A Survey on the method of moments

Werner Kirsch
Fakultät für Mathematik und Informatik
FernUniversität in Hagen
Germany

October 1, 2015

Contents

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

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

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

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

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 variable, 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 β 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 β. 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 μ on the Borel-σ-Algebra $\mathcal{B}(\mathbb{R}^d)$ with $\mu(\mathbb{R}^d) < \infty$. A *probability measure* μ on $\mathbb{R}^d$ is a bounded measure with $\mu(\mathbb{R}^d) = 1$.

Examples 2.1.

1. For any $x_0 \in \mathbb{R}^d$ $\big($and $A \in \mathcal{B}(\mathbb{R}^d)\big)$

$$\delta_{x_0}(A) = \begin{cases} 1, & \text{if } x_0 \in A; \\ 0, & \text{otherwise.} \end{cases}$$

   defines a probability measure on $\mathbb{R}^d$, the *Dirac measure* in x_0.

2. If λ 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 μ on $\mathcal{B}(\mathbb{R}^d)$. The function f is called the *Lebesgue density* of μ.

If μ is a bounded measure on $\mathbb{R}^d$ and f a bounded continuous function on $\mathbb{R}^d$, the integral

$$I_\mu(f) \ := \ \int_{\mathbb{R}^d} f(x)\, d\mu(x) \tag{1}$$

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

$$\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)| \tag{2}$$

$C_b(\mathbb{R}^d)$ is a Banach space.

Proposition 2.3. *For any bounded measure μ on $\mathbb{R}^d$ the mapping $I_\mu : C_b(\mathbb{R}^d) \to \mathbb{R}$, defined in (1), is a bounded, positive linear functional on $C_b(\mathbb{R}^d)$, i.e.*

1. *$I_\mu : C_b(\mathbb{R}^d) \to \mathbb{R}$ is a linear mapping:*
$$I_\mu(\alpha f + \beta g) = \alpha I_\mu(f) + \beta I_\mu(g) \tag{3}$$
for all $\alpha, \beta \in \mathbb{R}$, and $f, g \in C_b(\mathbb{R}^d)$.

2. *I_μ is positive: If $f \geq 0$, then $I_\mu(f) \geq 0$*

3. *I_μ is bounded: There is a constant $C < \infty$ such that for all $f \in C_b(\mathbb{R}^d)$*
$$|I_\mu(f)| \leq C \, \|f\|_\infty \ . \tag{4}$$
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 $I : C_b(\mathbb{R}^d) \to \mathbb{R}$ is a bounded, positive linear functional then there exists a bounded measure μ on $\mathbb{R}^d$, such that*
$$I(f) \ = \ I_\mu(f) \ = \ \int f(x) \, d\mu(x) \qquad \text{for all } f \in C_b(\mathbb{R}^d) \tag{5}$$

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

In the sense of Theorem 2.4 we may (and will) identity measures μ and the corresponding (bounded, positive) linear functionals I_μ.

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

We also define (for any dimension d) a function $\Phi_{R,L} : \mathbb{R}^d \to [0,1]$ by
$$\Phi_{R,L}(x) = \phi_{R,L}(\|x\|) \tag{7}$$

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

We remind the reader that the support 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 $f_y(x) = \Phi_{\frac{1}{10}, \frac{2}{10}}(x - y)$ then supp $f_i \cap$ 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 $m \in M$

$$F_m(x) = \sum_{i \in \mathbb{Z}^d} m_i f_i(x) \tag{8}$$

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

$$\|F_m - F_{m'}\|_\infty \geq \sup_{i \in \mathbb{Z}^d} \left| m_i - m_i^{'} \right| \geq 2 \tag{9}$$

Since M is uncountable this proves the claim. □

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 to strong requirement for most purposes. For example, it is quite desirable that $\delta_{\frac{1}{n}}$ should converge to δ_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 μ_n converge if applied to a Borel set, we require that the linear functionals I_{μ_n} converge, if applied to a bounded continuous function, more precisely:

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

$$\int f(x)\, d\mu_n(x) \to \int f(x)\, d\mu(x) \tag{10}$$

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

Remarks 2.10.

1. If $\mu_n \Rightarrow \mu$ then $\mu(\mathbb{R}^d) = \lim\ \mu_n(\mathbb{R}^d)$
2. $\delta_{\frac{1}{n}} \Longrightarrow \delta_0$

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

Theorem 2.11 (Portemanteau). *If μ_n and μ are bounded measures on $\mathbb{R}^d$, then $\mu_n \Longrightarrow \mu$ 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 μ_n and μ are bounded measures on $\mathbb{R}^d$, then $\mu_n \Longrightarrow \mu$ if and only if*

$$\int f(x)\, d\mu_n \ \to \ \int f(x)\, d\mu$$

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 μ_n is a sequence of bounded measures on $\mathbb{R}^d$ and that*

$$\lim_{n\to\infty} \int f(x)\, d\mu_n(x) = I(f) \tag{11}$$

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

$$\mu_n \Longrightarrow \mu \tag{12}$$

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 μ such that

$$I(f) = \int f(x)\, d\mu(x) \tag{13}$$

and $\mu_n \Longrightarrow \mu$. □

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

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

$C_0(\mathbb{R}^d)$ 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

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

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)

$$\begin{aligned} \int f(x)\, d\mu(x) &= \int (f_1(x) + i f_2(x))\, d\mu(x) \\ &= \int f_1(x)\, d\mu(x) + i \int f_2(x)\, d\mu(x)\,. \end{aligned}$$

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

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

$$\int f(x)\, d\mu_n(x) \to \int f(x)\, d\mu_n \tag{16}$$

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

Remark 2.18.

1. Vague convergence can be defined for all Borel measures μ 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*

$$\|f - f_n\|_\infty \to 0 \quad \text{as } n \to \infty \tag{17}$$

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

We define

$$D_0 = \Big\{ \Psi : \mathbb{R}^d \to \mathbb{R} \mid \Psi(x) = P(x)\,\Phi_{L,2L}(x) \\ \text{for some } P \in \mathcal{P}_\mathbb{Q} \text{ and some } L \in \mathbb{N} \Big\} \tag{18}$$

Now, take $f \in C_0(\mathbb{R}^d)$ and $L \in \mathbb{N}$ so large that $\operatorname{supp} f \subset \overline{B_L}$, hence $f = f \cdot \Phi_{L,2L}$.

By the Weierstrass approximation theorem there is a sequence $P_n \in \mathcal{P}_\mathbb{Q}$ such that

$$\sup_{x \in \overline{B_{2L}}} |f(x) - P_n(x)| \to 0 \qquad \text{as } n \to \infty \tag{19}$$

Set $f_n(x) = P_n(x)\Phi_{L,2L}$ then

$$\begin{aligned}\|f - f_n\|_\infty &= \sup_{x\in\mathbb{R}^d} |f(x) - f_n(x)| \\ &= \sup_{x\in\overline{B_{2L}}} |f(x) - P_n(x)\Phi_{L,2L}(x)| \\ &= \sup_{x\in\overline{B_{2L}}} |f(x) - P_n(x)| \to 0\end{aligned}$$

□

2.3 Tightness and Prohorov's Theorem

If μ 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

$$\mu(\complement K_\varepsilon) < \varepsilon\,. \tag{20}$$

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

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

Hence

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

Definition 2.20. A sequence μ_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

$$\mu_n(\complement K_\varepsilon) < \varepsilon \tag{21}$$

Remark 2.21.

1. We have seen above that $K_\varepsilon = K_\varepsilon^n$ with (21) exists for each n, tightness requires that K_ε can be chosen *independent* of n.

2. To prove tightness it is sufficient to check (21) for $n > N_0$. Since there are compact set K_j with $\mu_j(\complement K_j) < \varepsilon$ for $j = 1, \ldots, N_0$ we have that $\tilde{K} := K \cup \bigcup_{j=1}^{N_0} K_j$ is compact and

$$\mu_n(\complement\tilde{K}) < \varepsilon \qquad \text{for all } n\,.$$

Examples 2.22.

1. The sequence $\mu_n = \delta_n$ is not tight.
2. The sequence $\mu_n = (1 - \frac{1}{n})\delta_{\frac{1}{n}} + \frac{1}{n}\delta_n$ is tight.

There is a useful tightness criterion involving moments of measure:

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

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

Then the sequence μ_n is tight.

Proof: We estimate

$$\begin{aligned} \mu_n\Big(\complement\overline{B_R}\Big) \;&=\; \int \chi_{\|x\|>R}(x)\, d\mu_n(x) \\ &\leq\; \int \frac{\|x\|^k}{R^k}\, d\mu_n(x) \\ &=\; \frac{C}{R^k}\,. \end{aligned}$$

This implies tightness of the sequence μ_n. □

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

1. $\mu(\mathbb{R}^d) \leq \underline{\lim}\, \mu_n(\mathbb{R}^d)$
2. *If*

$$\mu(\mathbb{R}^d) \geq \overline{\lim}\, \mu_n(\mathbb{R}^d) \tag{23}$$

then μ_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 $\mu(\complement \overline{B_R}) < \frac{\varepsilon}{2}$. So

$$\begin{aligned}\mu(\mathbb{R}^n) &\le \mu(\overline{B_R}) + \frac{\varepsilon}{2} \\ &\le \int \Phi_{R,R+1}(x)\, d\mu(x) + \frac{\varepsilon}{2} \\ &\le \int \Phi_{R,R+1}(x)\, d\mu_n(x) + \varepsilon \quad \text{for } n \text{ large enough (by (16))} \\ &\le \mu_n(\mathbb{R}^d) + \varepsilon\end{aligned}$$

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

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

$$\begin{aligned}\mu_n(\complement \overline{B_{R+1}}) &= \mu_n(\mathbb{R}^d) - \mu_n(\overline{B_{R+1}}) \\ &\le \mu_n(\mathbb{R}^d) - \int \Phi_{R,R+1}(x)\, d\mu_n(x)\end{aligned}$$

using assumption (23) and the vague convergence of the μ_n to μ we get

$$\begin{aligned}&\le \mu(\mathbb{R}^d) - \int \Phi_{R,R+1}(x)\, d\mu(x) + \frac{\varepsilon}{2} \quad \text{for } n \text{ large enough} \\ &\le \mu(\complement \overline{B_R}) + \frac{\varepsilon}{2} < \varepsilon\end{aligned}$$

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

□

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

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

$$\mu(\complement \overline{B_R}) < \frac{\varepsilon}{3\,\|f\|_\infty} \quad \text{and} \tag{24}$$

$$\mu_n(\complement \overline{B_R}) < \frac{\varepsilon}{3\,\|f\|_\infty} \quad \text{for all } n \tag{25}$$

Then

$$
\begin{aligned}
&\left| \int f(x)\, d\mu(x) - \int f(x)\, d\mu_n(x) \right| \\
= &\Big| \int f(x) \Big(\Phi_{R,R+1}(x) \,+\, \big(1 - \Phi_{R,R+1}(x)\big)\Big)\, d\mu(x) \\
&- \int f(x) \Big(\Phi_{R,R+1}(x) \,+\, \big(1 - \Phi_{R,R+1}(x)\big)\Big)\, d\mu_n(x) \Big| \\
\leq &\left| \int f(x)\Phi_{R,R+1}(x)\, d\mu(x) - \int f(x)\Phi_{R,R+1}(x)\, d\mu_n(x) \right| \\
&+ \|f\|_\infty\, \mu_n(\complement \overline{B_R}) + \|f\|_\infty\, \mu(\complement \overline{B_R}) \\
\leq &\left| \int f(x)\Phi_{R,R+1}(x)\, d\mu(x) - \int f(x)\Phi_{R,R+1}(x)\, d\mu_n(x) \right| + \frac{2}{3}\varepsilon \\
< &\,\varepsilon
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
|I_n(f) - I_k(f)| \;=\; &\left| \int f(x)\, d\mu_n(x) - \int f(x)\, d\mu_k(x) \right| \\
\leq\; &\|f\|_\infty\, \mu_n(\complement \overline{B_L}) + \|f\|_\infty\, \mu_k(\complement \overline{B_L}) \\
+ &\left| \int f(x)\Phi_{L,2L}(x)\, d\mu_n(x) - \int f(x)\Phi_{L,2L}(x)\, d\mu_k(x) \right|
\end{aligned}
$$

