‘alignment’ is more likely than non alignment, i.e. the probability that two random variables (‘spins’) $X_i$ and $X_j$ agree (‘are aligned’) is higher than for independent random variables. Such a behavior is typical for ‘ferromagnets’.

The distribution of $N$ independent $\{-1, 1\}$-valued random variables with a symmetric distribution (i.e. with $P_0(X_i = 1) = P_0(X_i = -1) = \frac{1}{2}$) is given by:

$$P_0\left(X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N\right) = \frac{1}{2^N}$$  \hspace{1cm} (155)

where we chose to call the probability measures $P_0$ rather than the generic $P$ to distinguish it from the measure $P_\beta$ to be defined below.

In statistical physics the distribution of elementary quantities $X_1, X_2, \ldots, X_N$ like spins is usually defined via an ‘energy functional’, i.e. a real valued function

$$H = H(x_1, x_2, \ldots, x_N)$$  \hspace{1cm} (156)

In our case $H$ is a function on $\{-1, +1\}^N$, the ‘configuration space’ of all possible spin combinations. It represents the ‘energy’ of a physical system of $N$ spins when the spins take values $x_1, x_2, \ldots, x_N \in \{-1, +1\}^N$. Physical systems are more likely in a configuration with small energy $H$.

A second fundamental quantity of statistical physics is ‘temperature’ $T$, which enters the distribution we are going to define via the quantity $\beta = \frac{1}{T}$ the ‘inverse temperature’. Low temperature (large $\beta$) means that the fluctuations of the system are small, while high temperature (small $\beta$) results in strong fluctuations of the system.

The probability distribution corresponding to the energy function $H$ at inverse temperature $\beta \geq 0$ is given by

$$P_\beta\left(X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N\right)$$

$$= \frac{1}{Z} e^{-\beta H(x_1, x_2, \ldots, x_N)} P_0\left(X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N\right)$$

$$= \frac{1}{Z} \frac{1}{2^N} e^{-\beta H(x_1, x_2, \ldots, x_N)}$$  \hspace{1cm} (157)
In formula (157), \( Z \) is a normalization constant which forces the total probability to be equal to 1, i.e.

\[
Z = \frac{1}{2^N} \sum_{x_1, x_2, \ldots, x_N \in \{-1, +1\}} e^{-\beta H(x_1, x_2, \ldots, x_N)}.
\]

We may write (157) in the shorter form:

\[
P_\beta = Z^{-1} e^{-\beta H} P_0
\]

i.e. the measure \( P_\beta \) has the density \( Z^{-1} e^{-\beta H(\cdot)} \) with respect to \( P_0 \).

Although \( Z = Z_{\beta N} \) obviously depends both on \( \beta \) and on \( N \), it is common to suppress these indices in the notation. Moreover, even when we change the model or its representation we usually keep the bad, but common habit to call the normalization constant \( Z \) although its value may change from line to line. So, whenever we write \( P = Z^{-1} e^{-\beta H} P_0 \) we implicitly agree that \( Z \) is the summation constant which make \( P \) a probability measure.

If we add a constant \( C \) to an energy function \( H \) the corresponding measures are the same, more precisely:

**Lemma 5.1.** If

\[
P_\beta = Z^{-1} e^{-\beta H} P_0 \quad \text{and} \quad \tilde{P}_\beta = \tilde{Z}^{-1} e^{-\beta (H+C)} P_0
\]

then \( P_\beta = \tilde{P}_\beta \)

**Proof:** We have

\[
Z = \sum_{x_1, x_2, \ldots, x_N \in \{-1, +1\}} e^{-\beta H(x_1, x_2, \ldots, x_N)} \quad \text{and} \quad \tilde{Z} = \sum_{x_1, x_2, \ldots, x_N \in \{-1, +1\}} e^{-\beta (H(x_1, x_2, \ldots, x_N) - C)}
\]

\[
= Z e^{\beta C}
\]

Hence

\[
\tilde{P}_\beta = \tilde{Z}^{-1} e^{-\beta (H+C)} P_0 = Z^{-1} e^{\beta C} e^{\beta H} e^{\beta C} P_0 = P_\beta
\]

A measure of the form (158) is called a canonical ensemble in physics. Canonical ensembles are used to describe thermodynamical systems in thermal equilibrium.
with a ‘heat bath’, i.e. for systems with a temperature $T$ which is kept fixed. At absolute temperature $T = 0$ the system will be in the state (or states) which minimizes (or minimize) the energy $H$, thermal fluctuations which may distract the system from equilibrium are absent for $T = 0$.

For $T > 0$ thermal fluctuations will distract the system from its energy minimum, but the probability $\mathbb{P}_\beta$ still has its maximum at the minimum of the energy $H$. If we increase $T$, the thermal fluctuations become bigger and bigger. In particular, if $T = \infty$ (i.e. if $\beta = 0$) the system is completely dominated by thermal fluctuations, the energy landscape $H$ become completely irrelevant. In deed, as one sees from the definition of $\mathbb{P}_\beta$, the measure $\mathbb{P}_0$ makes the random variables ('spins') independent.

In statistical physics, one is interested in the behavior of the measure $\mathbb{P}_\beta$ and various quantities thereof in the limit $N \to \infty$. For example the mean magnetization $\frac{1}{N} \sum_{i=1}^{N} X_i$ is an interesting quantity which we will investigate in the limit $N \to \infty$.

The physical properties of the system under consideration will certainly depend on the particular energy function $H$. In the following we will deal with a specific energy function, namely the one for the Curie-Weiss model. There is, of course, a variety of other spin models, most notably the celebrated Ising model. For these models we refer to the literature, for example [10], [9], [2] and references therein.

The energy for the Curie-Weiss model is given by

$$\tilde{H}(x_1, x_2, \ldots, x_N) = -\frac{1}{2} \sum_{i=1}^{N} \left( x_i \cdot \frac{1}{N} \sum_{j=1 \atop j \neq i}^{N} x_j \right)$$

Expression (159) makes it more likely ('preferable') for each spin $X_i$ to take the value 1 (resp. $-1$) if the average of all other spins is positive (resp. negative). In fact, in these cases the energy is smaller (there is a minus sign in front of the sum), thus the probability given in (158) is bigger. Thus we expect that there is a tendency for the spins to be aligned.

We may rewrite $\tilde{H}$ as

$$\tilde{H}(x_1, \ldots, x_N) = -\frac{1}{2} \frac{1}{N} \sum_{i,j=1 \atop i \neq j}^{N} x_i x_j$$
\[
\begin{align*}
&= -\frac{1}{2N} \sum_{i,j=1}^{N} x_i x_j + \frac{1}{2N} \sum_{i=1}^{N} x_i^2 \\
&= -\frac{1}{2N} \sum_{i,j=1}^{N} x_i x_j + \frac{1}{2} \\
&= -\frac{1}{2N} \left( \sum_{i=1}^{N} x_i \right)^2 + \frac{1}{2}
\end{align*}
\]

So, if we set
\[
H(x_1, \ldots, x_N) = -\frac{1}{2N} \left( \sum_{i=1}^{N} x_i \right)^2
\] (160)

we may write according to Lemma 5.1
\[
P_\beta = Z^{-1} e^{\beta H}
\] (161)

for the probability distribution associated with (159). In the following we will always work with the energy function (160) rather than with (159).

**Remark 5.2 (Warning).** When we deal with independent identically distributed random \( X_i \), it's of no importance whether we start with an infinite sequence \( \{X_i\}_{i \in \mathbb{N}} \) and restrict to \( X_1, \ldots, X_N \) or start with a finite sequence \( X_1, \ldots, X_{N'} \) with \( N' \geq N \) and restrict it to \( X_1, \ldots, X_N \).

In other words, there is a probability measure \( P \) associated to the infinite sequence \( \{X_i\}_{i \in \mathbb{N}} \) on the infinite dimensional space \( \mathbb{R}^N \) and the measures corresponding to the finite sequences \( X_1, \ldots, X_N \) are just the projections of \( P \) to finite dimensional subspaces.

In the case of the Curie-Weiss model this is not true. Above we introduced for any \( N \) the random variables \( X_1, X_2, \ldots, X_N \) ‘the’ Curie-Weiss measure \( P_\beta \). We have to admit now that this is an abuse of notation as ‘the’ measure \( P_\beta = P_\beta^{(N)} \) depends (explicitly) on \( N \), and so does \( H \)!

In fact there is not a single Curie-Weiss model, but there is sequence of Curie-Weiss models, a different one for each \( N \).

Moreover, if \( X_1, \ldots, X_N \) are distributed according to \( P_\beta^{(N)} \), then it is not true that the subsequence \( X_1, \ldots, X_M \) for \( M < N \) is distributed according to \( P_\beta^{(M)} \).

Even worse: There is no probability measure on the infinite dimensional space the restrictions of which give the distribution of the finite subsequences. In fact, we can not define what it means for an infinite sequence to be Curie-Weiss distributed.
Consequently, it would make sense to make this dependence on $N$ explicit by writing $X^{(N)}_1, X^{(N)}_2, \ldots, X^{(N)}_N$ instead of $X_1, X_2, \ldots, X_N$ and $\mathbb{P}^{(N)}_\beta$ rather than $\mathbb{P}_\beta$.

Following the tradition we will not do so most of the time hoping that the confusion of the reader will be limited. In fact, we will be interested in properties of ‘the’ Curie-Weiss model which ‘hold true for large $N$’.

We formalize the above considerations in the following definition.

**Definition 5.3.** For any integer $N \in \mathbb{N}$ and any $\beta \geq 0$ we define the $N$-spin Curie-Weiss measure $\mathbb{P}^{(N)}_\beta$ with inverse temperature $\beta$ on $\{-1, +1\}^N$ by

$$\mathbb{P}^{(N)}_\beta \left( x_1, x_2, \ldots, x_N \right) = \frac{1}{Z^{(N)}_\beta} e^{\beta \frac{1}{N} \left( \sum_{i=1}^{N} x_i \right)^2}$$

(162)

where

$$Z^{(N)}_\beta = \sum_{x_1, x_2, \ldots, x_N \in \{-1, +1\}} e^{\beta \frac{1}{N} \left( \sum_{i=1}^{N} x_i \right)^2}$$

The Curie-Weiss model (for given $N$ and $\beta$) consists of $\mathbb{P}^{(N)}_\beta$-distributed random variables $X^{(N)}_1, X^{(N)}_2, \ldots, X^{(N)}_N$.

**Notation 5.4.** Whenever there is no danger of confusion we drop the superscript $N$ in the above defined quantities.

The Curie-Weiss model is interesting because on the one hand it is simple enough to allow many explicit calculations and on the other hand it is complex enough to show the important phenomenon of a phase transition. A ‘phase transition’ occurs if some physical quantity, for example the magnetization $\frac{1}{N} \sum X_i$, changes its behaviour (in the $N \to \infty$ limit) qualitatively if the inverse temperature $\beta$ exceeds a certain threshold. We will discuss this phenomenon in detail below.