Since $\{\mu_n\}$ is tight we may take L so large, that $\mu_j(\complement\overline{B_L}) < \frac{\varepsilon}{3\|f\|_\infty}$ 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 $f \in C_b(\mathbb{R}^d)$ 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,

$$\begin{aligned} |I_n(f)| &\leq \mu_n(\complement B_L)\,\|f\|_\infty + \left|\int f(x)\,\Phi_{L,L+1}(x)\, d\mu_n(x)\right| \\ &\leq \varepsilon\,\|f\|_\infty + \int f(x)\,\Phi_{L,L+1}(x)\, d\mu_n(x)\,\|f\|_\infty \\ &\leq \left(\int f(x)\,\Phi_{L,L+1}(x)\, d\mu_n(x) + \varepsilon\right)\,\|f\|_\infty \\ &\leq C\,\|f\|_\infty \end{aligned}$$

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 μ 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 μ_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 μ_{n_k} such that for any $f \in C_0(\mathbb{R}^d)$

$$I_k(f) = \int f(x)\, d\mu_{n_k}(x) \tag{26}$$

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 μ_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

$$\tilde{\mu}_n = \mu_n^{(n)} \tag{27}$$

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

$$\begin{aligned}
& \left| \int f(x)\, d\tilde{\mu}_n(x) - \int f(x)\, d\tilde{\mu}_m(x) \right| \\
\leq & \left| \int f(x)\, d\tilde{\mu}_n(x) - \int f_k(x)\, d\tilde{\mu}_n(x) \right| \\
& + \left| \int f_k(x)\, d\tilde{\mu}_n(x) - \int f_k(x)\, d\tilde{\mu}_m(x) \right| \\
& + \left| \int f_k(x)\, d\tilde{\mu}_m(x) - \int f(x)\, d\tilde{\mu}_m(x) \right| \\
\leq & \|f - f_k\|_\infty \left(\tilde{\mu}_n(\mathbb{R}^d) + \tilde{\mu}_m(\mathbb{R}^d) \right) \\
& + \left| \int f_k(x)\, d\tilde{\mu}_n(x) - \int f_k(x)\, d\tilde{\mu}_m(x) \right| \\
\leq & 2\,C\ \|f - f_k\|_\infty + \left| \int f_k(x)\, d\tilde{\mu}_n(x) - \int f_k(x)\, d\tilde{\mu}_m(x) \right|
\end{aligned}$$

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 $\left| \int f_k\, d\tilde{\mu}_n - \int f_k\, d\tilde{\mu}_m \right| < \frac{\varepsilon}{2}$. □

2.4 Separating classes of functions

From Theorem 2.27 we know that *for tight sequences* μ_n vague convergence and weak convergence are equivalent. In other words: If we know that μ_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 $C_0^\infty(\mathbb{R}^d)$ of infinitely differentiable functions with compact support is separating.
4. The set 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 μ_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 μ_n converges weakly to some bounded measure μ.

Proof: By Prohorov’s Theorem 2.29 we know that any subsequence of μ_n has a weakly convergent subsequence. Let ν_k and ρ_k be two weakly convergent subsequences of μ_n and suppose $\nu_k \Rightarrow \nu$ and $\rho_n \Rightarrow \rho$.

Take $f \in \mathcal{D}$. By assumption

$$\int f(x)\,d\mu_n(x) \ \to a \qquad \text{for some } a \in \mathbb{R}$$

and consequently for the susequences ν_k and ρ_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 $f \in \mathcal{D}$. Thus $\nu = \rho$.

Using the elementary lemma below we conclude that μ_n converges weakly. □

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 $\mathcal{D}_{trig} = \{f_p \in C_b(\mathbb{R}^d) \mid f_p(x) = e^{i\,p\cdot x}, p \in \mathbb{R}^d\}$ 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 μ we define the *characteristic function* $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ (Fourier transform) of μ by

$$\hat{\mu}(p) = \int e^{i\,p\cdot x}\; d\mu(x) \tag{30}$$

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 $\Lambda_L = \{x \in \mathbb{R}^d \mid |x_i| \leq L \text{ for } i = 1, \cdots, d\}$. A *trigonometric polynomial* on Λ_L is a (finite) linear combination of functions $e^{i\frac{\pi}{L}\, n\cdot x}$ with $n \in \mathbb{Z}^d$.

We call a continuous function $f \in C(\Lambda_L)$ *periodic* if

$$f(x_1, \ldots, x_{r-1}, L, x_{r+1}, \ldots, x_d) = f(x_1, \ldots, x_{r-1}, -L, x_{r+1}, \ldots, x_d)$$

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. *(Fejér) A continuous periodic function f on Λ_L can be uniformly approximated by trigonometric polynomials on Λ_L, i.e. there exists a sequence f_n of trigonometric polynomials on Λ_L, such that*

$$\sup_{x \in \Lambda_L} |f(x) - f_n(x)| \to 0 \qquad \text{as } n \to \infty \tag{31}$$

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

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

Without loss we may assume that

$$\operatorname{supp} f \subset \Lambda_{\frac{L}{2}} \tag{32}$$

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

Since T is periodic we also have $\|T\|_\infty \leq \|f\|_\infty + 1$. Hence

$$\begin{aligned}
\left| \int f(x)\, d\mu(x) - \int T(x)\, d\mu(x) \right| &\leq \int_{\Lambda_L} |f(x) - T(x)|\; d\mu(x) + \int_{\complement\Lambda_L} |T(x)|\; d\mu(x) \\
&\leq \mu(\mathbb{R}^d) \frac{\varepsilon}{4\mu(\mathbb{R}^d)} + \|T\|_\infty \frac{\varepsilon}{4(\|f\|_\infty + 1)} \\
&< \varepsilon
\end{aligned}$$

We conclude: Knowing $\int f(x)\, d\mu(x)$ for trigonometric polynomials f, allows us to compute the integral for $f \in C_0(\mathbb{R}^d)$ and this determines μ. □

2.5 Moments and weak convergence

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

$$\overline{m_k}(\mu) = \int |x|^k \, d\mu(x) \tag{33}$$

If $\overline{m_k}(\mu) < \infty$ we call

$$m_k(\mu) = \int x^k \, d\mu(x) \tag{34}$$

the k^{th} *moment* of μ.

Remarks 2.40.

1. Observe that $\overline{m_{2k}}(\mu) = m_{2k}(\mu)$.
2. If $\overline{m_k}(\mu) < \infty$ then $\overline{m_l}(\mu) < \infty$ for all $l \leq k$. This can be seen by applying either Hölder's inequality or Jensen's inequality.
3. The first moment $\mathbb{E}(\mu) := m_1(\mu)$ is usually called the *mean* or the *expectation* of μ and the quantity $\mathbb{V}(\mu) := m_2(\mu) - m_1(\mu)^2$ is called the *variance* of μ.

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

$$n_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\,\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{35}$$

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

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:

$$n!! \; := \; \begin{cases} n \cdot (n-2) \cdot (n-4) \cdot \ldots \cdot 1 & \text{for } n \text{ odd} \\ n \cdot (n-2) \cdot (n-4) \cdot \ldots \cdot 2 & \text{for } n \text{ even} \end{cases} \tag{36}$$

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:*

$$m_k \;=\; m_k\Big(\mathcal{N}(0,\sigma^2)\Big) = \begin{cases} (k-1)!!\;(\sigma^2)^{k/2} & \textit{for } k \textit{ even} \\ 0 & \textit{for } k \textit{ odd} \end{cases} \tag{37}$$

Proof: For k odd, the function $x^k\, e^{-\frac{x^2}{2}}$ is odd, hence $m_k = 0$.

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

For arbitrary k even we compute, using integration by parts

$$\begin{aligned}
m_{k+2} \;&=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{k+2}\, e^{-\frac{x^2}{2}}\, dx \\
&=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{k+1} \left(x\, e^{-\frac{x^2}{2}}\right) dx \\
&=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{k+1} \left(e^{-\frac{x^2}{2}}\right)' dx \\
&=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(x^{k+1}\right)'\, e^{-\frac{x^2}{2}}\, dx \\
&=\; (k+1)\, \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k\, e^{-\frac{x^2}{2}}\, dx \\
&=\; (k+1)\, m_k
\end{aligned}$$

The assertion follows by induction. □

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*

$$m_{2k} \;=\; \frac{(2k)!}{2^k\, k!} \tag{38}$$

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

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

$$(2n)!! \;=\; 2^n\, n! \tag{39}$$

$$(2n-1)!! \;=\; \frac{(2n)!}{n!\, 2^n} \tag{40}$$

Proof: Equation (39) is immediate.

We prove equation (40) by induction:

$$\begin{aligned}
\big(2(n+1)-1\big)!! &= (2n+1)\cdot(2n-1)!! \\
&= (2n+1)\cdot\frac{(2n)!}{n!\,2^n} \\
&= \frac{(2n+1)!}{n!\,2^n} \\
&= \frac{\big(2(n+1)\big)!}{(n+1)!\,2^{n+1}}
\end{aligned}$$

□

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 $\overline{m_k}(\mu) < \infty$.

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 μ 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 μ 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

$$\begin{aligned}
\overline{m}_{2k-1}(\mu) &= \int |x|^{2k-1}\,d\mu \\
&\leq \Big(\int |x|^{2k}\,d\mu\Big)^{1/2}\Big(\int |x|^{2k-2}\,d\mu\Big)^{1/2} \\
&= (\overline{m}_{2k})^{1/2}\quad(\overline{m}_{2k-2})^{1/2}
\end{aligned}$$

□

Proposition 2.50. *Suppose μ_n is a sequence of bounded measures with existing moments and*

$$\sup_n m_k(\mu_n) \;\leq\; C_k \;<\; \infty\,. \tag{41}$$

If $\mu_n \Longrightarrow \mu$ then all moments of μ exist and

$$m_k(\mu_n) \to m_k(\mu) \qquad \textit{for all } k.$$

Proof: We estimate:

$$\begin{aligned}
\int |x|^k\,\phi_{L,2L}(x)\,d\mu(x) \;&\leq\; \sup_n \int |x|^k\,\phi_{L,2L}(x)\,d\mu_n(x)\\
&\leq\; \sup_n\, \overline{m}_k(\mu_n)\\
&\leq\; C_k\,.
\end{aligned}$$

Thus

$$\overline{m}_k(\mu) \;=\sup_L \int |x|^k\,\phi_{L,2L}(x)\,d\mu(x) \;\leq\; C_k \;<\; \infty\,.$$

So, all moments of μ exist.

To prove the second assertion of the proposition we estimate

$$\begin{aligned}
&\Big|\int |x|^k\,d\mu \;-\; \int |x|^k\,d\mu_n\Big|\\
\leq\;& \Big|\int |x|^k\,\phi_{L,2L}(x)\,d\mu \;-\; \int |x|^k\,\phi_{L,2L}(x)\,d\mu_n\Big|\\
&+ \int_{|x|\geq L} |x|^k\,d\mu_n \;+\; \int_{|x|\geq L} |x|^k\,d\mu
\end{aligned}$$

The first expression in the above sum goes to zero due to weak convergence. The second summand can be estimated by

$$\begin{aligned}
\int_{|x|\geq L} |x|^k\,d\mu_n \;&\leq\; \Big(\int |x|^{2k}\,d\mu_n\Big)^{1/2}\,\Big(\mu_n(|x|\geq L)\Big)^{1/2}\\
&\leq\; {C_k}^{1/2}\quad \mu_n\big(|x|\geq L\big)^{1/2}
\end{aligned}$$

Due to the tightness of the sequence μ_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 tighness 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 μ.

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

$$\overline{m_k}(\mu) \;\leq\; A\, C^k \, k! \tag{42}$$

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:

$$\begin{aligned}
\overline{\mu_{2k-1}} &= \int |x|^{2k-1} \, d\mu \\
&\leq \Big(\int |x|^{2k} \, d\mu\Big)^{1/2} \Big(\int |x|^{2k-2} \, d\mu\Big)^{1/2} \qquad \text{(by Cauchy-Schwarz)} \\
&\leq A\, C^{2k-1} \, (2k)!^{1/2} \, (2k-2)!^{1/2} \\
&\leq A\, C^{2k-1} \, (2k-2)! \, (2k-1)^{1/2} \, (2k)^{1/2} \\
&\leq 2\, A\, C^{2k-1} \, (2k-1)!
\end{aligned}$$

□

Examples 2.53.

1. If μ has compact support, then μ is a measure with moderately growing moments.

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

Theorem 2.54. *Suppose μ is a bounded measure with moderately growing moments (satisfying* (42)*, then:*

1. *The characteristic function*

$$\hat{\mu}(z) \;=\; \int e^{izx}\, d\mu(x) \tag{43}$$

is well defined on the strip

$$S \;=\; \big\{ z = z_1 + i\, z_2 \in \mathbb{C} \;\big|\; |z_2| < \tfrac{1}{C} \big\} \tag{44}$$

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

$$\begin{aligned}
|\hat{\mu}(z)| \;&\leq\; \int |e^{i\,z\,x}|\, d\mu(x) \\
&\leq\; \int e^{|z_2|\,x}\, d\mu(x) \\
&=\; \int \quad \sum_{k=0}^{\infty} \frac{|z_2|^k}{k!}\, |x|^k\, d\mu(x) \\
&=\; \sum_{k=0}^{\infty} \frac{|z_2|^k}{k!}\, \overline{m_k}(\mu) \\
&\leq A \, \sum_{k=0}^{\infty} \, |z_2|^k\, C^k \\
&< \infty \qquad\qquad \text{for } |z_2| < C^{-1}
\end{aligned}$$

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

$$\begin{aligned}
\hat{\mu}(z_0 + \zeta) \;&=\; \int \; e^{iz_0x} \, \sum_{k=0}^{\infty} \frac{\zeta^k}{k!}\, x^k\, d\mu(x) \\
&=\; \sum_{k=0}^{\infty} \Big(\frac{\int\; e^{iz_0x}\, x^k\; d\mu}{k!} \Big)\, \zeta^k
\end{aligned}$$

Interchanging sum and integral is justified by the estimate

$$\left| \frac{\int e^{iz_0 x} \, x^k \, d\mu(x)}{k!} \right| \;\leq\; \frac{\int |x|^k \, d\mu(x)}{k!} \;\leq\; A\,C^k \,.$$

Above C denotes the constant from (42). □

Theorem 2.55. *Suppose μ is a bounded measure with moderately growing moments. If for some bounded measure ν*

$$m_k(\mu) = m_k(\nu) \tag{45}$$

for all $k \in \mathbb{N}$ then

$$\mu = \nu \tag{46}$$

Proof: Since for all even k we have $m_k(\nu) = m_k(\mu) < \infty$ we have that all moments of ν 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

$$\hat{\mu}(\zeta) \;=\; \sum_{k=0}^{\infty} m_k(\mu)\, \zeta^k \;=\; \sum_{k=0}^{\infty} m_k(\nu)\, \zeta^k \;=\; \hat{\nu}(\zeta) \tag{47}$$

Thus, $\hat{\mu}$ and $\hat{\nu}$ agree on a neighborhood of $z = 0$ and hence in the whole strip S. □

Theorem 2.56. *Suppose μ_n is a sequence of bounded measures and $m_k \in \mathbb{R}$ is a sequence with*

$$m_k \;\leq\; A\,C^k\, k! \tag{48}$$

for all $k \in \mathbb{N}$.

If

$$m_k(\mu_n) \;\to\; m_k \tag{49}$$

for all $k \in \mathbb{N}$ then there is a bounded measure μ with $m_k(\mu) = m_k$ and such that

$$\mu_n \;\Longrightarrow\; \mu \tag{50}$$

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

2.6 Random Variables and Their Distribution

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

$$P_X(A) \;=\; \mathbb{P}\,(X \in A) \;:=\; \mathbb{P}\,\{\,\omega \in \Omega \mid X(\omega) \in A\} \tag{51}$$

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 $\big(X_1, X_2, \ldots, X_N\big)$ is called the *joint distribution* of $X_1, \ldots, X_N$, i. e.

$$P_{X_1, \ldots, X_N}\,(A) \;=\; \mathbb{P}\Big(\big(X_1, X_2, \ldots, X_N\big) \in A\Big) \tag{52}$$

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

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

$$P_{X_1, X_2, \ldots, X_N}\Big(A_1 \times A_2 \times \ldots \times A_N\Big) \;=\; P_{X_1}(A_1) \cdot P_{X_2}(A_2) \cdot \ldots \cdot P_{X_N}(A_N) \tag{53}$$

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 $X_n \overset{\mathcal{D}}{\Longrightarrow} X$, if the corresponding distributions converge, i. e. if $P_{X_n} \Longrightarrow P_X$.

Remark 2.61. Sometimes we also say X_n converges in distribution to a probability measure μ or $X_n \overset{\mathcal{D}}{\Longrightarrow} \mu$, if $P_{X_n} \Longrightarrow \mu$.

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

$$\mathbb{E}\Big(f(X_n)\Big) \longrightarrow \mathbb{E}\Big(f(X_n)\Big) \qquad \text{for all } f \in C_b(\mathbb{R}^d) \tag{54}$$

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

$$\mathbb{P}(|X_n - X| > \varepsilon) \longrightarrow 0 \qquad \text{as } n \to \infty$$

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

Proof:

$$\begin{aligned} \mathbb{P}(|X_n - c| > \varepsilon) &= 1 - \mathbb{P}(|X_n - c| \leq \varepsilon) \\ &\leq 1 - \mathbb{E}\Big(\Phi_{\frac{\varepsilon}{2},\varepsilon}(X_n - c)\Big) \longrightarrow 0 \end{aligned}$$

□

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 $\mathbb{E}(|X_i|^N) < \infty$ for all $i = 1, \ldots, N$ then*

$$\mathbb{E}(|X_1 \cdot X_2 \cdot \ldots \cdot X_N|) \leq \max_{1 \leq i \leq N} \mathbb{E}(|X_i|^N) < \infty \tag{55}$$

Proof: By a repeated application of the Hölder inequality we get

$$
\begin{aligned}
mu \quad & \mathbb{E}\left(|X_1 X_2 \cdot \ldots \cdot X_N|\right) \\
\leq \quad & \mathbb{E}(|X_1|^N)^{1/N} \, \mathbb{E}(|X_2 \cdot \ldots \cdot X_N|^{N/(N-1)})^{(N-1)/N} \\
\leq \quad & \mathbb{E}(|X_1|^N)^{1/N} \, \mathbb{E}(|X_2|^N)^{1/N} \, \mathbb{E}(|X_3 \cdot \ldots \cdot X_N|^{N/(N-2)})^{(N-2)/N} \\
& \ldots \\
\leq \quad & \mathbb{E}(|X_1|^N)^{1/N} \, \mathbb{E}(|X_2|^N)^{1/N} \, \ldots \, \mathbb{E}(|X_N|^N)^{1/N} \\
\leq \quad & \max_{1 \leq i \leq N} \mathbb{E}(|X_i|^N) < \infty
\end{aligned}
$$

□

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 $(\Omega, \mathcal{F}, \mathbb{P})$. 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:

$$\mathbb{E}\big({S_N}^L\big) = \frac{1}{N^L} \sum_{i_1, i_2, \ldots i_L} \mathbb{E}\Big(X_{i_1} X_{i_2} \cdot \ldots \cdot X_{i_L}\Big) \tag{56}$$

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:

$$\mathbb{E}({S_N}^L) = \frac{1}{N^L} \sum_{\substack{i_1, i_2, \ldots i_L \\ \text{all indices distinct}}} \mathbb{E}(X_{i_1} X_{i_2} \cdot \ldots \cdot X_{i_L}) \tag{57}$$

$$+ \frac{1}{N^L} \sum_{\substack{i_1, i_2, \ldots i_L \\ \text{not all distinct}}} \mathbb{E}(X_{i_1} X_{i_2} \cdot \ldots \cdot X_{i_L}) \tag{58}$$

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 $N \to \infty$.

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

$$\mathbb{E}(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_L}) \;=\; \prod_{\ell=1}^{L} \mathbb{E}(X_{i_L}) \;=\; m^L \quad \text{for distinct } \; i_1, i_2, \ldots, i_L \tag{59}$$

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

$$(57) \;=\; \frac{1}{N^L} \frac{N!}{(N-L)!}\, m^L \quad \overset{N\to\infty}{\longrightarrow} \quad m^L \tag{60}$$

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

□

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

$$\mathbb{E}(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_L})\,. \tag{61}$$

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 $\underline{i} = (i_1, i_2, \ldots, i_L) \in \{1, 2, \ldots, N\}^L$.

1. For $j \in \{1, 2, \ldots, N\}$ we set

$$\nu_j(\underline{i}) = |\ \{k \in \{1, 2, \ldots, L\} \mid i_k = j\ \}\ | \,. \tag{62}$$

   Here $|M|$ denotes the number of elements in the set M.

2. For $\ell = 0, 1, \ldots, L$ we define

$$\rho_\ell(\underline{i}) \ = \ |\{j \mid \nu_j(\underline{i}) = \ell\}| \tag{63}$$

   and

$$\underline{\boldsymbol{\rho}}(\underline{i}) \ = \ (\rho_1(\underline{i}), \rho_2(\underline{i}), \ldots, \rho_L(\underline{i})) \tag{64}$$

In words: $\nu_j(\underline{i})$ tells us how frequently the index j occurs among the indices $(i_1, i_2, \ldots i_L)$ and $\rho_k(\underline{i})$ gives the number of indices that occur exactly k times in $(i_1, i_2, \ldots i_L)$.

Remark 3.5. Setting $\underline{\boldsymbol{r}} = (r_1, r_2, \ldots, r_L) = \underline{\boldsymbol{\rho}}(\underline{i})$ we have

$$\begin{aligned}
\mathbb{E}\left(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_L}\right) \ &= \ \mathbb{E}\left(\prod_{j=1}^{N} X_j^{\nu_j(\underline{i})}\right) \\
= \mathbb{E}&\left(\prod_{j_1=1}^{r_1} X_{j_1} \ \cdot \ \prod_{j_2=1}^{r_2} X^2_{r_1+j_2} \ \cdot \ \prod_{j_L=1}^{r_L} X^L_{\sum_{\ell=1}^{L-1} r_\ell + j_L}\right) \\
= \prod_{\ell=1}^{L} \mathbb{E}&\left(X_\ell^\ell\right)^{r_\ell}
\end{aligned}$$

Thus, we showed

Proposition 3.6. *If $X_n, n \in \mathbb{N}$ are independent and identically distributed random variables, then*

$$\mathbb{E}\left(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_L}\right) \ = \ \mathbb{E}\left(X_{j_1} \cdot X_{j_2} \cdot \ldots \cdot X_{j_L}\right) \tag{65}$$

whenever $\underline{\boldsymbol{\rho}}(i_1, i_2, \ldots, i_L) = \underline{\boldsymbol{\rho}}(j_1, j_2, \ldots, j_L)$.

Therefore the following abbreviation is well defined.

Notation 3.7. If $\underline{\boldsymbol{r}} = (r_1, r_2, \ldots, r_L) = \underline{\boldsymbol{\rho}}(\underline{i})$ we set

$$\mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \ = \ \mathbb{E}\big(X_{i_1}\, X_{i_2}\, \ldots\, X_{i_L}\big) \tag{66}$$

To evaluate sums of expressions as in (61) we investigate the combinatorics of $L-$tuples of indices.

Lemma 3.8. *For each $\underline{i} \in \{1, 2, \dots, N\}^L$ we have:*

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

2. $0 \leq \rho_\ell(\underline{i}) \leq L,$ *for* $\ell \neq 0$ *and* $\rho_\ell(\underline{i}) = 0$ *for* $\ell > L,$

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

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

   i.e. the sum over the ρ_ℓ is the number of distinct *indices in $\underline{\boldsymbol{i}}$.*

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

The proof of the Lemma is left to the reader.