The Curie-Weiss model can also be used to model voting behavior in a large set of voters which can only vote ‘Yes’ or ‘No’. This model describes an interaction between the voters who try to convince each other. For further reading on this aspect see [5] and [6].

### 5.2 Curie-Weiss as a model of de Finetti type

From the very definition of the Curie-Weiss model it is clear that it is a sequence of exchangeable random variables. However, it is not so obvious that it is actually
a sequence of de Finetti type. That this is indeed the case is the main topic of the current section.

We start with a lemma which will allow us to rewrite the Curie-Weiss probability in the desired form.

**Lemma 5.5.** For each \( a \in \mathbb{R} \) we have

\[
e^{\frac{a^2}{2}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{s^2}{2} + as} ds
\]  

(163)

**Proof:**

\[
\sqrt{2\pi} = \int_{-\infty}^{+\infty} e^{-\frac{s^2}{2}} ds = \int_{-\infty}^{+\infty} e^{-\frac{(s-a)^2}{2}} ds = e^{-\frac{a^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{s^2}{2} + as} ds
\]

\( \square \)

**Theorem 5.6.** If \( \mathbb{P}_{\beta}^{(N)} \) denotes the Curie-Weiss measure on \( \{-1, 1\}^N \) then

\[
\mathbb{P}_{\beta}^{(N)}(x_1, \ldots, x_N) = Z^{-1} \int_{-1}^{1} P_t^{(N)}(x_1, \ldots, x_N) \frac{e^{-\frac{N}{2} F_\beta(t)}}{1 - t^2} dt
\]  

(164)

where

\[
F_\beta(t) = \frac{1}{\beta} \left( \frac{1}{2} \ln \frac{1 + t}{1 - t} \right)^2 + \ln (1 - t^2)
\]  

(165)

and

\[
Z = \int_{-1}^{1} e^{-\frac{N}{2} F_\beta(t)} dt.
\]  

(166)

**Remark 5.7.** Theorem 5.6 says that \( \mathbb{P}_{\beta}^{(N)} \) is a measure of de Finetti type with de Finetti measure

\[
\mu(dt) = \frac{1}{Z} \frac{e^{-\frac{N}{2} F_\beta(t)}}{1 - t^2} dt \quad \text{on} \quad [-1, 1].
\]
Proof: Take \((x_1, \ldots, x_N) \in \{-1, +1\}^N\). Using Lemma 5.5 we write

\[
e^{\frac{\beta}{2\pi} (\sum x_i)^2} = (2\pi)^{-\frac{N}{2}} \int_{-\infty}^{+\infty} e^{-\frac{\beta}{2\pi} \sqrt{\frac{\beta}{N}} (\sum x_i)s} ds
\]

\[
= \left( \frac{N}{2\pi\beta} \right)^{\frac{N}{2}} \int_{-\infty}^{+\infty} e^{-\frac{N}{2\pi} \beta y^2} \prod_{i=1}^{N} e^{x_iy} dy
\]

where we put \(y = \sqrt{\frac{\beta}{N}} s\).

Now we change variables setting \(t = \tanh y\). Below (Lemma 5.8) we compute

\[
y = \tanh^{-1} t = \frac{1}{2} \ln \left( \frac{1 + t}{1 - t} \right)
\]

\[
dt = (1 - \tanh^2 y) dy
\]

\[
\ln \cosh y = -\frac{1}{2} \ln(1 - \tanh^2 y)
\]

\[
e^y \frac{2}{\cosh y} = \frac{1}{2} (1 + t) \quad \text{and} \quad e^{-y} \frac{2}{\cosh y} = \frac{1}{2} (1 - t)
\]

Thus we obtain

\[
\int e^{-\frac{N}{2\pi} \beta y^2 - \ln \cosh y} \prod_{i=1}^{N} \frac{e^{x_iy}}{2 \cosh y} dy
\]

\[
= \int e^{-\frac{N}{2\pi} \beta \left( \frac{1}{2} \ln \frac{1 + t}{1 - t} \right)^2 + \ln(1 - t^2)} \left( \prod_{i=1}^{N} \frac{1}{2} \left( 1 + x_i t \right) \right) \cdot \frac{1}{1 - t^2} dt
\]

\[
= \int \mathcal{P}_t^{(N)}(x_1, \ldots, x_N) \frac{e^{-\frac{N}{2\pi} F_\beta(t)}}{1 - t^2} dt \quad (167)
\]

With the correct normalizing constant \(Z\) this proves the assertion. \(\Box\)

Lemma 5.8. For \(x \in \mathbb{R}\) and \(|t| < 1\) we have

1. \(\tanh^{-1} t = \frac{1}{2} \ln \left( \frac{1 + t}{1 - t} \right)\)

2. \(\tanh'(x) = 1 - \tanh^2 x\)
3. \[ \ln \cos x = \frac{1}{2} \ln (1 - \tanh^2 x) \]

4. \[ \frac{e^x}{2 \cosh x} = \frac{1}{2} (1 + \tanh x) \quad \frac{e^{-x}}{2 \cosh x} = \frac{1}{2} (1 - \tanh x) \]

**Proof:**

1. \[ \frac{1}{2} \ln \left( \frac{1 + \tanh x}{1 - \tanh x} \right) = \frac{1}{2} \ln \left( \frac{\cosh x + \sinh x}{\cosh x - \sinh x} \right) = \frac{1}{2} \ln \frac{2e^x}{2e^{-x}} = x \]

2. \[ \tanh'(x) = \frac{d}{dx} \left( \frac{\sinh(x)}{\cosh(x)} \right) = \frac{\cosh^2(x) - \sinh^2(x)}{\cosh^2(x)} = 1 - \tanh^2(x) \]

3. \[ e^{-\frac{1}{2} \ln (1 - \tanh^2 x)} = e^{-\ln(1 - \tanh^2 x)\frac{1}{2}} = \frac{1}{(1 - \tanh^2 x)^{\frac{1}{2}}} = \cosh x \]

4. \[ \frac{e^x}{\cosh x} = \frac{1}{2} \frac{(e^x + e^{-x}) + (e^x - e^{-x})}{\cosh x} = \frac{\cosh x + \sinh x}{\cosh x} = 1 + \tanh x \]
Proposition 5.9. Suppose $X_1, X_2, \ldots, X_K$ are $P^{(N)}_\beta$-distributed Curie-Weiss random variables, then

$$E^{N}_\beta (X_1 X_2 \ldots X_K) = Z^{-1} \int_{-1}^{1} t^K \frac{e^{-\frac{2\pi}{N} F_\beta(t)}}{1-t^2} \, dt$$

where again

$$F_\beta(t) = \frac{1}{\beta} \left( \frac{1}{2} \ln \frac{1+t}{1-t} \right)^2 + \ln(1-t^2)$$

and $E^{N}_\beta$ denotes expectation with respect to $P^{(N)}_\beta$.

Proof: From Theorem 5.6 we know that

$$E^{N}_\beta (X_1 X_2 \ldots X_K) = Z^{-1} \int_{-1}^{1} E^{(N)}_t(X_1, \ldots, X_N) \frac{e^{-\frac{2\pi}{N} F_\beta(t)}}{1-t^2} \, dt$$

= $$Z^{-1} \int_{-1}^{1} t^K \frac{e^{-\frac{2\pi}{N} F_\beta(t)}}{1-t^2} \, dt$$

where we used that the random variables $X_1, X_2, \ldots, X_K$ are independent under the probability measure $P^{(N)}_t$ and $E^{(N)}_t(X_j) = t$ for all $j$.

5.3 Laplace’s method

The Laplace method is a technique to analyze integrals of the form

$$\int e^{-\frac{2\pi}{N} F(x)} \phi(x) \, dx$$

asymptotically as $N \to \infty$.

The main idea of the method is the insight that the leading contribution of the integral comes from the minimum of the function $F$.

For example, if $F(x) \approx F''(0) x^2$ and $\phi(x) \equiv 1$ then we might expect that

$$\int_{0}^{\infty} e^{-\frac{2\pi}{N} F(x)} \, dx \approx \int_{0}^{\infty} e^{-\frac{2\pi}{N} F''(0) x^2} \, dx$$

$$= \frac{1}{\sqrt{N F''(0)}} \int_{0}^{\infty} e^{y^2/2} \, dy$$

$$= \frac{1}{2} \sqrt{\frac{2\pi}{F''(0)}} \frac{1}{\sqrt{N}}.$$
The following theorem makes the above intuition rigorous.

**Theorem 5.10** (Laplace Method). Suppose \( a \in \mathbb{R}, b \in \mathbb{R} \cup \{\infty\}, a < b \) and \( F : [a, b) \to \mathbb{R} \) is a function whose absolute minimum is at \( x = a \).

Assume:

1. For every \( \delta > 0 \)
   \[
   \inf_{x > a + \delta} F(x) > F(a) \tag{170}
   \]

2. There are constants \( A \) and \( C \) and \( \delta_0 > 0 \) and \( k > 0 \) and \( \eta > 0 \) such that for all \( x \) with \( a \leq x \leq a + \delta_0 \)
   \[
   \left| (F(x) - F(a)) - A(x - a)^k \right| \leq C(x - a)^{k+\eta} \tag{171}
   \]

3. \( \phi_N, \phi : [a, b) \to \mathbb{R} \) are functions such that for a \( \delta_1 > 0 \)
   \[
   \sup_{x \in [a, a + \delta_1]} |\phi_N(x) - \phi(x)| \to 0 \tag{172}
   \]
   and \( \phi(x) \) is continuous at \( x = a \)

4. There is a function \( \tilde{\phi} : [a, b) \to \mathbb{R} \) such that
   \[
   |\phi(x)| + |\phi_N(x)| \leq \tilde{\phi}(x) \tag{173}
   \]
   for all \( N \) and \( x \in [a, b) \) and
   \[
   \int_a^b e^{-\frac{N}{2}F(x)}(x - a)^\ell \tilde{\phi}(x) \, dx < \infty \tag{174}
   \]
   for all \( N \geq N_0 \) and an \( \ell \geq 0 \).

Then we have:

\[
\lim_{N \to \infty} e^{\frac{N}{2}F(a)}(AN)^{\frac{\ell+1}{k}} \int_a^b e^{-\frac{N}{2}F(x)}(x - a)^\ell \phi_N(x) \, dx
\]

\[= \phi(a) \int_0^\infty e^{-\frac{y^k}{2}} y^\ell \, dy \tag{175}\]

\[= \phi(a) \frac{2^{\frac{\ell+1}{k}}}{k} \Gamma\left(\frac{\ell + 1}{k}\right) \tag{176}\]
For the reader’s convenience we recall

**Definition 5.11.** For \( x > 0 \) we define the \textit{Gamma function} \( \Gamma \) by

\[
\Gamma(x) := \int_{0}^{\infty} t^{x-1} e^{-t} \, dt
\]

(177)

**Proof (Theorem 5.10):** Without loss of generality we assume that \( a = 0 \) and \( F(a) = 0 \).

We have

\[
\int_{0}^{b} e^{-\frac{N}{2} F(x)} x^{\ell} \phi_{N}(x) \, dx = \int_{0}^{\delta} e^{-\frac{N}{2} F(x)} x^{\ell} \phi_{N}(x) \, dx + \int_{\delta}^{b} e^{-\frac{N}{2} F(x)} x^{\ell} \phi_{N}(x) \, dx
\]

(178)

for a \( 0 < \delta < \delta_0 \) to be chosen later.

The second integral can be estimated using Assumptions [1] and [4] assuming in addition \( F(x) \geq B > 0 \) for \( x > \delta \)

\[
\int_{\delta}^{b} e^{-\frac{N}{2} F(x)} x^{\ell} \tilde{\phi}(x) \, dx \leq e^{-\frac{N}{2} B} \int_{\delta}^{b} e^{-\frac{N}{2} F(x)} x^{\ell} \tilde{\phi}(x) \, dx.
\]

(179)

By Assumption [4] this term goes to zero, even if multiplied by \( N^{\ell+1} \).

We turn to the first integral in (178): For \( x \leq \delta_0 \) we may write \( F(x) = Ax^k + r(x) \) with \( |r(x)| \leq Cx^k + \eta \) according to Assumption [2].

We have

\[
\int_{0}^{\delta} e^{-\frac{N}{2} F(x)} x^{\ell} \phi_{N}(x) \, dx = \int_{0}^{\delta} e^{-\frac{N}{2} (Ax^k + r(x))} x^{\ell} \phi_{N}(x) \, dx
\]

\[
= (NA)^{-\frac{\ell+1}{2}} \delta (NA)^{\frac{k}{2}} \int_{0}^{\delta (NA)^{\frac{1}{k}}} e^{-\frac{N}{2} y^k} e^{-\frac{N}{2} r(\frac{y}{(NA)^{\frac{1}{k}}})} y^{\ell} \phi_{N}(\frac{y}{(NA)^{\frac{1}{k}}}) \, dy
\]

(180)

where we put \( y = (NA)^{\frac{1}{k}} x \).

Since \( \phi_{N} \rightarrow \phi \) uniformly in a neighborhood of 0 and \( \phi \) is continuous there we conclude that

\[
\phi_{N}(\frac{y}{(NA)^{\frac{1}{k}}}) \rightarrow \phi(0) \quad \text{as} \quad N \rightarrow \infty.
\]
Moreover, since
\[ r \left( \frac{y}{(NA)\frac{k}{2}} \right) \leq C \frac{y^{k+\eta}}{(NA)^{\frac{k+\eta}{2}}} \]  
(181)

We learn that
\[ Nr \left( \frac{y}{(NA)\frac{k}{2}} \right) \to 0 \]

thus
\[ e^{-\frac{N}{2}r \left( \frac{y}{(NA)\frac{k}{2}} \right)} \to 1 \quad \text{for each } y. \]

Consequently, the integral
\[
\int_0^{\delta(NA)\frac{k}{2}} e^{-\frac{y^k}{2}} e^{-\frac{N}{2}r \left( \frac{y}{(NA)\frac{k}{2}} \right)} y^\ell \phi_N \left( \frac{y}{(NA)\frac{k}{2}} \right) dy
\]
\[= \int_0^\infty e^{-\frac{y^k}{2}} \chi_{[0,\delta(NA)\frac{k}{2}]}(y) e^{-\frac{N}{2}r \left( \frac{y}{(NA)\frac{k}{2}} \right)} y^\ell \phi_N \left( \frac{y}{(NA)\frac{k}{2}} \right) dy \]  
(182)

converges to \( \phi(0) \int_0^\infty e^{-\frac{y^k}{2}} y^\ell dy \) provided we can show that the integrand in (182) is dominated by an integrable function.

Due to the uniform convergence of \( \phi_N \) and the continuity (at 0) of \( \phi \) the term \( \phi_N \left( \frac{y}{(NA)\frac{k}{2}} \right) \) is uniformly bounded for \( y \in [0, \delta(NA)\frac{k}{2}] \), if \( \delta \) is small enough.

Again for \( y < \delta(NA)\frac{k}{2} \) we estimate using Assumption 2 and (181)
\[
N \left| r \left( \frac{y}{(NA)\frac{k}{2}} \right) \right| \leq \frac{CN}{A^{\frac{k+n}{2}} N^{\frac{k+n}{2}}} y^\eta y^k \leq \frac{C}{A^{\frac{k+n}{2}} N^{\frac{k+n}{2}}} \delta^\eta N^\frac{\eta}{2} A^\frac{\eta}{2} y^k \leq \frac{1}{4} y^k \]  
(183)

if \( \delta \) is small enough.

Thus we have shown that for suitable \( \delta > 0 \) the integrand in (182) is dominated by the integrable function
\[ Dy^\ell e^{-\frac{y^k}{2}} \]

where \( D \) is a constant. This justifies the interchanging of the limit and the integral above and thus finishes the proof (175). To prove (176) we make the change of variables \( t = \frac{y^k}{2} \) in (176).
The conditions on the function $F$ in Theorem 5.10 can be checked easily for smooth functions. The following corollary is a typical example for the use of this strategy.

**Corollary 5.12.** Suppose $F : (c, d) \to \mathbb{R}$ (with $c \in \mathbb{R} \cup \{-\infty\}$ and $d \in \mathbb{R} \cup \{\infty\}$) is $k + 1$-times continuously differentiable for an even $k$ and assume that for some $a \in (c, d)$:

$$F(a) = F'(a) = \cdots = F^{(k-1)}(a) = 0 \quad \text{and} \quad F^k(a) > 0 \quad (184)$$

and $F'(x) \neq 0$ for $x \neq a$.

Suppose furthermore that $\phi : (c, d) \to \mathbb{R}$ is continuous at $a$ with $\phi(a) \neq 0$ and such that

$$\int_c^d e^{-F(x)} |\phi(x)| |x - a|^\ell \, dy < \infty.$$

Then

1. For even $\ell$

$$\int_c^d e^{-\frac{N}{2} F(x)} (x - a)^\ell \phi(x) \, dx \approx \phi(a) \left( \frac{k!}{F^k(a)} \right)^{\frac{\ell + 1}{2}} \left( \frac{2}{N} \right)^{\frac{\ell + 1}{2}} \frac{2}{k} \Gamma\left( \frac{\ell + 1}{k} \right) \quad \text{as} \quad N \to \infty \quad (185)$$

2. For odd $\ell$

$$N^{\frac{\ell + 1}{k}} \int_c^d e^{-\frac{N}{2} F(x)} (x - a)^\ell \phi(x) \, dx \to 0 \quad (186)$$

**Notation 5.13.** Above we used (and will use throughout) the expression

$$a(N) \approx b(N) \quad \text{as} \quad N \to \infty \quad (187)$$

as a shorthand notation for

$$\lim_{N \to \infty} \frac{a(N)}{b(N)} = 1 \quad (188)$$

**Proof (Corollary):** The assumptions on $F$ imply that $F$ has a unique minimum at $x = a$ and $\inf_{|x - a| \geq \delta} F(x) > 0$ for every $\delta > 0$. 

75
By Taylor’s formula we have
\[ F(x) = \frac{F^{(k)}(a)}{k!} (x - a)^k + r(x - a) \]

with \( |r(x)| \leq C |x|^{k+1} \) near \( a \) since \( F^{(k+1)} \) exists and is continuous.

Thus we may apply Theorem 5.10 to both the integrals
\[ \int_a^d e^{-\frac{N}{2} F(x)} (x - a)^\ell \phi(x) dx \quad \text{and} \quad \int_c^a e^{-\frac{N}{2} F(x)} (x - a)^\ell \phi(x) dx \]

For even \( \ell \) we have
\[ \int_c^a e^{-\frac{N}{2} F(x)} (x - a)^\ell \phi(x) dx = -\int_{-a}^{-c} e^{-\frac{N}{2} F(x)} (x - a)^\ell \phi(-x) dx \]

Since these integrals give the same asymptotic result, part (i) of the Corollary follows from Theorem 5.10.

If \( \ell \) is even then
\[ \int_c^a e^{-\frac{N}{2} F(x)} (x - a)^\ell \phi(x) dx = -\int_{-a}^{-c} e^{-\frac{N}{2} F(x)} (x - a)^\ell \phi(-x) dx \]

and Theorem 5.10 implies part (ii). \( \square \)

As an illustration of the Laplace method we make a slight detour and prove the Stirling formula, a well known asymptotic expression for \( N! \).

We first need:

**Lemma 5.14.**

1. For all \( x > 0 \) we have \( \Gamma(x + 1) = x \Gamma(x) \).
2. For \( N \in \mathbb{N} \) we have \( N! = \Gamma(N + 1) \)

**Proof:** For all \( x > 0 \) we compute using integration by parts
\[ \Gamma(x + 1) = \int_0^\infty t^x e^{-t} dt = x \int_0^\infty t^{x-1} e^{-t} dt \] (189)
This proves 1.

We prove 2. by induction

\[
\Gamma(1) = \int_0^{\infty} e^{-t} dt = 1 = 0! \quad (190)
\]

For the induction step we use 1.:

\[
\Gamma(N + 1) = N \cdot \Gamma(N) = N \cdot (N - 1)! = N!
\]

\[\square\]

**Proposition 5.15. (Stirling Formula)**

\[
N! \approx N^N e^{-N} \sqrt{2\pi N} \quad \text{as} \quad N \to \infty. \quad (191)
\]

**Proof:**

\[
N! = \int_0^{\infty} t^N e^{-t} dt = N^{N+1} \int_0^{\infty} s^N e^{-Ns} ds
\]

where we changed variables \( t = Ns \)

\[
= N^{N+1} \int_0^{\infty} e^{-N(s-\ln s)} ds
\]

Setting \( F(s) = s - \ln s \) we find that \( F \) has a strict minimum at \( s = 1 \) and \( F''(1) = 1 \).