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

$$\sum_{\ell=1}^{L} \ell \, r_\ell = L \, . \tag{67}$$

Given a profile $\underline{\boldsymbol{r}}$ we define

$$W(\underline{\boldsymbol{r}}) = W_L(\underline{\boldsymbol{r}}) \;=\; \left\{ \underline{\boldsymbol{i}} \in \mathcal{N}^k \mid \underline{\boldsymbol{\rho}}(\underline{\boldsymbol{i}}) = \underline{\boldsymbol{r}} \right\} \tag{68}$$

$$\text{and } w(\underline{\boldsymbol{r}}) = w_L(\underline{\boldsymbol{r}}) \;=\; \left| \left\{ \underline{i} \in \mathcal{N}^k \mid \underline{\boldsymbol{\rho}}(\underline{\boldsymbol{i}}) = \underline{\boldsymbol{r}} \right\} \right| \tag{69}$$

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 $(i_1, i_2, \dots i_L)$. 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 $\mathcal{L}^L$ by $\Pi^{(L)}$ or simply by Π, if L is clear from the context, thus

$$\Pi \ := \ \Pi^{(L)} \ := \ \left\{ (r_1, r_2, \ldots, r_L) \in \{0, 1, \ldots, L\}^L \mid \sum_{\ell=1}^{L} \ell \, r_\ell = L \right\} \tag{70}$$

For later use, we define some subsets of Π:

$$\Pi_k \ := \ \{ \underline{\boldsymbol{r}} \in \Pi^{(L)} \mid r_1 = k \} \tag{71}$$

$$\Pi_+ \ := \ \{ \underline{\boldsymbol{r}} \in \Pi^{(L)} \mid r_1 > 0 \} \tag{72}$$

$$\Pi^0 \ := \ \{ \underline{\boldsymbol{r}} \in \Pi^{(L)} \mid r_\ell = 0 \text{ for all } \ell \geq 3 \} \tag{73}$$

$$\Pi^+ \ := \ \{ \underline{\boldsymbol{r}} \in \Pi^{(L)} \mid r_\ell > 0 \text{ for some } \ell \geq 3 \} \tag{74}$$

Remark 3.11. We may combine the sub- and superscripts in the previous definition, for example

$$\Pi_k^+ \ = \ \{ \underline{\boldsymbol{r}} \in \Pi^{(L)} \mid r_1 = k \text{ and } r_\ell > 0 \text{ for some } \ell \geq 3 \} \tag{75}$$

In particular the set Π_0^0 consists of those $\underline{\boldsymbol{r}}$ with $r_\ell = 0$ for all $\ell \neq 2$. Note that $\Pi_0^0 = \emptyset$ if L is odd.

The above defined sets decompose Π into disjoint subsets in various ways, for example

$$\begin{aligned} \Pi \ &= \ \bigcup_{k=0}^{L} \Pi_k \\ &= \ \Pi_+ \cup \Pi_0^0 \cup \Pi_0^+ \end{aligned}$$

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 $\sum\limits_{\ell=1}^{L} \ell \, r_\ell = L$ 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 $(r_1, r_2, \ldots, r_L)$. $\square$

We note the following important observation

Theorem 3.13. *For independent identically distributed random variables X_i we have*

$$\sum_{i_1, i_2, \ldots i_L = 1}^{N} \mathbb{E}\left(X_{i_1} X_{i_2} \cdot \ldots \cdot X_{i_L}\right) = \sum_{\underline{\boldsymbol{r}} \in \Pi} w_L(\underline{\boldsymbol{r}})\, \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big)$$
$$= \sum_{\underline{\boldsymbol{r}} \in \Pi} w_L(\underline{\boldsymbol{r}}) \prod_{\ell=1}^{L} \mathbb{E}\left(X_1^{\ell}\right)^{r_\ell} \tag{76}$$

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

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

$$w_L(\underline{\boldsymbol{r}}) = \frac{N!}{r_1!\, r_2! \ldots r_L!\, r_0!} \; \frac{L!}{1!^{r_1}\; 2!^{r_2}\; 3!^{r_3}\; \ldots\; L!^{r_L}} \tag{77}$$

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*

$$\begin{pmatrix} & M & \\ m_1 \; m_2 & \ldots & m_K \end{pmatrix} = \frac{M!}{m_1!\, m_2! \ldots m_K!} \,. \tag{78}$$

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 object ('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*

$$\begin{pmatrix} & M & \\ m_1 \; m_2 & \ldots & m_K \end{pmatrix} \tag{79}$$

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

$$\binom{M}{m_1} \cdot \binom{M - m_1}{m_2} \cdot \binom{M - m_1 - m_2}{m_3} \cdot \ldots \cdot \binom{M - \sum_{k=1}^{K-1} m_k}{m_K} \tag{80}$$

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

$$\binom{M}{m_1\ m_2\ \ldots\ m_K} \tag{81}$$

□

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 prescribes 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*

$$\binom{M}{m_1\ m_2\ \ldots\ m_K} \tag{82}$$

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

To implement this idea, we partition the set $\mathcal{M}$ 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

$$w_\ell \;=\; a_k \qquad \text{if and only if } \; a_k \in A_\ell \tag{83}$$

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 $\mathcal{M}$ is given by

$$\begin{pmatrix} M \\ m_1 \; m_2 \; \ldots \; m_K \end{pmatrix} \tag{84}$$

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 $\mathcal{N}$ with profile $(r_1, r_2, \ldots, r_L)$. To do so we first partition the set $\mathcal{N}$ into classes $R_1, R_2, \ldots, R_L$ and R_0. The class R_ℓ will be interpreted as those numbers in $\mathcal{N}$ which are going to occur exactly ℓ 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_ℓ elements in the class R_ℓ, 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

$$\begin{pmatrix} N \\ r_1 \; r_2 \; \ldots r_L \; r_0 \end{pmatrix} \;=\; \frac{N!}{r_1! \, r_2! \; \ldots r_L! \, r_0!} \tag{85}$$

which is the first fraction in (77).

Next, we have to construct an L-tuple $\underline{\boldsymbol{i}} = (i_1, i_2, \ldots, i_L)$ 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:

$$\begin{pmatrix} L \\ \underbrace{1\,1 \ldots 1}_{r_1\text{-times}} \; \underbrace{2\,2 \ldots 2}_{r_2\text{-times}} \; \underbrace{3\,3 \ldots 3}_{r_3\text{-times}} \; \ldots \; \underbrace{L\,L \ldots L}_{r_L\text{-times}} \end{pmatrix} \;=\; \frac{L!}{1!^{r_1} \, 2!^{r_2} \, 3!^{r_3} \; \ldots \; L!^{r_L}} \tag{86}$$

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* $\underline{\boldsymbol{r}} \in \Pi^{(L)}$ *we have*

$$w_L(\underline{\boldsymbol{r}}) \;\approx\; C(\underline{\boldsymbol{r}})\, N^{\sum_{\ell=1}^{L} r_\ell} \qquad \text{as } N \to \infty \tag{87}$$

where the constant $C(\underline{\boldsymbol{r}})$ *is given by*

$$C(\underline{\boldsymbol{r}}) \;=\; \frac{1}{r_1!\, r_2!\, \ldots\, r_L!}\, \frac{L!}{\prod_{\ell=1}^{L} \ell!^{r_\ell}} \;\leq\; L! \tag{88}$$

Notation 3.19. The expression

$$a_N \;\approx\; b_N \qquad \text{as} \quad N \longrightarrow \infty$$

is an abbreviation for

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

We used this Corollary tacitely and in a weak form already in the proof of Theorem 3.1, when we concluded that the contribution of those $\underline{\boldsymbol{i}}$ with $\sum_{\ell=2}^{L} \rho_\ell(\underline{\boldsymbol{i}}) > 0$ 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 in dependent 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

$$\Sigma_N \;=\; \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Big(X_i - E\big(X_i\big) \Big) \tag{89}$$

converges in distribution to a normal distribution $\mathcal{N}(0, \sigma^2)$ *with mean zero and variance* σ^2*.*

Proof: By theorem 2.56 it is sufficient to prove that the moments of Σ_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:

$$\begin{aligned}
\mathbb{E}\Big((\Sigma_N)^L\Big) &= \frac{1}{N^{L/2}} \sum_{i_1, i_2, \dots i_L} \mathbb{E}\big(X_{i_1} X_{i_2} \cdot \ldots \cdot X_{i_L}\big) \\
&= \frac{1}{N^{L/2}} \sum_{\underline{\boldsymbol{r}} \in \Pi} w_L(\underline{\boldsymbol{r}})\, \mathbb{E}\Big(X(\underline{\boldsymbol{r}})\Big) \\
&= \frac{1}{N^{L/2}} \sum_{\underline{\boldsymbol{r}} \in \Pi_+} w_L(\underline{\boldsymbol{r}})\, \mathbb{E}\Big(X(\underline{\boldsymbol{r}})\Big) \qquad (90) \\
&\quad + \frac{1}{N^{L/2}} \sum_{\underline{\boldsymbol{r}} \in \Pi_0^0} w_L(\underline{\boldsymbol{r}})\, \mathbb{E}\Big(X(\underline{\boldsymbol{r}})\Big) \qquad (91) \\
&\quad + \frac{1}{N^{L/2}} \sum_{\underline{\boldsymbol{r}} \in \Pi_0^+} w_L(\underline{\boldsymbol{r}})\, \mathbb{E}\Big(X(\underline{\boldsymbol{r}})\Big) \qquad (92)
\end{aligned}$$

where the sets Π_+, Π_0^0 and Π_0^+ are defined in Definition 3.10.

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

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

   $\mathbb{E}(X_i) = 0$ we have

   $$\mathbb{E}\big(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_L}\big) = \mathbb{E}\big(X_{i_j}\big)\, \mathbb{E}\Big(\prod_{\ell \neq j} X_{i_\ell}\Big) = 0$$

   Therefore, the summand (90) vanishes.

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

   Consequently, since

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

we get

$$\begin{aligned}
\sum_{\ell=1}^{L} r_\ell &= \frac{1}{2}\,\big(2r_2 + 2\sum_{\ell=3}^{L} r_\ell\big)\\
&\leq \frac{1}{2}\,\big(2r_2 + \frac{2}{3}\sum_{\ell=3}^{L} \ell\, r_\ell\big)\\
&\leq \frac{1}{2}\,\big(2r_2 + \sum_{\ell=3}^{L} \ell\, r_\ell - \frac{1}{3}\sum_{\ell=3}^{L} \ell\, r_\ell\big)\\
&\leq \frac{1}{2}\,\big(2r_2 + \sum_{\ell=3}^{L} \ell\, r_\ell - 1\big)\\
&= \frac{1}{2}\sum_{\ell=1}^{L} \ell\, r_\ell \;-\frac{1}{2} = \; L/2 - 1/2 \qquad (93)
\end{aligned}$$

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

$$\Big|\,\frac{1}{N^{L/2}}\sum_{\underline{\boldsymbol{r}}\in\Pi_0^+} w_L(\underline{\boldsymbol{r}})\;\mathbb{E}\Big(X(\underline{\boldsymbol{r}})\Big)\;\Big| \;\leq\; \frac{N^{L/2-1/2}}{N^{L/2}}L!\;(L+1)^L\;\mathbb{E}\big(|X_1|^L\big)$$

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 Π_0^0 consist of those $\underline{\boldsymbol{r}}$ 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 $(0, L/2, 0, \dots, 0)$ for *even* L.

   Thus, for odd L

$$\frac{1}{N^{L/2}}\sum_{\underline{\boldsymbol{r}}\in\Pi_0^0} w_L(\underline{\boldsymbol{r}})\;\mathbb{E}\Big(X(\underline{\boldsymbol{r}})\Big) \;=\; 0 \qquad (94)$$

For even $L = 2\,K$ we have

$$
\begin{aligned}
& \frac{1}{N^{L/2}} \sum_{\underline{\boldsymbol{r}} \in \Pi_0^0} w_L(\underline{\boldsymbol{r}})\, \mathbb{E}\Big(X(\underline{\boldsymbol{r}})\Big) \\
= & \frac{1}{N^K}\, w_L\big((0, K, 0, \ldots, 0)\big)\, \mathbb{E}\Big({X_1}^2\Big)^K \\
= & \frac{1}{N^K}\, \frac{N!}{(K)!\,(N-K)!}\, \frac{(2K)!}{2^K}\, \big(\sigma^2\big)^K \\
= & \frac{1}{N^K}\, \frac{N!}{(N-K)!}\, (2K-1)!!\, \big(\sigma^2\big)^K \qquad \text{by Lemma 2.46} \\
\to & \; (2K-1)!!\, \big(\sigma^2\big)^K \qquad\qquad \text{as } N \to \infty
\end{aligned}
\tag{95}
$$

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. □

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 $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a non decreasing function then

$$\mathbb{P}(|X| > a) \leq \frac{\mathbb{E}(f(X))}{f(a)} \tag{96}$$

Proof:

$$\mathbb{P}(|X| > a) = \int \chi_{\{|X|>a\}}(w)\, d\mathbb{P}(w) \tag{97}$$

$$\leq \int \chi_{\{|X|>a\}}(w) \frac{f(|X|)(w)}{f(a)}\, d\mathbb{P}(w) \tag{98}$$

$$\leq \int \frac{f(|X|)}{f(a)}\, d\mathbb{P} = \frac{\mathbb{E}(f(X))}{f(a)} \tag{99}$$

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

Corollary 3.23. *For all* $p > 0$:

$$\mathbb{P}(|X| > a) \leq \frac{1}{a^p}\mathbb{E}(|X|^p) \tag{100}$$

Chebyshev's inequality

$$\mathbb{P}(|X - E(X)| > a) \leq \frac{1}{a^2}E(|X - E(X)|^2) \tag{101}$$

$$= \frac{\mathbb{V}(X)}{a^2} \tag{102}$$

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 $\mathbb{P}(|\frac{1}{N}E(X_i - E(X_i))| > a)$.

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* $a > 0$

$$\mathbb{P}\Big(\Big|\frac{1}{N}\sum_{i=1}^{N}(X_i - E(X_i))\Big| > a\Big) \leq \frac{C_L}{a^{2L}N^L} \tag{103}$$

Remark 3.25. The constant C_L can be chosen to be

$$C_L = \frac{(2L+1)^{2L}(2L)!}{2^{2L}} \sup_{i\in\mathbb{N}} \mathbb{E}((X_i - E(X_i))^{2L}) \tag{104}$$

Remark 3.26. Theorem 3.24 is a kind of a quantitive version of the weak law of large numbers. It asserts that $\mathbb{P}(|\frac{1}{N}\sum_{i=1}^{N}(X_i - E(X_i))| > a)$ not only converges to zero, but it converges faster than any power of N.

Proof: Using the Chebyshev-Markov inequality we obtain

$$\mathbb{P}\Big(\Big|\frac{1}{N}\sum_{i=1}^{N}(X_i - E(X_i))\Big| > a\Big) \tag{105}$$

$$= \mathbb{P}\Big(\Big|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(X_i - E(X_i))\Big| > a\, N^{\frac{1}{2}}\Big) \tag{106}$$

$$\leq \frac{1}{a^{2L}N^L}\ \mathbb{E}\Big(\Big|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\big(X_i - E(X_i)\big)\Big|^{2L}\Big) \tag{107}$$

In the proof of Theorem 3.20 we have shown that the above expecation 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 $Y_i = X_i - E(X_i)$.

$$\mathbb{E}\Big((\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Y_i)^{2L}\Big) \leq \frac{1}{N^L}\sum_{\underline{r}\in\Phi^{(2L)}} w(\underline{r})\Big|\mathbb{E}\big(Y(\underline{r})\big)\Big| \tag{108}$$

$$= \frac{1}{N^L}\sum_{r\in\Phi_0} w(\underline{r})\ \Big|\mathbb{E}(Y(\underline{r}))\Big|\ \text{(since } \mathbb{E}(Y_i) = 0) \tag{109}$$

$$\leq \sum_{r\in\Phi_0} \frac{(2L)!}{2^{2L}} \sup_{i\in\mathbb{N}} \mathbb{E}(Y_i^{2L})\ \text{(due to Corollary 3.18 and Lemma 2.65)} \tag{110}$$

$$\leq \frac{(2L+1)^{2L}(2L)!}{2^{2L}} \sup_{i\in\mathbb{N}} \mathbb{E}(Y_i^{2L})\ \text{(by Lemma 3.12)} \tag{111}$$

□

We single out the following result:

Corollary 3.27. *Under the assumptions of Theorem 3.24 we have*

$$\mathbb{E}\Big((\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(X_i - E(X_i))^{2L}\Big) \leq \frac{(2L+1)^{2L}(2L)!}{2^{2L}} \quad \sup_{i\in\mathbb{N}} \mathbb{E}\Big(\big(X_i - E(X_i)\big)^{2L}\Big) \tag{112}$$

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 $\sup_{i\in\mathbb{N}} \mathbb{E}(X_i^K) \le M_K < \infty$ for all K.

Then

$$S_N \;:=\; \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}(X_i - E(X_i)) = \; 0 \qquad \mathbb{P}\text{-almost surely,} \tag{113}$$

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

Remark 3.29. If all $\mathbb{E}(X_i)$ are equal then $\frac{1}{N}\sum_{i=1}^{N} X_i \to E(X_1)$ $\mathbb{P}$-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. $\limsup_n A_n \;:=\; \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m$ is called the *lim-sup of the A_n*.

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

Lemma 3.31. *1.* $\limsup_{n\to\infty} A_n = \{\omega \mid \omega \in A_n$ *for infinitely many* $n\}$.

2. $\liminf_{n\to\infty} A_n = \{\omega \mid \omega \in A_n$ *for all n 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 $m \ge n$ with $\omega \in A_m$. By definition

$$\limsup_{n\to\infty} A_n \;=\; \bigcap_{n=1}^{\infty} B_n \tag{115}$$

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. □

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* $\mathbb{P}\big(\limsup_{n\to\infty} A_n\big) = 0$.

2. *If* $\sum_{n=1}^{\infty} \mathbb{P}(\complement A_n) < \infty$, *then* $\mathbb{P}\big(\liminf_{n\to\infty} A_n\big) = 1$.

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

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

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. $\liminf_{n\to\infty} A_n = \complement \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \complement A_m$.

Thus 2. follows from 1. □

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:

$$A = \{\omega \mid \forall k \in \mathbb{N}\ \ \exists n \in \mathbb{N}\ \ \forall m \geq n\ \ |Y_m| < \frac{1}{k}\} \tag{116}$$

$$= \bigcap_{k\in\mathbb{N}} \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} A_{m,k} \tag{117}$$

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

$$\mathbb{P}(A) = \lim_{k\to\infty} \mathbb{P}(M_k) \tag{118}$$

$$\mathbb{P}(M_k) = \mathbb{P}(\bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} A_{m,k}) \tag{119}$$

$$= \mathbb{P}(\liminf_{n\to\infty} A_{n,k}) \tag{120}$$

$$= 1 \tag{121}$$

since $\sum_{n=1}^{\infty} \mathbb{P}(\complement A_{n,k}) = \sum_{n=1}^{\infty} \mathbb{P}(|Y_n| > \frac{1}{k}) < \infty$ by assumption, hence $\mathbb{P}(A) = 1$. □

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

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

which proves the assertion. □

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 π of $\{1, \ldots, N\}$ we have

$$\mathbb{P}\Big((X_1, \ldots, X_N) \in A\Big) \;=\; \mathbb{P}\Big((X_{\pi(1)}, \ldots, X_{\pi(N)}) \in A\Big) \tag{122}$$

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 ℓ (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*

$$\text{Cov}\big(X_i, X_j\big) \;:=\; \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) \tag{123}$$

of X_i and X_j is bounded from below by

$$\text{Cov}\big(X_i, X_j\big) \;=\; \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) \;\geq\; -\frac{1}{N-1}\,\mathbb{V}(X_i)$$

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

$$\begin{aligned}
\mathbb{E}(\tilde{X}_1\tilde{X}_2) &= \frac{1}{N(N-1)}\sum_{\substack{i,j=1\\ i\neq j}}^{N}\mathbb{E}(\tilde{X}_i\tilde{X}_j)\\
&= \frac{1}{N(N-1)}\sum_{i,j=1}^{N}\mathbb{E}(\tilde{X}_i\tilde{X}_j) - \frac{1}{N(N-1)}\sum_{i=1}^{N}\mathbb{E}(\tilde{X}_i^2)\\
&= \frac{1}{N(N-1)}\mathbb{E}((\sum_{i=1}^{N}\tilde{X}_i)^2) - \frac{1}{N-1}\mathbb{E}(\tilde{X}_1^2)\\
&\geq -\frac{1}{N-1}\mathbb{E}(\tilde{X}_1^2)
\end{aligned}$$

□

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

$$\mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\,\mathbb{E}(X_2) \;=\; -\frac{1}{N-1}\,\mathbb{V}(X_1)\,. \tag{124}$$

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

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

We compute

$$\begin{aligned}\mathbb{E}(X_1 X_2) \;&=\; \frac{P}{P+M}\frac{P-1}{P+M-1} + \frac{M}{P+M}\frac{M-1}{P+M-1}\\ &\quad - \frac{P}{P+M}\frac{M}{P+M-1} - \frac{M}{P+M}\frac{P}{P+M-1}\\ &= \frac{(P-M)^2 - (P+M)}{(P+M)^2 - (P+M)}\end{aligned}$$

So, we conclude

$$\mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\,\mathbb{E}(X_2)$$

$$= \frac{(P-M)^2-(P+M)}{(P+M)(P+M-1)} - \frac{(P-M)^2}{(P+M)^2}$$

$$= -\ \frac{1}{N-1}\left(\frac{(P-M)^2\,(P+M-1)}{(P+M)^2}\ -\ \frac{(P-M)^2-(P+M)}{P+M}\right)$$

$$= -\ \frac{1}{N-1}\left(1-\frac{(P-M)^2}{(P+M)^2}\right)$$

$$= -\ \frac{1}{N-1}\,\mathbb{V}\big(X_1\big)$$

□

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*

$$S_N = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{125}$$

converge in distribution to a probability measure μ.

The support supp(μ) *of μ is contained in the (possibly infinite) interval*

$$\big[\inf(\text{supp}P_{X_1}), \sup(\text{supp}P_{X_1})\big]$$

and the moments $m_k(\mu)$ of μ are given by:

$$m_k(\mu) \;=\; \mathbb{E}(X_1\cdot\ldots\cdot X_k) \tag{126}$$

Definition 4.6. We call the measure μ 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)

$$\big|\,\mathbb{E}(X_1, X_2, \ldots, X_K)\,\big| \;\leq\; \mathbb{E}(|X_1|^K) \tag{127}$$

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

$$\begin{aligned} \mathbb{E}(S_N^K) \;&=\; \frac{1}{N^K} \quad \mathbb{E}\Big(\Big(\sum_{i=1}^{N} X_i\Big)^K \Big) \\ &=\; \frac{1}{N^K} \sum_{i_1,\ldots,i_K=1}^{N} \mathbb{E}\big(X_{i_1}, X_{i_2}, \ldots, X_{i_K}\big) \\ &=\; \frac{1}{N^K} \sum_{\substack{i_1,\ldots,i_K \\ \text{all distinct}}} \mathbb{E}\big(X_1, X_2, \ldots, X_K\big) \\ &\quad+\; \frac{1}{N^K} \sum_{\substack{i_1,\ldots,i_K \\ \text{not all distinct}}} \mathbb{E}\big(X_{i_1}, X_{i_2}, \ldots, X_{i_K}\big) \end{aligned} \tag{128}$$

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

$$\mathbb{E}(X_1, X_2, \ldots, X_K) \qquad \text{as } N \to \infty$$

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 $-1 \le r \le 1$ let $P_r^{(1)}$ be the probability measure on $\{-1,1\}$ with

$$P_r^{(1)}(1) = \frac{1}{2}(1+r) \qquad \text{and} \qquad P_r^{(1)}(-1) = \frac{1}{2}(1-r) \tag{129}$$