Applying Corollary 5.12 we obtain

\[
\int_0^{\infty} e^{-N(s-\ln s)} ds \approx e^{-N} \frac{\sqrt{2}}{N^{\frac{3}{2}}} \int_0^{\infty} e^{-s^2} ds
\]

We compute

\[
\int_0^{\infty} e^{-s^2} ds = \frac{1}{\sqrt{2}} \int_0^{\infty} e^{-x^2} dx = \frac{1}{2} \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} e^{-x^2} dx = \frac{1}{2} \sqrt{\pi}
\]

This gives the result. \[\square\]
5.4 Correlations for the Curie-Weiss model

In this section we compute the correlations for Curie-Weiss random variables asymptotically. To do so we use (168) of Proposition 5.9 to estimate the correlations

\[ E^{(N)}_\beta(X_1 \cdot X_2 \cdot \ldots \cdot X_k) \]  

via Laplace’s method.

In order to apply Corollary 5.12 we have to find and analyze the minima of the function

\[ F_\beta(t) = \frac{1}{\beta} \left( \frac{1}{2} \ln \frac{1 + t}{1 - t} \right)^2 + \ln(1 - t^2) \]

for \( t \in (-1, 1) \).

We show:

**Proposition 5.16.**

1. For \( \beta < 1 \) the function \( F_\beta \) has a unique minimum at \( t = 0 \). \( F'_\beta(0) = 0 \) and \( F''_\beta(0) = 2 \left( \frac{1}{\beta} - 1 \right) > 0 \)

2. For \( \beta = 1 \) the function \( F_\beta \) has a unique minimum at \( t = 0 \). \( F'_\beta(0) = F''_\beta(0) = F'''_\beta(0) = 0 \) and \( F^{(iv)}_\beta(0) = 4 > 0 \)

3. For \( \beta > 0 \) the function \( F_\beta \) has a unique minimum in \([0, 1]\) at \( t_0 > 0 \) and a unique minimum in \((-1, 0]\) at \(-t_0 < 0\). \( F'_\beta(t_0) = F'_\beta(-t_0) = 0 \) and \( F''_\beta(t_0) = F''_\beta(-t_0) > 0 \). \( t_0 \) is the unique strictly positive solution of \( t \) = \( \tanh(\beta t) \) (see 5.8). \( F_\beta \) has a local maximum at \( t = 0 \) with \( F'_\beta(0) = 0 \) and \( F''_\beta(0) = 2 \left( \frac{1}{\beta} - 1 \right) < 0 \).

**Proof:** We compute two derivatives of \( F_\beta \):

\[ F'_\beta(t) = \frac{1}{1 - t^2} \left( \frac{1}{\beta} \ln \frac{1 + t}{1 - t} - 2t \right) \]

\( F'_\beta(t) = 0 \) is equivalent to

\[ \frac{1}{2} \ln \frac{1 + t}{1 - t} = \beta t \]

Since \( \frac{1}{2} \ln \frac{1 + t}{1 - t} = \tanh^{-1} t \) (see 5.8) this is equivalent to

\[ \tanh(\beta t) = t \]
This is satisfied for $t = 0$ for all $\beta$.

$$F_{\beta}''(t) = \frac{2}{(1 - t^2)^2} \left( \left( \frac{1}{\beta} - (1 - t^2) \right) - 2t \left( \frac{1}{2\beta} \ln \frac{1 + t}{1 - t} - t \right) \right)$$

(197)

Thus

$$F_{\beta}''(0) = 2 \left( \frac{1}{\beta} - 1 \right) \begin{cases} > 0 & \text{for } \beta < 1 \\ = 0 & \text{for } \beta = 1 \\ < 0 & \text{for } \beta > 1 \end{cases}$$

(198)

Consequently $t = 0$ is a local minimum for $\beta < 1$ and a local maximum for $\beta > 1$.

Let us analyze the case $\beta < 1$. Due to the symmetry $F_{\beta}(t) = F_{\beta}(-t)$ it suffices to look at $t \geq 0$. Setting $g(t) = \tanh \beta t$ and $f(t) = t$ we see that

$$g(0) = f(0) = 0$$

(199)

and

$$g'(t) = \frac{\beta}{\cosh^2(\beta t)} < 1 = f'(t)$$

(200)

for all $t \in [0, 1)$ and $\beta < 1$ thus there is no strictly positive solution for

$$\tanh \beta t = t$$

(201)

if $\beta < 1$.

Moreover, the same argument shows that $t > \tanh \beta t$ for $\beta < 1$ and all $t \in (0, 1)$. It follows that $F_{\beta}(t)$ is strictly positive for $t \in (0, 1)$ thus $F_{\beta}(t)$ is strictly monotone there.

We turn to the case $\beta > 1$. Under this assumption

$$g(0) = f(0) = 0$$

(202)

$$g'(0) = \beta > 1 = f'(t)$$

(203)

and

$$g(t) \to 1, f(t) \to \infty \text{ as } t \to \infty$$

(204)

Thus there exists a strictly positive solution $t_0$ of

$$\tanh \beta t = t$$

(205)

Since $\tanh \beta t < 1$ we conclude $t_0 < 1$. 

79
Moreover, we have that $g'(t) = \frac{\beta}{\cosh^2(\beta t)}$ is decreasing for $t \geq 0$, hence the positive solution $t_0$ is unique. We have $g(t) > t$ for $0 < t < t_0$ and $g(t) < t$ for $t > t_0$. It follows that $g'(t_0) < 1$.

Using that $\frac{1}{2} \ln \frac{1 + t_0}{1 - t_0} = \beta t_0$ we compute

$$F^{\prime\prime}_\beta(t_0) = \frac{2}{(1 - t_0^2)^2} \left( \frac{1}{\beta} - (1 - t_0^2) \right)$$

and

$$\frac{1}{\beta} - (1 - t_0^2) = \frac{1}{\beta} - (1 - \tanh^2 \beta t_0) = \frac{1}{\beta} - \frac{1}{\beta} g'(t_0) > 0$$

The case $\beta = 1$ involves a straight forward but tedious computation. We got to compute $F^{\prime\prime}_1(t)$ and $F^{iv}_1(t)$. It turns out that $F^{\prime\prime}_1(0) = 0$ and $F^{iv}_1(0) = 4$.

Moreover, $F_1(t) > 0$ for $t > 0$, hence $F_1(t)$ is strictly monotone for $t \in (0, 1)$.

**Theorem 5.17.** Suppose $X_1, X_2, \cdots, X_\ell$ are $\mathbb{P}_\beta^{(N)}$-distributed Curie-Weiss random variables.

If $\ell$ is even, then

1. if $\beta < 1$

$$\mathbb{E}_\beta^{(N)}(X_1 \cdot X_2 \cdot X_\ell) \approx (\ell - 1)!! \left( \frac{\beta}{1 - \beta} \right)^\ell \frac{1}{N^{\ell/2}} \quad \text{as } N \to \infty$$

2. if $\beta = 1$ there is a constant $c_\ell$ such that

$$\mathbb{E}_\beta^{(N)}(X_1 \cdot X_2 \cdot X_\ell) \approx c_\ell \frac{1}{N^{\ell/2}}$$

3. if $\beta > 1$

$$\mathbb{E}_\beta^{(N)}(X_1 \cdot X_2 \cdot X_\ell) \approx m(\beta)^\ell$$

where $t = m(\beta)$ is the strictly positive solution of $\tanh \beta t = t$.

If $\ell$ is odd then $\mathbb{E}_\beta^{(N)}(X_1 \cdot X_2 \cdot X_\ell) = 0$ for all $\beta$.

**Remark 5.18.** In the case $\beta = 1$ the constant $c_\ell$ is given by $12^\ell \frac{\Gamma(\ell + 1)}{\Gamma(\ell + 4)}$ as the proof will show.
Proof: We set
\[ Z_N(\ell) = \int_{-1}^{1} \frac{e^{-t N F_\beta(t)}}{1 - t^2} t^\ell \, dt \]  
(211)

Since \( F_\beta(t) = F_\beta(-t) \) we know that \( Z_k(\ell) = 0 \) for \( \ell \) odd. For even \( \ell \) we apply Corollary 5.12. Together with Proposition 5.16 we obtain the following:

For \( \beta < 1 \)
\[ Z_N(\ell) \approx \left( \frac{4}{F_\beta'(0)} \right)^{\frac{\ell + 1}{2}} \Gamma(\frac{\ell + 1}{2}) N^{-\frac{\ell + 1}{2}} \]  
(212)
\[ = \left( \frac{\beta}{1 - \beta} \right)^{\frac{\ell + 1}{2}} \Gamma\left( \frac{\ell + 1}{2} \right) N^{-\frac{\ell + 1}{2}} 2^{\frac{\ell + 1}{2}} \]  
(213)

Consequently
\[ E_\beta^{(N)}(X_1 \cdot X_2 \cdot X_\ell) = \frac{Z_N(\ell)}{Z_N(0)} \]  
(214)
\[ \approx \frac{\Gamma\left( \frac{\ell + 1}{2} \right)}{\Gamma\left( \frac{1}{2} \right)} \left( \frac{\beta}{1 - \beta} \right)^{\frac{\ell + 1}{4}} N^{-\frac{\ell + 1}{4}} 2^{\frac{\ell + 1}{2}} \]  
(215)

Using the fact that for even \( \ell \)
\[ 2^\frac{\ell}{2} \frac{\Gamma\left( \frac{\ell + 1}{2} \right)}{\Gamma\left( \frac{1}{2} \right)} = (\ell - 1)!! \]  
(216)
(a proof will be given below).

We arrive at (208). For \( \beta = 1 \) we know from Proposition 5.16 that \( F_\beta^{(k)}(0) = 0 \) for \( k = 0, 1, 2, 3 \) and \( F_\beta^{(4)}(0) = 4 > 0 \).

Hence Corollary 5.12 gives
\[ Z_N(\ell) \approx 12^\frac{\ell + 1}{4} \cdot \frac{1}{2} \cdot \Gamma\left( \frac{\ell + 1}{4} \right) \cdot N^{-\frac{\ell + 1}{4}} \]  
(217)

This gives the result for \( \beta = 1 \) with \( c_\ell = 12^\frac{\ell}{4} \frac{\Gamma\left( \frac{\ell + 1}{4} \right)}{\Gamma\left( \frac{1}{4} \right)} \). The proof for \( \beta > 1 \) goes along the same lines (and is somewhat easier than the two other cases).

Lemma 5.19. For even \( \ell \geq 2 \) we have
\[ 2^\frac{\ell}{2} \frac{\Gamma\left( \frac{\ell + 1}{2} \right)}{\Gamma\left( \frac{1}{2} \right)} = (\ell - 1)!! \]  
(218)
Proof: Set $\ell = 2r$. We do induction over $r$.

For $r = 1$ we use Lemma 5.14 to prove
\[
2 \frac{\Gamma(1 + \frac{1}{2})}{\Gamma(\frac{1}{2})} = 2 \frac{1}{2} \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2})} = 1 = 0!
\] (219)

The induction step is similar:
\[
2^{r+1} \frac{\Gamma(r + 1 + \frac{1}{2})}{\Gamma(\frac{1}{2})} = 2^{r+1} \frac{1}{2} \frac{\Gamma(r + \frac{1}{2})}{\Gamma(\frac{1}{2})} = (2r + 1)(2r - 1)!! = (2r + 1)!!
\] (220)

5.5 The law of large numbers for the Curie-Weiss model

In this section we investigate the large-N-behavior of normalized sums
\[
m_N = \frac{1}{N} \sum_{i=1}^{N} X_i
\] (222)
of Curie-Weiss distributed random variables $X_i$.

In physics, where Curie-Weiss random variables serve as models for (ferro-) magnetic systems, the quantity $m_N$ gives the mean magnetization of the system. In the theory of voting systems [5] $m_N$ models the result of the voting of $N$ voters. For example, a simple majority is established if $m_N > 0$. More generally, a qualified majority, meaning the number of "Yes"-votes is bigger than $qN$, is established if $m_N > 2q - 1$.

In the following we will show that the behavior of $m_N$ (as $N \to \infty$) changes drastically at $\beta = 1$. More precisely, if $\beta \leq 1$ then
\[
m_N \xrightarrow{D} \delta_0
\] (223)

while for $\beta > 1$ we prove
\[
m_N \xrightarrow{D} \frac{1}{2}(\delta_{-m(\beta)} + \delta_{m(\beta)})
\] (224)

where $t = m(\beta)$ is the (strictly) positive solution of $t = \tanh(\beta t)$ (see Theorem 5.17.3).

In physical terms: At inverse temperature $\beta = 1$ there is a phase transition from paramagnetism to ferromagnetism.
Theorem 5.20. Suppose the random variables \( X_1, \cdots, X_N \) are \( P^{(N)}_{\beta} \)-distributed Curie-Weiss random variables and set \( m_N = \frac{1}{N} \sum_{i=1}^{N} X_i \) then

1. If \( \beta \leq 1 \) then

\[
m_N \overset{D}{\to} \delta_0 \tag{225}
\]

2. If \( \beta > 1 \) then

\[
m_N \overset{D}{\to} \frac{1}{2} (\delta_{-m(\beta)} + \delta_{m(\beta)}) \tag{226}
\]

where \( m(\beta) \) is the unique (strictly) positive solution of

\[
tanh(\beta t) = t \tag{227}
\]

Remark 5.21. If we set \( m(\beta) = 0 \) for \( \beta \leq 1 \) we obtain \( m_N \overset{D}{\to} \frac{1}{2} (\delta_{-m(\beta)} + \delta_{m(\beta)}) \) for all \( \beta \).

Proof: As usual we investigate the moments of \( m_N \):

\[
\mathbb{E}(m_N^\ell) = \frac{1}{N^\ell} \mathbb{E}\left( \left( \sum_{i=1}^{N} X_i \right)^\ell \right) = \frac{1}{N^\ell} \sum_{i_1, \cdots, i_\ell = 1}^{N} \mathbb{E}(X_{i_1} X_{i_2} \cdots X_{i_\ell}) \tag{228}
\]

We note that \( |\mathbb{E}(X_{i_1} X_{i_2} \cdots X_{i_\ell})| \leq 1 \). We split the sum \(\text{\eqref{228}}\) in two parts. First, we consider those tuples \((i_1, \cdots, i_\ell)\) for which at least one index occurs at least twice. The number of such indices is at most \( C_\ell N^{\ell-1} \) (with \( C_\ell \) independent of \( N \)).

Thus

\[
\frac{1}{N^\ell} \sum_{i_1, \cdots, i_\ell = 1}^{N} \mathbb{E}(X_{i_1} X_{i_2} \cdots X_{i_\ell}) \leq C_\ell \frac{1}{N} \to 0 \tag{229}
\]

some index occurs twice
Hence asymptotically (as $N \to \infty$)

$$E(m_N^\ell) \approx \frac{1}{N^{\ell}} \sum_{\text{no index repeated}}^N E(X_{i_1} \cdots X_{i_\ell})$$

$$= \frac{1}{N^{\ell}} \sum_{\text{no index repeated}}^N E(X_1 X_2 \cdots X_\ell)$$

$$\approx \frac{1}{N^{\ell}} N^{\ell} E(X_1 X_2 \cdots X_\ell)$$

$$\approx E(X_1 X_2 \cdots X_\ell) \quad (230)$$

the last term goes to zero if $\beta \leq 1$ by Theorem 5.17 and to $m(\beta)^\ell$ for $\beta > 1$ and even $\ell$, it equals zero for odd $\ell$. This proves the Theorem. □

Remark 5.22. All we needed for part 1 of Theorem 5.20 was that for all $\ell$ the expectations satify that $E(X_1 X_2 \cdots X_\ell) \to 0$ (together with the existence of all moments).

5.6 Central limit theorems for the Curie-Weiss model

We turn to central limit theorems for Curie-Weiss random variables. Since $\frac{1}{N} \sum_{i=1}^N X_i$ does not converge to zero if $\beta > 1$, it is clear that for this case $\frac{1}{\sqrt{N}} \sum_{i=1}^N X_i$ has no chance to converge.

Thus, we suppose $\beta \leq 1$ for the rest of this section postponing a closer look at fluctuation for $\beta > 1$ to the next section.

We will see in the following that, in deed,

$$\Sigma_N := \frac{1}{\sqrt{N}} \sum_{i=1}^N X_i \quad (231)$$

converges to a normal distribution as long as $\beta < 1$.

Although $V(X_i) = E(X_i^2) = 1$ (in fact $X_i^2 = 1$) the limit distribution of $\Sigma_N$ is not $\mathcal{N}(0,1)$ but $\mathcal{N}(0, \frac{1}{1-\beta})$. 
In fact,

\[ \mathbb{E}(\Sigma_N^2) = \frac{1}{N} \sum_{i,j=1}^{N} \mathbb{E}(X_i X_j) \]

\[ = \frac{1}{N} \left( N + \sum_{i \neq j} \mathbb{E}(X_i X_j) \right) \]

\[ \approx 1 + \frac{N(N - 1)}{N} \frac{\beta}{1 - \beta} \]

where we used \( \mathbb{E}(X_i^2) = 1 \) and \( \mathbb{E}(X_i X_j) \approx \frac{\beta}{1 - \beta} \) for \( i \neq j \) by Theorem 5.17.1. So

\[ \mathbb{V}(\Sigma_N) = \mathbb{E}(\Sigma_N^2) \approx \frac{1}{1 - \beta}. \]  

This calculation explains why to expect \( \mathcal{N}(0, \frac{1}{1 - \beta}) \) as the limit distribution. It also suggests that \( \Sigma_N \) will not converge for \( \beta = 1 \).

In fact, to make \( \mathbb{V}(\frac{1}{N^\alpha} \sum X_i) \) converge for \( \beta = 1 \), we got to choose \( \alpha = \frac{3}{4} \) as the following computation shows (we use Theorem 5.17.2).

\[ \mathbb{E}\left( \left( \frac{1}{N^\alpha} \sum X_i \right)^2 \right) = \frac{1}{N^{2\alpha}} \sum_{i,j=1}^{N} \mathbb{E}(X_i X_j) \]

\[ = \frac{1}{N^{2\alpha}} \left( N + \sum_{i \neq j} \mathbb{E}(X_i X_j) \right) \]

\[ \approx N^{1-2\alpha} + \frac{N(N - 1)}{N^{2\alpha}} \cdot \frac{\beta}{1 - \beta} \cdot \frac{1}{N^{1/2}} \]

It turns out that the fluctuations at the critical (inverse) temperature \( \beta = 1 \) are, in deed, of order \( N^{\frac{3}{4}} \) (rather than \( N^{\frac{1}{2}} \) as for \( \beta < 1 \)), i.e.

\[ \Sigma'_N = \frac{1}{N^{\frac{3}{4}}} \sum_{i=1}^{N} X_i \]

(233)

converge to a limit distribution. Moreover, the limit measure is not a normal distribution.

**Recalling notations and strategy of the proof**

We start our discussion with the case \( \beta < 1 \). It should not come as a surprise, that again we have to deal with expectations of the form

\[ \mathbb{E}(X_{i_1} \cdots X_{i_L}) \]  

(234)

85
and a substantial part of the proofs consists of careful bookkeeping.

In the following, we recall and apply notations from section 3.2.

For each multiindex $\vec{i} = (i_1, \ldots, i_L)$ the quantities $\rho_\ell(\vec{i})$ count the number of indices $i_1, \ldots, i_L$ that occur exactly $\ell$-times in $\vec{i}$.

Since Curie-Weiss random variables are exchangeable it is clear that

$$E(X_{i_1}X_{i_2}\cdots X_{i_L})$$

depends only on $\vec{r} = \rho(\vec{i}) = (\rho_1(\vec{i}), \ldots, \rho_L(\vec{i}))$.

We set

$$E(X(\vec{r})) = E(X_{i_1}X_{i_2}\cdots X_{i_L})$$

if $\vec{r} = \rho(\vec{i})$.

Therefore, we can write

$$E\left(\left(\frac{1}{N^{1/2}} \sum_{i=1}^{N} X_i\right)^L\right) = \frac{1}{N^{L/2}} \sum_{\vec{r} \in \Pi} w_L(\vec{r}) E(X(\vec{r}))$$

(237)

Here, as in section 3.2, $\Pi$ denotes the set of all $L$-profiles and $w_L(\vec{r})$ counts the number of multi-indices $\vec{i}$ with $\rho(\vec{i}) = \vec{r}$ (see Definitions 3.9 and 3.10 for details).

From these definitions we also recall that $\Pi^0_k$ denotes the set of profiles $\vec{r}$ with $r_1 = k$ and $r_\ell = 0$ for all $\ell > 2$. Also, $\Pi^+_k$ denotes the set of profiles $\vec{r}$ with $r_1 = k$ and $r_\ell > 0$ for some $\ell > 2$.

The strategy of our proof follows the one for independent, identically distributed random variables in Section 3.3. We split the sum in (237) into the sums over the sets $\Pi^0_k$ with $k = 0, 1, \ldots, L$ and the sets $\Pi^+_k$.

We saw already in Section 3.3 (cf. Remark 3.21) that the sum over $\Pi^+_k$ is negligible in the limit. Independence was not used in the corresponding estimate.

Again as in the independent case, the contribution from the set $\Pi^0_0$ is easily seen to be 1.