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

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

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

$$\begin{aligned} P_r^{(N)}\Big((x_1,x_2,\ldots,x_N)\Big) \;&=\; \prod_{i=1}^{N} \, P_r^{(1)}(x_i) \\ &=\; \frac{1}{2^N}\,(1+r)^{n_+(x_1,\ldots,x_N)}\,(1-r)^{n_-(x_1,\ldots,x_N)} \end{aligned} \tag{131}$$

where $x_i \in \{-1,1\}$ and $n_\pm(x_1,\ldots,x_N) = \#\{i \in \{1,\ldots,N\} \mid x_i = \pm 1\}$. 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 ξ_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 μ 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* μ if

$$\begin{aligned} &\mathbb{P}(X_1 = a_1, X_2 = a_2, \ldots, X_N = a_N) \\ =& \int_{-1}^{1} P_t^{(N)}\Big((a_1,a_2,\ldots,a_N)\Big)\,d\mu(t) \\ =& \frac{1}{2^N}\int_{-1}^{1} (1+t)^{n_+(a_1,\ldots,a_N)}\,(1-t)^{n_-(a_1,\ldots,a_N)}\,d\mu(t) \end{aligned} \tag{132}$$

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

$$\mathbb{P}_\mu(x_1, x_2, \ldots, x_N) \;=\; \int_{-1}^{1} P_t^{(N)}\Big((x_1, x_2, \ldots, x_N)\Big)\, d\mu(t) \tag{133}$$

where

$$\mathbb{P}_\mu(x_1, x_2, \ldots, x_N) \tag{134}$$

is a short hand notation for

$$\mathbb{P}_\mu\Big(\{\omega \in \{1, +1\}^N \mid \omega_1 = x_1, \omega_2 = x_2, \ldots, \omega_N = x_N\}\Big) \tag{135}$$

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 μ, then*

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

$$\mathbb{E}_\mu\Big(F(X_1, X_2, \ldots, X_N)\Big) = \int E_t(F(X_1, \ldots, X_N))\, d\mu(t) \tag{136}$$

2.

$$\mathbb{E}_\mu\left(X_1, X_2, \ldots, X_K\right) = \int_{-1}^{1} t^K\, d\mu(t) = m_K(\mu) \tag{137}$$

3. $$\begin{aligned} Cov_\mu(X_1, X_2) :&= \mathbb{E}_\mu(X_1\, X_2) - \mathbb{E}_\mu(X_1)\, \mathbb{E}_\mu(X_2) \\ &= \int t^2\, d\mu(t) - \Big(\int t\; d\mu(t)\Big)^2 \geq 0 \end{aligned} \tag{138}$$

4. *$X_1, \ldots, X_N$ are independent if and only if μ 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 μ.*

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_r)

$$\begin{aligned} E_t(X_1, X_2, \ldots, X_K) &= E_t(X_1) E_t(X_2) \ldots E_t(X_K) \\ &= t^K \end{aligned} \tag{139}$$

3) With $m := \int t\, d\mu(t)$ we compute

$$\text{Cov}_\mu(X_1, X_2) = \int t^2\, d\mu(t) - \Big(\int t\; d\mu(t)\Big)^2 \tag{140}$$

$$= \int (t-m)^2\, d\mu(t) \tag{141}$$

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 $\mathbb{P}$) the covariance Cov(X_1, X_2) must vanish, so μ has to be δ_m.

On the other hand, if $\mu = \delta_m$ then

$$\mathbb{P}_\mu(X_1 = a_1, \ldots, X_N = a_N) = P_m(X_1 = a_1, \ldots, X_N = a_N) \tag{142}$$

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 ν with

$$m_K(\nu) \;=\; \mathbb{E}_\mu(X_1 \cdot X_2 \cdot \ldots \cdot X_K) \;= m_K(\mu)$$

Hence $\nu = \mu$. □

4.4 Sequences of de-Finetti-type

In this section we study sequences $\{X_1^{(N)}, \dots, X_N^{(N)}\}_{N\in\mathbb{N}}$ of de-Finetti-type with corresponding de-Finetti-measures μ_N and their behavior for $N \to \infty$.

Definition 4.13. For each $N \in \mathbb{N}$ let $X_1^{(N)}, \dots, X_N^{(N)}$ be a sequence of de-Finetti-type with de-Finetti-measure μ_N.

Then the scheme

$$\{\{X_1^{(N)}\}_{i=1}^N\}_{N=1,2,\dots}$$

is called a *de-Finetti sequence*.

We say that a de Finetti sequence converges in distribution to a probability measure $\mathbb{P}_\mu$ on $\{-1,1\}^N$ if for each $M \le N$

$$\mathbb{P}_{\mu_N}(x_1, \dots, x_M) \to \mathbb{P}_\mu(x_1, \dots, x_M) \tag{143}$$

In this case we write $\mathbb{P}_{\mu_N} \Longrightarrow \mathbb{P}_\mu$.

Proposition 4.14. *If $\mu_N \Longrightarrow \mu$ then $\mathbb{P}_{\mu_N} \Longrightarrow \mathbb{P}_\mu$.*

The proof follows immediately from the definition.

Theorem 4.15. *If $\mu_N \Longrightarrow \mu$ then*

$$\frac{1}{N}\sum_{i=1}^N X_i^{(N)} \stackrel{\mathcal{D}}{\Longrightarrow} \mu \tag{144}$$

Proof: We compute the moments of the random variables $S_N = \frac{1}{N}\sum\limits_{I=1}^N X_i^{(N)}$. As in (128) we write

$$\begin{aligned}\mathbb{E}_{\mu_N}(S_N^K) = \frac{1}{N^K}\frac{N!}{(N-K)!} \quad & \mathbb{E}_{\mu_N}(X_1 \cdot X_2 \cdot \ldots \cdot X_K) \\ + \frac{1}{N^K} \sum_{i_1,\dots,i_K \text{not all distinct}} & \mathbb{E}_{\mu_N}(X_{i_1} \cdot X_{i_2} \cdot \ldots \cdot X_{i_K})\end{aligned} \tag{145}$$

Since by assumption $\mu_N \Longrightarrow \mu$ we know that the first term of (145) converges to $\mathbb{E}_\mu(X_1 \cdot X_2 \cdot \ldots \cdot X_K)$. The second term can be bounded by $\frac{1}{N^K}(1 - \frac{N!}{(N-K)!})$ 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,

$$\mathbb{P}\big(\#\{i \leq L | X_i = -1\} = k\big) = \binom{N}{k} p_k^N \quad (147)$$

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 μ its associated de Finetti measure then*

$$\mathbb{P}(X_1 = a_1, X_2 = a_2, \ldots, X_N = a_N) \;=\; \int_{-1}^{1} P_t^{(N)}\Big((a_1, a_2, \ldots, a_N)\Big)\, d\mu(t)\,.$$

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

Proof: Given the measure μ we know due to Theorem 4.5 that

$$E_k \;:=\; \mathbb{E}\Big(X_1 \cdot X_2 \cdot \ldots \cdot X_k\Big) = \int t^k \, d\mu\,. \quad (149)$$

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

$$p_0^1 \;=\; \frac{1}{2}\big(1 + \mathbb{E}(X_1)\big)$$

and

$$p_1^1 \;=\; \frac{1}{2}\big(1 - \mathbb{E}(X_1)\big)$$

Induction step: Suppose, we can compute the numbers p_k^N from the knowledge of the $E_k = \mathbb{E}\Big(X_1 \cdot X_2 \cdot \ldots \cdot X_k\Big)$ for $k = 0, \ldots, N$.

We have:

$$p_k^{N+1} = \mathbb{P}\Big(X_1 = -1, \ldots, X_k = -1, X_{k+1} = 1, \ldots, X_N = 1, X_{N+1} = 1\Big)$$

and

$$\begin{aligned} p_{k+1}^{N+1} &= \mathbb{P}\Big(X_1 = -1, \ldots, X_k = -1, X_{k+1} = -1, X_{k+2} = 1 \ldots, X_N = 1, X_{N+1} = 1\Big) \\ &= \mathbb{P}\Big(X_1 = -1, \ldots, X_k = -1, X_{k+1} = 1, X_{k+2} = 1 \ldots, X_N = 1, X_{N+1} = -1\Big) \end{aligned}$$

hence

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

It follows that the numbers p_k^{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 $p_k^{N+1} = p_{k-1}^N - p_{k-1}^{N+1}$ 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

$$\begin{aligned} \mathbb{E}(X_1 \cdot \cdots \cdot X_{N+1}) &= \sum_{k=0}^{N+1} (-1)^k \binom{N+1}{k} p_k^{N+1} \\ &= \sum_{k=0}^{N} A_k^N + \sum_{k=0}^{N+1} \binom{N+1}{k} p_0^{N+1} \\ &= \sum_{k=0}^{N} A_k^N + 2^{N+1} p_0^{N+1} \end{aligned} \tag{154}$$

Note that A_k^N depends only on the p_k^N, hence, by induction hypothesis, A_k^N can be computed by $\mathbb{E}(X_1 \cdot \cdots \cdot X_L)$ 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). $\square$

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

5 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 $\big($i. e. with $\mathbb{P}_0(X_i = 1) = \mathbb{P}_0(X_i = -1) = \frac{1}{2}\big)$ is given by:

$$\mathbb{P}_0\Big(X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N\Big) \;=\; \frac{1}{2^N} \tag{155}$$

where we chose to call the probability measures $\mathbb{P}_0$ rather than the generic $\mathbb{P}$ to distinguish it from the measure $\mathbb{P}_\beta$ 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

$$H \;=\; H(x_1, x_2, \ldots, x_N)\,. \tag{156}$$

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 β) means that the fluctuations of the system are small, while high temperature (small β) 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

$$\begin{aligned}
&\mathbb{P}_\beta\Big(X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N\Big)\\
&= \; Z^{-1}\, e^{-\,\beta\, H(x_1, x_2, \ldots, x_N)}\; \mathbb{P}_0\Big(X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N\Big)\\
&= \; Z^{-1}\, \frac{1}{2^N}\; e^{-\beta H(x_1, x_2, \ldots, x_N)}
\end{aligned} \tag{157}$$

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:

$$\mathbb{P}_\beta = Z^{-1} e^{-\beta H} \mathbb{P}_0 \tag{158}$$

i. e. the measure $\mathbb{P}_\beta$ has the density $Z^{-1} e^{-\beta H(\cdot)}$ with respect to $\mathbb{P}_0$.

Although $Z = Z_{\beta N}$ obviously depends both on β 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 $\mathbb{P} = Z^{-1} e^{-\beta H} \mathbb{P}_0$ we implicitly agree that Z is the summation constant which make $\mathbb{P}$ 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*

$$\mathbb{P}_\beta = Z^{-1} e^{-\beta H} \mathbb{P}_0 \qquad \textit{and} \qquad \widetilde{\mathbb{P}}_\beta = \widetilde{Z}^{-1} e^{-\beta(H+C)} \mathbb{P}_0$$

then $\mathbb{P}_\beta = \tilde{\mathbb{P}}_\beta$

Proof: We have

$$\begin{aligned} Z &= \sum_{x_1, x_2, \ldots, x_N \in \{-1,+1\}} e^{-\beta H(x_1, x_2, \ldots, x_N)} \qquad \text{and} \\ \widetilde{Z} &= \sum_{x_1, x_2, \ldots, x_N \in \{-1,+1\}} e^{-\beta (H(x_1, x_2, \ldots, x_N) - C)} \\ &= Z e^{\beta C} \end{aligned}$$

Hence

$$\widetilde{\mathbb{P}}_\beta = \widetilde{Z}^{-1} e^{-\beta(H+C)} \mathbb{P}_0 = Z^{-1} e^{\beta C} e^{\beta H} e^{\beta C} \mathbb{P}_0 = \mathbb{P}_\beta$$

□

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

$$\tilde{H}(x_1, x_2, \ldots, x_N \;=\; -\frac{1}{2}\sum_{i=1}^{N}\Big(x_i \;\cdot\; \frac{1}{N}\sum_{\substack{j=1\\ j\neq i}}^{N} x_j\Big) \tag{159}$$

In this expression the magnetization x_i of the spin i correlates ('interacts') with the mean magnetization $\frac{1}{N}\sum x_j$ of all the other spins. 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

$$\tilde{H}(x_1, \ldots, x_N) \;=\; -\frac{1}{2}\,\frac{1}{N}\sum_{\substack{i,j=1\\ i\neq j}}^{N} x_i\, x_j$$

$$
\begin{aligned}
&= -\frac{1}{2N} \sum_{i,j=1}^{N} x_i\, x_j \;+\; \frac{1}{2\,N} \sum_{i=1}^{N} {x_i}^2 \\
&= -\frac{1}{2N} \sum_{i,j=1}^{N} x_i\, x_j \;+\; \frac{1}{2} \\
&= -\frac{1}{2N} \Big(\sum_{i=1}^{N} x_i \Big)^2 \;+\; \frac{1}{2}
\end{aligned}
$$

So, if we set

$$
H(x_1, \ldots, x_N) \;=\; -\frac{1}{2N} \Big(\sum_{i=1}^{N} x_i \Big)^2 \tag{160}
$$

we may write according to Lemma 5.1

$$
\mathbb{P}_\beta \;=\; Z^{-1}\, e^{\beta H} \tag{161}
$$

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 $\mathbb{P}$ associated to the infinite sequence $\{X_i\}_{i\in\mathbb{N}}$ on the infinite dimensional space $\mathbb{R}^{\mathbb{N}}$ and the measures corresponding to the finite sequences $X_1, \ldots, X_N$ are just the projections of $\mathbb{P}$ 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 $\mathbb{P}_\beta$. We have to admit now that this is an abuse of notation as 'the' measure $\mathbb{P}_\beta = \mathbb{P}_\beta^{(N)}$ 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 $\mathbb{P}_\beta^{(N)}$, then it is *not* true that the subsequence $X_1, \ldots, X_M$ for $M < N$ is distributed according to $\mathbb{P}_\beta^{(M)}$.