For independent random variables (with expectation 0) we have $E(X(\vec{r})) = 0$ if $r_1 > 0$, hence the sets $\Pi^0_k$ and $\Pi^+_k$ for $k > 0$ do not contribute at all. This is different in the current context. It turns out that the sets $\Pi^+_k$ do not contribute to the limit, but the sum over the sets $\Pi^0_k$ does. It is here where the correlation estimates of Theorem 5.17 will play a crucial role.
Results and proofs

Theorem 5.23. Let \( X_1, \ldots, X_N \) be \( \mathbb{P}_\beta^N \)-distributed Curie-Weiss random variable with \( \beta < 1 \) then

\[
\sum_{N=1}^N \frac{1}{\sqrt{N}} \sum_{i=1}^N X_i \xrightarrow{D} \mathcal{N}(0, \frac{1}{1-\beta}) \tag{238}
\]

Proof: As above we write:

\[
E\left(\left( \frac{1}{N^{1/2}} \sum_{i=1}^N X_i \right)^L \right) = \frac{1}{N^{L/2}} \sum_{\mathbf{r} \in \Pi} w_L(\mathbf{r}) E(X(\mathbf{r}))
\]

\[
= \frac{1}{N^{L/2}} \sum_{k=0}^L \sum_{\mathbf{r} \in \Pi_k} w_L(\mathbf{r}) E(X(\mathbf{r})) \tag{239}
\]

\[
+ \frac{1}{N^{L/2}} \sum_{k=0}^L \sum_{\mathbf{r} \in \Pi_k^+} w_L(\mathbf{r}) E(X(\mathbf{r})) \tag{240}
\]

To discuss the expressions \( w_L(\mathbf{r}) E(X(\mathbf{r})) \) we start with the case \( \mathbf{r} \in \Pi_k^+ \):

Lemma 5.24. Let \( X_1, \ldots, X_N \) be \( \mathbb{P}_\beta^N \)-distributed Curie-Weiss random variable with \( \beta < 1 \). Suppose the \( L \)-profile \( \mathbf{r} \) belongs to \( \Pi_k^+ \). Then

1. There is a constant \( C_1 \) independent of \( N \) such that

\[
\left| \mathbb{E}(X(\mathbf{r})) \right| \leq C_1 \frac{1}{N^{k/2}}. \tag{241}
\]

2. The number \( w_L(\mathbf{r}) \) of index tuples \( \mathbf{i} \) with profile \( \mathbf{r} \) satisfies

\[
w_L(\mathbf{r}) \leq L! N^{(L+k-1)/2} \tag{242}
\]

Proof (Lemma 5.24): We have

\[
X_i^\ell = \begin{cases} 
1, & \text{if } \ell \text{ is even;} \\
X_i, & \text{if } \ell \text{ is odd.}
\end{cases}
\]

Thus, if \( \mathbf{r} \in \Pi_k^+ \), we obtain

\[
\mathbb{E}(X(\mathbf{r})) = \mathbb{E}(X_1 X_2 \ldots X_i)
\]

for some \( \ell \geq k. \)
From this Theorem \[\text{5.17}\] gives the claim 1.

From Corollary \[\text{3.18}\] we learn

\[
\begin{align*}
    w_L(r) &\leq L! \sum_{\ell=1}^{L} r_\ell \\
    &\leq L! N^{k+\frac{1}{2}} \sum_{\ell=2}^{L} \ell r_\ell^{-\frac{1}{2}} \\
    &\leq L! N^{\frac{k}{2}+\frac{1}{2}} \sum_{\ell=1}^{L} \ell r_\ell^{-\frac{1}{2}} \\
    &\leq L! N^{\frac{1}{2}(L+k)-\frac{1}{2}}
\end{align*}
\]

(we used that \(r_\ell > 0\) for some \(\ell > 2\))

\[\square\]

Continuing the proof of Theorem \[\text{5.23}\] we observe that for \(r \in \Pi^+_k\)

\[
    w_L(r) \mathbb{E}(X(r)) \leq L! N^{\frac{1}{2}(L+k)-\frac{1}{2}} C_1 \frac{1}{N^{k/2}} \leq C_2 N^{\frac{k}{2}-\frac{1}{2}}
\]

Thus

\[
\frac{1}{NL^2} \sum_{k=0}^{L} \sum_{r \in \Pi^+_k} w_L(r) \mathbb{E}(X(r))
\]

converges to zero as \(N \to \infty\) (and \(L\) is fixed).

We turn to the case \(r \in \Pi^0_k\). Since \(r_\ell = 0\) for \(\ell \geq 3\) we have \(k + 2r_2 = L\). Thus either both \(k\) and \(L\) are odd or both are even. If \(k\) is odd then

\[
\mathbb{E}(X(r)) = \mathbb{E}(X_1 X_2 \ldots X_k) = 0
\]

(243)

by Theorem \[\text{5.17}\]. Hence we may suppose that both \(k\) and \(L\) are even.

**Lemma 5.25.** Let \(X_1, \cdots, X_N\) be \(\mathbb{P}_\beta^N\)-distributed Curie-Weiss random variable with \(\beta < 1\). Let \(k\) and \(L\) be even. Suppose the \(L\)– profile \(r\) belongs to \(\Pi^0_k\). Then

1. For large \(N\)

\[
\mathbb{E}(X(r)) \approx (k - 1)!! \left(\frac{\beta}{1 - \beta}\right)^{k/2} N^{-k/2}. \tag{244}
\]

2. The number \(w_L(r)\) of index tuples \(i\) with profile \(r\) satisfies

\[
w_L(r) \approx \frac{L!}{k! (\frac{L-k}{2})!} 2^{(L-k)/2} N^{(L+k)/2} \tag{245}\]
Proof (Lemma 5.25): Since for \( r \in \Pi^0_0 \) we have \( E(X(r)) = E(X_1 X_2 \ldots X_k) \) assertion 1 follows immediately from Theorem 5.17.

Assertion 2 is a simple consequence of Theorem 3.14.

Finally, we collect the various observations and get for even \( L \)

\[
E \left( \left( \frac{1}{N^{1/2}} \sum_{i=1}^{N} X_i \right)^L \right)
\approx \frac{1}{N^{L/2}} \sum_{\ell=0}^{L/2} \sum_{r \in \Pi^0_\ell} w_L(r) \, E(X(r))
\approx \frac{1}{N^{L/2}} \sum_{\ell=0}^{L/2} \frac{L!}{(2\ell)! (L/2 - \ell)! 2^{L/2-\ell}} N^{L/2+\ell} (2\ell - 1)!! \left( \frac{\beta}{1-\beta} \right)^\ell N^{-\ell}
\]

\[
= \sum_{\ell=0}^{L/2} \frac{L!}{(2\ell)! (L/2 - \ell)! 2^{L/2-\ell}} (2\ell - 1)!! \left( \frac{\beta}{1-\beta} \right)^\ell
\]

we use Lemma 2.46 to express \((2\ell - 1)!!:\)

\[
= \frac{L!}{(L/2)! 2^{L/2}} \sum_{\ell=0}^{L/2} \frac{(L/2)!}{(L/2 - \ell)! \ell!} \left( \frac{\beta}{1-\beta} \right)^\ell
\]

\[
= (L - 1)!! \left( \frac{2}{1-\beta} \right)^{L/2}
\]  

(246)

Thus we have proved that the moments of \( \Sigma_N \) converge to the moments of \( \mathcal{N}(0, \frac{1}{1-\beta}) \).

This completes the proof of Theorem 5.23.

We turn to the case of the ‘critical’ inverse temperature \( \beta = 1 \).

**Theorem 5.26.** Let \( X_1, \ldots, X_N \) be \( \mathbb{P}_0^N \)-distributed Curie-Weiss random variable with \( \beta = 1 \) and denote by \( \mu_1 \) the measure on \( \mathbb{R} \) with density

\[
\phi_1(x) = 2 \left( \frac{1}{12} \right)^{-1/4} \Gamma(1/4)^{-1} e^{-\frac{1}{12} x^4}
\]

then

\[
\Sigma_N' = \frac{1}{N^{3/4}} \sum_{i=1}^{N} X_i \xrightarrow{D} \mu_1.
\]  

(247)
Remark 5.27. The measure $\mu_1$ is indeed a probability measure and its moments $m_k(\mu_1)$ are given by

\[
m_k(\mu_1) = \begin{cases} 
12^k \frac{\Gamma\left(\frac{k+1}{4}\right)}{\Gamma\left(\frac{k}{4}\right)}, & \text{for even } k; \\
0, & \text{for odd } k.
\end{cases}
\] (248)

Proof (Remark): For even $k$ we have

\[
\int_{-\infty}^{+\infty} x^k e^{-\frac{1}{12}x^4} \, dx = \frac{1}{2} (12)^{\frac{k+1}{4}} \int_0^{\infty} t^{\frac{k-3}{4}} e^{-t} \, dt
\]

by a change of variable $t = \frac{1}{12} x^4$.

Proof (Theorem 5.26): For $r \in \Pi_k$ we have by Theorem 5.17:

\[
\left| \mathbb{E}(X(r)) \right| \leq C_1 \frac{1}{N^{k/4}}.
\]

and

\[
w_L(r) \leq C N^{\sum_{\ell=1}^L r_\ell} \leq C N^{\frac{1}{4}(k+L)}
\]

since

\[
\sum_{\ell=1}^L r_\ell = k + \sum_{\ell=2}^L r_\ell \leq k + \frac{1}{2} \sum_{\ell=2}^L 2r_\ell
\]

\[
\leq \frac{1}{2} k + \frac{1}{2} \sum_{\ell=1}^L 2r_\ell = \frac{1}{2} (k + L)
\]

Thus

\[
\frac{1}{N^{L/4}} w_L(r) \left| \mathbb{E}(X(r)) \right| \leq C' N^{-\frac{1}{4}(L-k)}.
\]

This term goes to zero unless $k = L$.

We conclude

\[
\mathbb{E}\left(\left(\frac{1}{N^{3/4}} \sum_{i=1}^N X_i\right)^L\right)
\]

\[
= \frac{1}{N^{L/2}} \sum_{k=0}^{L-1} \sum_{r \in \Pi_k} w_L(r) \mathbb{E}(X(r)) + \frac{1}{N^{L/2}} \sum_{r \in \Pi_L} w_L(r) \mathbb{E}(X(r))
\]

\[
\approx \frac{1}{N^{L/2}} \sum_{r \in \Pi_L} w_L(r) \mathbb{E}(X(r))
\]

90
For \( r \in \Pi_L \) we have
\[
\omega_L(r) \approx N^k \quad \text{(by Theorem 3.14)}
\]
and
\[
\mathbb{E}(X(r)) \approx \begin{cases} 
12\frac{L+1}{L}\frac{\Gamma\left(\frac{L+1}{4}\right)}{\Gamma\left(\frac{L}{4}\right)} \frac{1}{N^{5/4}}, & \text{for even } L; \\
0, & \text{for odd } L.
\end{cases} \quad \text{(by Theorem 5.17)}
\]
From this the theorem follows.

5.7 **Fluctuation of the magnetization for large \( \beta \)**

The distribution of \( \frac{1}{N} \sum X_i \) converges to the measure \( \frac{1}{2}(\delta_{-m(\beta)} + \delta_{m(\beta)}) \). \( m = m(\beta) \) is the biggest solution to
\[
\tanh(\beta m) = m
\]
\( m(\beta) > 0 \) if \( \beta > 1 \) (\( m(\beta) = 0 \) if \( \beta \leq 1 \)).

Consequently for \( \beta > 1 \), the random variables
\[
\frac{1}{\sqrt{N}} \sum (X_i - c)
\]
cannot converge no matter how we choose the constant \( c \).

Nevertheless, it would be interesting to know how the mean magnetization \( \frac{1}{N} \sum X_i \) fluctuates around its "limit points" \( \pm m(\beta) \). To formalize this we consider the distribution of \( \Sigma_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i - m(\beta)) \) under the condition that \( \frac{1}{N} \sum_{i=1}^{N} X_i > 0 \).

This conditional probability is defined as
\[
P_{\beta} \left( \Sigma_N \in A \mid \frac{1}{N} \sum_{i=1}^{N} X_i > 0 \right) = \frac{P_{\beta} \left( \Sigma_N \in A \wedge \frac{1}{N} \sum_{i=1}^{N} X_i > 0 \right)}{P_{\beta} \left( \frac{1}{N} \sum_{i=1}^{N} X_i > 0 \right)}
\]
We know that \( P_{\beta} \left( \frac{1}{N} \sum_{i=1}^{N} X_i > 0 \right) \) is strictly positive and, in fact, converges to \( \frac{1}{2} \).

We remark that we could as well consider
\[
P_{\beta} \left( \Sigma_N \in A \mid \frac{1}{N} \sum_{i=1}^{N} X_i - m(\beta) < \alpha \right)
\]
\[
91
\]
which gives the same result as the choice above as long as $0 < \alpha < 2m(\beta)$.

It will be sufficient to investigate the convergence behaviour of the random variable

$$
\Sigma_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i - m(\beta)) \chi_{\left\{ \frac{1}{N} \sum_{i=1}^{N} X_i > 0 \right\}}
$$

(253)

To shorten notation we set

$$
P_{\Sigma_N}^+(A) = P_{\beta}\left( \Sigma_N \in A \mid \frac{1}{N} \sum_{i=1}^{N} X_i > 0 \right)
$$

(254)

and

$$
\chi_N^+ = \chi_{\left\{ \frac{1}{N} \sum_{i=1}^{N} X_i > 0 \right\}}.
$$

(255)

**Theorem 5.28.** Suppose $\beta > 1$ and $m = m(\beta)$ is the unique positive solution of $\tanh(\beta m) = m$. Then the distribution $P_{\Sigma_N}^+$ of $\Sigma_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i - m(\beta))$ conditioned on $\frac{1}{N} \sum_{i=1}^{N} X_i > 0$ converges to a normal distribution $N(0, \sigma^2)$ with

$$
\sigma^2 = \frac{1 - m(\beta)}{1 - \beta(1 - m(\beta))}
$$

(256)

**Proof:** We set

$$
\zeta_N(L) = \int_{-1}^{1} \frac{e^{-\frac{N}{2} F_\beta(t)}}{1 - t^2} E_t\left( \sum_{N}^L \chi_N^+ \right) dt
$$

(257)

where as usual $F_\beta(t) = \frac{1}{\beta} \left( \frac{1}{2} \ln\left( \frac{1-t}{1+t} \right) \right)^2 + \ln(1 - t^2)$ since

$$
\mathbb{E}_\beta\left( \Sigma_N^L \mid \frac{1}{N} \sum_{i=1}^{N} X_i > 0 \right) = \frac{\zeta_N(L)}{\zeta_N(0)}
$$

(258)

we have to analyze the large-$N$-behaviour of $\zeta_N(L)$. For $\beta > 1$ the function $F_\beta$ has two minima $\pm m(\beta)$ with $F_\beta(-m(\beta)) = F_\beta(m(\beta))$.

From Theorem [5.10] we expect that the asymptotics of (257) is dominated by the behavior of the integrand near $t = \pm m(\beta)$, in fact, we will see that $t = -m(\beta)$ does not play a role due to the condition $\frac{1}{N} \sum_{i=1}^{N} X_i > 0$. To simplify notation we will write $m$ instead of $m(\beta)$. We will also assume that $F_\beta(m) = F_\beta(-m) = 0.$

92
This can be achieved by multiplying numerator and denominator of (258) by the same constant, thus not changing the value of the quotient.

To evaluate (257) we cannot use Theorem (257) directly since we have no explicit expression for \( E_t\left(\Sigma N^L\chi_N^\pm\right) \).

Our strategy will be to evaluate

\[
\xi_N(L) = \int_{\frac{m}{2}}^{1} e^{-\frac{N^2}{2} F_0(t)} \frac{1}{1-t^2} E_t\left(\Sigma N^L\right) dt
\]

instead and prove that the difference to \( \zeta_N(L) \) is asymptotically negligible.

First, we decompose \( \zeta_N(L) \) in three parts

\[
\zeta_N(L) = \int_{-\frac{m}{2}}^{-1} e^{-\frac{N^2}{2} F_0(t)} \frac{1}{1-t^2} E_t\left(\Sigma N^L\chi_N^\pm\right) dt
\]

\[
+ \int_{\frac{m}{2}}^{-\frac{m}{2}} e^{-\frac{N^2}{2} F_0(t)} \frac{1}{1-t^2} E_t\left(\Sigma N^L\chi_N^\pm\right) dt
\]

\[
+ \int_{\frac{m}{2}}^{1} e^{-\frac{N^2}{2} F_0(t)} \frac{1}{1-t^2} E_t\left(\Sigma N^L\chi_N^\pm\right) dt
\]

Now, we prove that (260) and (260) do not contribute asymptotically.

\[\square\]

**Lemma 5.29.** For each \( M \in \mathbb{N} \) there is a constant \( C_M < \infty \) such that

\[
\tilde{\zeta}_N(L) = \int_{-1}^{-\frac{m}{2}} e^{-\frac{N^2}{2} F_0(t)} \frac{1}{1-t^2} E_t\left(\Sigma N^L\chi_N^\pm\right) dt
\]

\[
\leq \frac{C_M}{N^M}
\]

**Proof:** (Lemma):

By the Cauchy-Schwarz inequality

\[
E_t\left(\Sigma N^L\chi_{\{x_i > 0\}}\right) \leq E_t\left(\Sigma N^{2L}\right)^{\frac{1}{2}} P_t\left(\frac{1}{N} \sum X_i > 0\right)^{\frac{1}{2}}
\]
Due to Theorem 3.24 we have

$$P_t\left(\frac{1}{N} \sum X_i > 0\right) \leq \frac{C_M}{N^{2M}}$$

(266)

for all $t \in ]-1, \frac{1}{2}m[$.

Thus

$$\zeta_N(L) \leq \frac{C}{NM} \int_{-1}^{0} \frac{e^{-\frac{N}{2} F_\beta(t)}}{1-t^2} E_t\left(\Sigma_N^{2L}\right) \, dt$$

(267)

$$\leq \frac{C^l}{N^{2M}}$$

(268)

We used that the moments $E_t(\Sigma_N^{2L})$ are bounded (in $t$ and in $N$) (see Corollary 3.27).

Next we consider the ”portion” of $\zeta_N$ between $-\frac{\beta m}{2}$ and $\frac{\beta m}{2}$.

**Lemma 5.30.** There is a constant $C_L$ and a $\delta > 0$ such that

$$\zeta_N^0(L) = \int_{-\frac{\beta m}{2}}^{\frac{\beta m}{2}} \frac{e^{-\frac{N}{2} F_\beta(t)}}{1-t^2} E_t\left(\Sigma_N^{2L}\chi_N^+\right) \, dt$$

(269)

satisfies

$$|\zeta_N^0(L)| \leq C_L e^{-N\delta}$$

(270)

**Proof:** Since $F_\beta(t)$ has unique minima at $\pm \beta m$ there exists a $\delta > 0$ such that $F_\beta(t) > 2\delta$ for all $t \in [-\frac{\beta m}{2}, \frac{\beta m}{2}]$.

Consequently

$$\zeta_N^0(L) \leq e^{-N\delta} \int_{-\frac{\beta m}{2}}^{\frac{\beta m}{2}} \frac{1}{1-t^2} \, dt \sup_{i,N} E_t\left(\Sigma_N^{2L}\right)^\frac{1}{2}$$

(271)

It remains to estimate

$$\zeta_N^+(L) = \int_{\frac{\beta m}{2}}^{1} \frac{e^{-\frac{N}{2} F_\beta(t)}}{1-t^2} E_t\left(\Sigma_N^{L}\chi_N^+\right) \, dt$$

(272)

94
We write
\[
\zeta_N^+(L) = \int_{m(\beta)}^1 \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2} E_t\left(\Sigma_N^L\right) \, dt
\]  
\[273\]
\[
- \int_{m(\beta)}^1 \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2} E_t\left(\Sigma_N^L(1 - \chi_N^+)\right) \, dt
\]  
\[274\]

The final expression can be estimated in a similar way as \(\zeta_N(L)\), since
\[
E_t(1 - \chi_N^+) = P_t\left(\frac{1}{N} \sum_{i=1}^N X_i \leq 0\right)
\]  
\[275\]

which is very small for \(t \geq \frac{m(\beta)}{2}\).

Thus it remains to estimate
\[
\zeta_N^{++}(L) = \int_{m(\beta)}^1 \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2} E_t\left(\Sigma_N^L\right) \, dt
\]  
\[276\]

We have
\[
E_t\left(\Sigma_N^L\right) = \frac{1}{N^{\frac{L}{2}}} E_t\left(\left(\sum_{i=1}^N (X_i - m(\beta))\right)^L\right)
\]  
\[277\]

setting \(Y_i = X_i - m(\beta)\) we see that
\[
E_t(Y_i \cdots Y_{i_L}) = E_t(Y_{j_1} \cdots Y_{j_L})
\]  
\[278\]

whenever \(\underline{r} = \rho(\underline{i}) = \rho(\underline{j})\) hence the above expressions depend only on \(r = \rho(\underline{i})\) so we write
\[
E_t\left(Y(\underline{r})\right) = E_t(Y_{i_1} \cdots Y_{i_L})
\]  
\[279\]

for \(\underline{r} = \rho(\underline{i})\).