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 $X_1^{(N)}, X_2^{(N)}, \ldots, X_N^{(N)}$ instead of $X_1, X_2, \ldots, X_N$ and $\mathbb{P}_\beta^{(N)}$ 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* $\mathbb{P}_\beta^{(N)}$ with inverse temperature β on $\{-1,+1\}^N$ by

$$\mathbb{P}_\beta^{(N)}\Big(x_1, x_2, \ldots, x_N\Big) \;=\; \frac{1}{Z_\beta^{(N)}}\, e^{\frac{\beta}{2N}(\sum_{i=1}^N x_i)^2} \tag{162}$$

where

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

The Curie-Weiss model (for given N and β) consists of $\mathbb{P}_\beta^{(N)}$-distributed random variables $X_1^{(N)}, X_2^{(N)}, \ldots, X_N^{(N)}$.

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 β 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*

$$e^{\frac{a^2}{2}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{s^2}{2}+sa} ds \tag{163}$$

Proof:

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

□

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

$$\mathbb{P}_\beta^{(N)}(x_1,\ldots,x_N) \;=\; Z^{-1} \int_{-1}^{1} P_t^{(N)}(x_1,\ldots,x_N) \;\; \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2} \;dt \tag{164}$$

where

$$F_\beta(t) \;=\; \frac{1}{\beta} \left(\frac{1}{2} \ln \frac{1+t}{1-t} \right)^2 + \; \ln\left(1-t^2\right) \tag{165}$$

and

$$Z \;=\; \int_{-1}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2} \;dt\,. \tag{166}$$

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

$$\mu(dt) \;=\; \frac{1}{Z}\, \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2} \;dt \qquad \text{on} \quad [-1,1].$$

Proof: Take $(x_1, \ldots, x_N) \in \{-1, +1\}^N$. Using Lemma 5.5 we write

$$e^{\frac{\beta}{2N}(\sum x_i)^2} = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{+\infty} e^{-\frac{s^2}{2}} e^{\sqrt{\frac{\beta}{N}}(\sum x_i)s} \; ds$$
$$= \Big(\frac{N}{2\pi\beta}\Big)^{\frac{1}{2}} \int_{-\infty}^{+\infty} e^{-\frac{N}{2\beta}y^2} \prod_{i=1}^{N} e^{x_i y} \; dy$$

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

$$= \Big(\frac{N}{2\pi\beta}\Big)^{\frac{1}{2}} \int_{-\infty}^{\infty} e^{-N(\frac{1}{2\beta}y^2 - \ln\cosh y)} \prod_{i=1}^{N} \frac{e^{x_i y}}{\cosh y} \; dy$$

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

$$y = \tanh^{-1} t = \frac{1}{2}\ln(\frac{1+t}{1-t})$$
$$\frac{dt}{dy} = (1 - \tanh^2 y)$$
$$\ln\cosh y = -\frac{1}{2}\ln(1 - \tanh^2 y)$$

$$\frac{e^y}{2\cosh y} = \frac{1}{2}(1+t) \qquad \text{and} \qquad \frac{e^{-y}}{2\cosh y} = \frac{1}{2}(1-t)$$

Thus we obtain

$$\int e^{-N(\frac{1}{2\beta}y^2 - \ln\cosh y)} \prod_{i=1}^{N} \frac{e^{x_i y}}{2\cosh y} \; dy$$
$$= \int e^{-\frac{N}{2}(\frac{1}{\beta}(\frac{1}{2}\ln\frac{1+t}{1-t})^2 + \ln(1-t^2))} \Big(\prod_{i=1}^{N} \frac{1}{2}(1 + x_i t)\Big) \cdot \frac{1}{1-t^2} \; dt$$
$$= \int P_t^{(N)}(x_1, \ldots, x_N) \; \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2} \; dt \tag{167}$$

With the correct normalizing constant Z this proves the assertion. $\square$

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.
$$\frac{e^x}{2\cosh x} = \frac{1}{2}(1+\tanh x) \qquad \frac{e^{-x}}{2\cosh x} = \frac{1}{2}(1-\tanh x)$$

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 $\mathbb{P}_\beta^{(N)}$-distributed Curie-Weiss random variables, then*

$$\mathbb{E}_\beta^N \big(X_1 \, X_2 \, \ldots \, X_K\big) \;=\; Z^{-1} \int_{-1}^{1} t^K \; \frac{e^{-\frac{N}{2} F_\beta(t)}}{1-t^2} \; dt \tag{168}$$

where again

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

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

Proof: From Theorem 5.6 we know that

$$\begin{aligned}
\mathbb{E}_\beta^N \big(X_1 \, X_2 \, \ldots \, X_K\big) \;&=\; Z^{-1} \int_{-1}^{1} E_t^{(N)}(X_1, \ldots, X_N) \; \frac{e^{-\frac{N}{2} F_\beta(t)}}{1-t^2} \; dt \\
&=\; Z^{-1} \int_{-1}^{1} t^K \; \frac{e^{-\frac{N}{2} F_\beta(t)}}{1-t^2} \; dt
\end{aligned}$$

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

5.3 Laplace's method

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

$$\int e^{-\frac{N}{2}\,F(x)}\,\phi(x)\,dx \tag{169}$$

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

$$\begin{aligned}
\int_0^\infty e^{-\frac{N}{2}\,F(x)}\,dx \;&\approx\; \int_0^\infty e^{-\frac{N}{2}\,F''(x_0)\,x^2}\,dx \\
&=\; \frac{1}{\sqrt{N\,F''(0)}} \int_0^\infty e^{y^2/2}\,dy \\
&=\; \frac{1}{2}\sqrt{\frac{2\pi}{F''(0)}}\;\frac{1}{\sqrt{N}}\,.
\end{aligned}$$

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 $\delta_0 > 0, k > 0$ and $\eta > 0$ such that for all x with $a \le x \le a + \delta_0$*
$$\left|\big(F(x) - F(a)\big) - A(x-a)^k\right| \le C(x-a)^{k+\eta} \tag{171}$$
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*
$$|\phi(x)| + |\phi_N(x)| \le \tilde{\phi}(x) \tag{173}$$
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 $N \ge N_0$ and an $\ell \ge 0$.

Then we have:

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

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

$$= \phi(a) \; \frac{2^{\frac{\ell+1}{k}}}{k} \; \Gamma\Big(\frac{\ell+1}{k}\Big) \tag{176}$$

For the reader's convenience we recall

Definition 5.11. For $x > 0$ we define the *Gamma function* Γ by

$$\Gamma(x) \ := \ \int_0^\infty t^{x-1}\, e^{-t}\, dt \tag{177}$$

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

We have

$$\int_0^b e^{-\frac{N}{2}F(x)} x^\ell \phi_N(x) dx = \int_0^\delta e^{-\frac{N}{2}F(x)} x^\ell \phi_N(x) dx + \int_\delta^b e^{-\frac{N}{2}F(x)} x^\ell \phi_N(x) dx \tag{178}$$

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 $x > \delta$

$$\int_\delta^b e^{-\frac{N}{2}F(x)} x^\ell \phi_N(x) dx \leq e^{-\frac{N-N_0}{2}B} \int e^{-\frac{N_0}{2}F(x)} x^\ell \tilde{\phi}(x) dx\,. \tag{179}$$

By Assumption 4 this term goes to zero, even if multiplied by $N^{\frac{\ell+1}{k}}$.

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

We have

$$\int_0^\delta e^{-\frac{N}{2}F(x)} x^\ell \phi_N(x)\ dx = \int_0^\delta e^{-\frac{N}{2}(Ax^k + r(x))} x^\ell \phi_N(x)\ dx$$

$$= (NA)^{-\frac{\ell+1}{k}} \int_0^{\delta (NA)^{\frac{1}{k}}} e^{-\frac{y^k}{2}}\ e^{-\frac{N}{2}\ r(\frac{y}{(NA)^{1/k}})}\ y^\ell\ \phi_N(\frac{y}{(NA)^{\frac{1}{k}}}) \quad dy \tag{180}$$

where we put $y = (NA)^{\frac{1}{k}}\, x$.
Since $\phi_N \to \phi$ uniformly in a neighborhood of 0 and ϕ is continuous there we conclude that

$$\phi_N\Big(\frac{y}{(NA)^{\frac{1}{k}}}\Big) \to \phi(0) \quad \text{as} \quad N \to \infty\,.$$

Moreover, since

$$r\left(\frac{y}{(NA)^{\frac{1}{k}}}\right) \leq C\frac{y^{k+\eta}}{(NA)^{\frac{k+\eta}{k}}} \tag{181}$$

We learn that

$$N\,r(\frac{y}{(NA)^{\frac{1}{k}}}) \to 0$$

thus

$$e^{-\frac{N}{2}r\left(\frac{y}{(NA)^{1/k}}\right)} \to 1 \qquad \text{for each } y.$$

Consequently, the integral

$$\int\limits_0^{\delta(NA)^{\frac{1}{k}}} e^{-\frac{y^k}{2}} \quad e^{-\frac{N}{2}\,r\left(\frac{y}{(NA)^{\frac{1}{k}}}\right)}\; y^\ell\; \phi_N(\frac{y}{(NA)^{\frac{1}{k}}})\;dy$$

$$= \int\limits_0^{\infty} e^{-\frac{y^k}{2}}\; \chi_{[0,\delta(NA)^{\frac{1}{k}}]}(y) \quad e^{-\frac{N}{2}r\left(\frac{y}{(NA)^{\frac{1}{k}}}\right)}\; y^\ell\; \phi_N(\frac{y}{(NA)^{\frac{1}{k}}})\;dy \tag{182}$$

converges to $\phi(0)\int\limits_0^\infty e^{-\frac{y^k}{2}}y^\ell dy$ *provided* we can show that the integrand in (182) is dominated by an integrable function.

Due to the uniform convergence of ϕ_N and the continuity (at 0) of ϕ the term $\phi_N(\frac{y}{(NA)^{\frac{1}{k}}})$ is uniformly bounded for $y \in [0, \delta(NA)^{\frac{1}{k}}]$, if δ is small enough.

Again for $y < \delta(NA)^{\frac{1}{k}}$ we estimate using Assumption 2 and (181)

$$N\left|r(\frac{y}{(NA)^{\frac{1}{k}}})\right| \leq \frac{CN}{A^{\frac{k+\eta}{k}}N^{\frac{k+\eta}{k}}}y^\eta y^k \leq \frac{C}{A^{\frac{k+\eta}{k}}N^{\frac{1}{k}}}\delta^\eta N^{\frac{\eta}{k}}A^{\frac{\eta}{k}}y^k \leq \frac{1}{4}y^k \tag{183}$$

if δ is small enough.