Now, we can expand
\[
E_t\left(\Sigma_N^L\right) = \frac{1}{N^{\frac{L}{2}}} \sum_{\underline{r} \in \Pi} w_L(\underline{r}) E_t\left(Y(\underline{r})\right)
\]  
\[280\]

For \(k = r_1\) we may write
\[
E_t(Y(\underline{r})) = E_t(Y_{i_1})^k E_t(Y(\underline{r}'))
\]  
\[281\]

with \(\underline{r}' = (0, r_2, r_3, \cdots, r_L)\).
Lemma 5.31. For \( r \in \Pi_k \) we have for even \( k \)

\[
\left| \int_{\frac{m}{2}}^{1} \frac{e^{-\frac{N}{2}F_{\beta}(t)}}{1 - t^2} \ E_t(Y(r)) \ dt \right| \leq \frac{C_L}{N^\frac{k}{2}}
\]  

(282)

and for odd \( k \)

\[
N^\frac{k}{2} \int_{\frac{m}{2}}^{1} \frac{e^{-\frac{N}{2}F_{\beta}(t)}}{1 - t^2} \ E_t(Y(r)) \ dt \to 0
\]  

(283)

Proof: The Lemma follows from Corollary 5.12. \( \square \)

So far we proved that

\[
\zeta_N(L) \approx \zeta_N^+(L) = \int_{\frac{m}{2}}^{1} \frac{e^{-\frac{N}{2}F_{\beta}(t)}}{1 - t^2} \ E_t\left(\Sigma_N^L\right) \ dt
\]  

(284)

\[
= \frac{1}{N^\frac{k}{2}} \sum_{r \in \Pi} w_L(r) \int_{\frac{m}{2}}^{1} \frac{e^{-\frac{N}{2}F_{\beta}(t)}}{1 - t^2} \ E_t(X(r)) \ dt
\]  

(285)

We set \( \mathcal{E}(r) = \int_{\frac{m}{2}}^{1} \frac{e^{-\frac{N}{2}F_{\beta}(t)}}{1 - t^2} \ E_t(X(r)) \ dt \). If \( r \in \Pi_k^+ \) we have by Lemma 5.31

\[
|\mathcal{E}(r)| \leq CN^{-\frac{k}{2}}
\]  

(286)

and by Corollary 3.18 (see also (93))

\[
w_L(r) \leq CN^{\frac{k+L}{2} - \frac{1}{2}}
\]  

(287)

Consequently, the terms \( \frac{1}{N^\frac{k}{2}} w_L(r) \mathcal{E}(r) \) go to zero for \( r \in \Pi_k^+ \).

Thus we have

\[
\zeta_N(L) \approx \frac{1}{N^\frac{k}{2}} \sum_{k=0}^{L} \sum_{r \in \Pi_k^0} w_L(r) \mathcal{E}(r)
\]  

(288)

For \( r \in \Pi_k^0 \) we know \( k + 2r_2 = L \), so either \( k \) and \( L \) are both odd or they are both even. If \( k \) is odd \( N^\frac{k}{2} \mathcal{E}(r) \to 0 \), so \( \zeta_N(L) \to 0 \) if \( L \) is odd.

Finally, we arrive at

\[
\zeta_N(2M) \approx \frac{1}{N^M} \sum_{k=0}^{M} \sum_{r \in \Pi_{2k}^0} w_N(r) \mathcal{E}(r)
\]  

(289)
Lemma 5.32. For $r \in \Pi_{2k}^0$ with $L = 2M$

$$\frac{\mathcal{E}(r)}{\zeta_N(0)} \approx (1 - m^2)^M \left( \frac{\beta(1 - m^2)}{1 - \beta(1 - m^2)} \right)^k (2k - 1)!! \frac{1}{N^k} \quad (290)$$

**Proof:** The function $F_\beta$ has a unique minimum in $[\frac{m}{2}, 1]$, namely $t = m$ (see the computation in Proposition 5.16).

We computed

$$F_\beta''(m) = \frac{2}{1 - m^2} \frac{1 - \beta(1 - m^2)}{\beta(1 - m^2)} \quad (> 0) \quad (291)$$

To apply Corollary 5.12 we compute for $r \in \Pi_{2k}^0$, so

$$r = (2k, M - k, 0, \ldots, 0) \quad (292)$$

$$E_t(X(r)) = E_t(X_1)^{2k} E_t(X_2)^{M - k} \quad (293)$$

$$= (t - m)^{2k} (1 - 2mt + m^2)^{M - k} \quad (294)$$

Applying Corollary 5.12 we get

$$\mathcal{E}(r) \approx (1 - m^2)^{M - k} (1 - m^2) \left( \frac{\beta(1 - m^2)}{1 - \beta(1 - m^2)} \right)^{2k + 1} 2^{2k + 1} \Gamma \left( \frac{2k + 1}{2} \right) \cdot \frac{1}{N^{2k + 1}} \quad (295)$$

Since

$$\zeta_N(0) = \mathcal{E}(0) \quad (0 \text{ is the unique element in } \Pi \text{ for } L = 0) \quad (296)$$

we obtain

$$\frac{\mathcal{E}(r)}{\zeta_N(0)} \approx (1 - m^2)^M \left( \frac{\beta(1 - m^2)}{1 - \beta(1 - m^2)} \right)^k \frac{2^k \Gamma \left( \frac{2k + 1}{2} \right) \cdot \frac{1}{N^k}}{\Gamma \left( \frac{k}{2} \right)} \quad (297)$$

$$= (1 - m^2)^M \left( \frac{\beta(1 - m^2)}{1 - \beta(1 - m^2)} \right)^k (2k - 1)!! \cdot \frac{1}{N^k} \quad (298)$$

where we used Lemma 5.19. ∎
To finish the proof of Theorem 5.28 we sum over $k$:

\[
\frac{1}{N^M} \sum_{k=0}^{M} \sum_{r \in \Pi_{2k}^0} w_{2M}(r) \frac{\mathcal{E}(r)}{\mathcal{E}(0)} \quad \approx \\
\frac{1}{N^M} \sum_{k=0}^{M} N^{M+k} \frac{(2M)!}{(2k)!(M-k)!} 2^{M-k} (1 - m^2)^M \cdot \left( \frac{\beta(1 - m^2)}{1 - \beta(1 - m^2)} \right)^k \cdot (2k - 1)!! \cdot \frac{1}{N^k} \quad \approx \\
(1 - m^2)^M \sum_{k=0}^{M} \frac{M!}{k!(M-k)!} \frac{(2M)!}{M!} \frac{k!}{(2k)!} 2^{M-k} \cdot (2k - 1)!! \cdot \left( \frac{\beta(1 - m^2)}{1 - \beta(1 - m^2)} \right)^k \\
= (1 - m^2)^M (2M - 1)!! \sum_{k=0}^{M} \binom{M}{k} \left( \frac{\beta(1 - m^2)}{1 - \beta(1 - m^2)} \right)^k \\
= (1 - m^2)^{\frac{L}{2}} (L - 1)!! \left( \frac{1}{1 - \beta(1 - m^2)} \right)^{\frac{L}{2}}
\]
Index

\begin{itemize}
  \item \(\Pi_k\) ................................................. 35
  \item \(C_0(\mathbb{R}^d)\) ........................................ 11
  \item \(C_b(\mathbb{R}^d)\) ........................................ 7
  \item \(C_\infty(\mathbb{R}^d)\) .................................... 11
  \item \(E_r\) .................................................. 55
  \item \(E_r^{(1)}\) ............................................. 54
  \item \(E_r^{(N)}\) ............................................. 54
  \item \(F_\beta(t)\) ........................................... 66
  \item \(I_\mu(f)\) .......................................... 7
  \item \(P_{X_1 \ldots X_N}\) ................................... 29
  \item \(P_X\) ................................................ 29
  \item \(P_r\) ................................................ 54
  \item \(P_r^{(1)}\) ........................................... 54
  \item \(P_r^{(N)}\) ........................................... 54
  \item \(W(\mathbf{r})\) ......................................... 34
  \item \(\text{Cov}\) ........................................... 50
  \item \(\Gamma(x)\) ........................................... 71
  \item \(\mathbb{E}(X)\) ........................................ 5
  \item \(\mathbb{E}_\mu\) ......................................... 55
  \item \(\mathbb{P}\)-almost surely ......................... 45
  \item \(\mathbb{P}(X \in A)\) ..................................... 29
  \item \(\mathbb{P}_\beta\) ....................................... 65
  \item \(\mathbb{P}_\beta^{(N)}\) .................................... 65
  \item \(\nu(\mathbf{r})\) ........................................ 22
  \item \(\Lambda L\) ........................................... 20
  \item \(\implies\) ........................................... 8
  \item \(\Phi_{R,L}(x)\) ...................................... 35
  \item \(\Pi\) ................................................. 35
  \item \(\Pi^{(L)}\) ........................................... 35
  \item \(\Pi^+\) ............................................... 35
  \item \(\Pi^0\) .............................................. 35
  \item \(\Pi_+\) .............................................. 35
  \item \(\Pi^+_k\) ............................................ 35
  \item \(\approx\) ............................................... 39
  \item \(\delta_{x_0}\) ......................................... 7
  \item \(\delta_x\) ........................................... 7
  \item \(\hat{\mu}\) .......................................... 20
  \item \(\|f\|_\infty\) ...................................... 7
  \item \(\lim \inf_n A_n\) .................................. 45
  \item \(\lim \sup_n A_n\) .................................. 45
  \item \(D_\infty\) ........................................... 19
  \item \(D_{\text{trig}}\) ...................................... 20
  \item \(\nu_\mathcal{I}^{(i)}\) .................................. 33
  \item \(B_R\) ............................................... 9
  \item \(m_k(\mu)\) ........................................ 21
  \item \(\Rightarrow\) ....................................... 11
  \item \(\phi_{R,L}\) ......................................... 8
  \item \(\rho_\mathcal{I}^{(i)}\) .................................. 33
  \item \(\implies\) ........................................... 29
  \item \(\text{supp}\) ........................................ 9
  \item \(\rho(\mathbf{i})\) ...................................... 33
  \item \(m_k(\mu)\) ........................................ 21
  \item \(m_N\) ............................................. 80
  \item \(w(\mathbf{r})\) ........................................ 34
  \item absolute moment .................................. 21
  \item Borel-Cantelli Lemma ........................... 45
  \item bounded measure .................................. 7
  \item canonical ensemble ................................ 62
  \item Central Limit Theorem ........................... 39
  \item characteristic function .......................... 20
  \item Chebyshev’s inequality .......................... 43
  \item Chebyshev-Markov inequality .................... 42
  \item convergence in distribution ..................... 29
  \item converges in distribution ....................... 29
  \item covariance ........................................ 50
  \item Curie–Weiss measure .............................. 65
  \item Curie–Weiss model ............................... 65
  \item de Finetti measure ................................ 53
  \item de Finetti Theorem ................................ 58
  \item de Finetti type ................................... 54
  \item de-Finetti sequence .............................. 57
\end{itemize}


References


https://www.ocf.berkeley.edu/~wwu/articles/fejer_theorem.pdf