Thus we have shown that for suitable $\delta > 0$ the integrand in (182) is dominated by the integrable function

$$Dy^\ell e^{-\frac{y^k}{4}}$$

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 $a\in(c,d)$:*

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

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

Then

1. *For even ℓ*

$$\int_c^d e^{-\frac{N}{2}F(x)}\,(x-a)^{\ell}\,\phi(x)\,dx$$

$$\approx\ \phi(a)\,\Big(\frac{k!}{F^{(k)}(a)}\Big)^{\frac{\ell+1}{k}}\,\Big(\frac{2}{N}\Big)^{\frac{\ell+1}{k}}\,\frac{2}{k}\,\Gamma(\frac{\ell+1}{k}) \qquad \textit{as} \quad N\to\infty \tag{185}$$

2. *For odd ℓ*

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

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

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

as a shorthand notation for

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

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

By Taylor's formula we have

$$F(x) \;=\; \frac{F^{(k)}(a)}{k!}(x-a)^k + r(x-a)$$

$$\text{with} \qquad |r(x)| \;\leq\; C\,|x|^{k+1} \qquad \text{near } a$$

since $F^{(k+1)}$ exists and is continuous.

Thus we may apply Theorem 5.10 to both the integrals

$$\int_a^d e^{-\frac{N}{2}F(x)}(x-a)^\ell \phi(x)dx \qquad \text{and} \qquad \int_c^a e^{-\frac{N}{2}F(x)}(x-a)^\ell \phi(x)dx$$

For even ℓ we have

$$\int_c^a e^{-\frac{N}{2}F(x)}(x-a)^\ell \phi(x)dx = \int_{-a}^{-c} e^{-\frac{N}{2}F(-x)}\big(x-(-a)\big)^\ell \phi(-x)dx$$

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

If ℓ is even then

$$\int_c^a e^{-\frac{N}{2}F(x)}(x-a)^\ell \phi(x)dx = -\int_{-a}^{-c} e^{-\frac{N}{2}F(-x)}\big(x-(-a)\big)^\ell \phi(-x)dx$$

and Theorem 5.10 implies part (ii). □

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

$$\Gamma(x+1) = \int_0^\infty t^x\, e^{-t}dt = x\int_0^\infty t^{x-1}e^{-t}dt \tag{189}$$

This proves 1.

We prove 2. by induction

$$\Gamma(1) \;=\; \int_0^\infty e^{-t}dt \;=\; 1 \;=\; 0! \tag{190}$$

For the induction step we use 1.:

$$\Gamma(N+1) = N \cdot \Gamma(N) = N \cdot (N-1)! = N!$$

□

Proposition 5.15. *(Stirling Formula)*

$$N! \;\approx\; N^N e^{-N}\sqrt{2\pi N} \qquad \textit{as} \quad N \to \infty\,. \tag{191}$$

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

$$\int_0^\infty e^{-N(s-\ln s)}ds \approx e^{-N}\frac{\sqrt{2}}{N^{\frac{1}{2}}}2\int_0^\infty e^{-s^2}ds$$

We compute

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

This gives the result. □

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

$$\mathbb{E}_{\beta}^{(N)}\big(X_1 \cdot X_2 \cdot \ \dots \ \cdot X_k\big) \tag{192}$$

via Laplace's method.

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

$$F_{\beta}(t) = \frac{1}{\beta}\,\big(\frac{1}{2}\ln\frac{1+t}{1-t}\big)^2 + \ln(1-t^2) \tag{193}$$

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

We show:

Proposition 5.16.

1. *For $\beta < 1$ the function F_β 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_β has a unique minimum at $t = 0$. $F_{\beta}^{'}(0) = F_{\beta}^{''}(0) = F_{\beta}^{'''}(0) = 0$ and $F_{\beta}^{(iv)}(0) = 4 > 0$*

3. *For $\beta > 0$ the function F_β 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 $t_0 = \tanh\beta t_0$. F_β 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_β:

$$F_{\beta}{}'(t) = \frac{1}{1-t^2}\,\left(\frac{1}{\beta}\ln\frac{1+t}{1-t} - 2t\right) \tag{194}$$

$F_{\beta}{}'(t) = 0$ is equivalent to

$$\frac{1}{2}\ln\frac{1+t}{1-t} = \beta t \tag{195}$$

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

$$\tanh\beta t = t \tag{196}$$

This is satisfied for $t = 0$ for all β.

$$F_\beta''(t) = \frac{2}{(1-t^2)^2}\left(\left(\frac{1}{\beta} - \left(1-t^2\right)\right) - 2t\left(\frac{1}{2\,\beta}\,\ln\frac{1+t}{1-t} - t\right) \right) \quad (197)$$

Thus

$$F_\beta''(0) = 2\,\left(\frac{1}{\beta} - 1\right) \quad \begin{cases} > 0 & \text{for } \beta < 1 \\ = 0 & \text{for } \beta = 1 \\ < 0 & \text{for } \beta > 1 \end{cases} \quad (198)$$

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

$$g(0) = f(0) = 0 \quad (199)$$

and

$$g'(t) = \frac{\beta}{\cosh^2(\beta t)} < 1 = f'(t) \quad (200)$$

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

$$\tanh\beta t = t \quad (201)$$

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

$$g(0) = f(0) = 0 \quad (202)$$

$$g'(0) = \beta > 1 = f'(t) \quad (203)$$

$$\text{and} \qquad g(t) \to 1, f(t) \to \infty \text{ as } t \to \infty \quad (204)$$

Thus there exists a strictly positive solution t_0 of

$$\tanh\beta t = t \quad (205)$$

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

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

$$F_\beta''(t_0) = \frac{2}{(1-t_0^2)^2} \left(\frac{1}{\beta} - (1-t_0^2) \right) \tag{206}$$

and

$$\frac{1}{\beta} - (1-t_0^2) = \frac{1}{\beta} - (1 - \tanh^2 \beta t_0) = \frac{1}{\beta} - \frac{1}{\beta} g'(t_0) > 0 \tag{207}$$

The case $\beta = 1$ involves a straight forward but tedious computation. We got to compute $F_1'''(t)$ and $F_1^{(iv)}(t)$. It turns out that $F_1'''(0) = 0$ and $F_1^{iv}(0) = 4$.

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 ℓ is even, then

1. *if $\beta < 1$*

$$\mathbb{E}_\beta^{(N)}\big(X_1 \cdot X_2 \cdot X_\ell\big) \;\approx\; (l-1)!! \left(\frac{\beta}{1-\beta}\right)^{\frac{\ell}{2}} \frac{1}{N^{\frac{\ell}{2}}} \qquad \text{as } N \to \infty \tag{208}$$

2. *if $\beta = 1$ there is a constant c_ℓ such that*

$$\mathbb{E}_\beta^{(N)}\big(X_1 \cdot X_2 \cdot X_\ell\big) \approx\; c_\ell \, \frac{1}{N^{\frac{\ell}{4}}} \tag{209}$$

3. *if $\beta > 1$*

$$\mathbb{E}_\beta^{(N)}\big(X_1 \cdot X_2 \cdot X_\ell\big) \approx\; m(\beta)^\ell \tag{210}$$

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

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

Remark 5.18. In the case $\beta = 1$ the constant c_ℓ is given by $12^{\frac{\ell}{4}} \frac{\Gamma(\frac{\ell+1}{4})}{\Gamma(\frac{1}{4})}$ as the proof will show.

Proof: We set

$$Z_N(\ell) = \int_{-1}^{1} \frac{e^{-\frac{NF_\beta(t)}{2}}}{1-t^2} t^\ell dt \tag{211}$$

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

For $\beta < 1$

$$Z_N(\ell) \approx (\frac{4}{F''_\beta(0)})^{\frac{\ell+1}{2}} \Gamma(\frac{\ell+1}{2}) N^{-\frac{\ell+1}{2}} \tag{212}$$

$$= (\frac{\beta}{1-\beta})^{\frac{\ell+1}{2}} \Gamma(\frac{\ell+1}{2}) N^{-\frac{\ell+1}{2}} 2^{\frac{\ell+1}{2}} \tag{213}$$

Consequently

$$\mathbb{E}_\beta^{(N)}(X_1 \cdot X_2 \cdot X_\ell) = \frac{Z_N(\ell)}{Z_N(0)} \tag{214}$$

$$\approx \frac{\Gamma(\frac{\ell+1}{2})}{\Gamma(\frac{1}{2})} (\frac{\beta}{1-\beta})^{\frac{\ell+1}{2}} N^{-\frac{\ell}{2}} 2^{\frac{\ell}{2}} \tag{215}$$

Using the fact that for even ℓ

$$2^{\frac{\ell}{2}} \frac{\Gamma(\frac{\ell+1}{2})}{\Gamma(\frac{1}{2})} = (\ell-1)!! \tag{216}$$

(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

$$Z_N(\ell) \approx 12^{\frac{\ell+1}{4}} \cdot \frac{1}{2} \cdot \Gamma(\frac{\ell+1}{4}) \cdot N^{-\frac{\ell+1}{4}} \tag{217}$$

This gives the result for $\beta = 1$ with $c_\ell = 12^{\frac{\ell}{4}} \frac{\Gamma(\frac{\ell+1}{4})}{\Gamma(\frac{1}{4})}$. 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*

$$2^{\frac{\ell}{2}} \frac{\Gamma(\frac{\ell+1}{2})}{\Gamma(\frac{1}{2})} = (\ell-1)!! \tag{218}$$

Proof: Set $\ell = 2r$. We do induction over r.

For $r = 1$ we use Lemma 5.14 to prove

$$2\frac{\Gamma(1+\frac{1}{2})}{\Gamma(\frac{1}{2})} = 2\frac{1}{2}\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2})} = 1 = 0! \tag{219}$$

The induction step is similar:

$$2^{r+1}\frac{\Gamma(r+1+\frac{1}{2})}{\Gamma(\frac{1}{2})} = 2(r+\frac{1}{2})2^r\frac{\Gamma(r+\frac{1}{2})}{\Gamma(\frac{1}{2})} \tag{220}$$

$$= (2r+1)(2r-1)!! = (2r+1)!! \tag{221}$$

□

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

In this section we investigate the large-N-behavior of normalized sums

$$m_N = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{222}$$

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 $q\,N$, 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

$$m_N \stackrel{\mathcal{D}}{\Longrightarrow} \delta_0 \tag{223}$$

while for $\beta > 1$ we prove

$$m_N \stackrel{\mathcal{D}}{\Longrightarrow} \frac{1}{2}(\delta_{-m(\beta)} + \delta_{m(\beta)}) \tag{224}$$

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 $\mathbb{P}_\beta^{(N)}$-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*

$$m_N \overset{\mathcal{D}}{\Longrightarrow} \delta_0 \tag{225}$$

2. *If $\beta > 1$ then*

$$m_N \overset{\mathcal{D}}{\Longrightarrow} \frac{1}{2}(\delta_{-m(\beta)} + \delta_{m(\beta)}) \tag{226}$$

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 $m_N \overset{\mathcal{D}}{\Longrightarrow} \frac{1}{2}(\delta_{-m(\beta)} + \delta_{m(\beta)})$ for all β.

Proof: As usual we investigate the moments of m_N:

$$\begin{aligned} \mathbb{E}({m_N}^{\ell}) &= \frac{1}{N^{\ell}}\mathbb{E}\Big(\big(\sum_{i=1}^{N} X_i\big)^{\ell}\Big) \\ &= \frac{1}{N^{\ell}} \sum_{i_1,\cdots,i_\ell=1}^{N} \mathbb{E}\big(X_{i_1}X_{i_2}\cdots X_{i_\ell}\big) \end{aligned} \tag{228}$$

We note that $|\mathbb{E}(X_{i_1}X_{i_2}\cdots X_{i_\ell})| \leq 1$. We split the sum (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_ℓ independent of N).

Thus

$$\frac{1}{N^{\ell}} \sum_{\substack{i_1,\cdots,i_\ell=1 \\ \text{some index occurs twice}}}^{N} \mathbb{E}\big(X_{i_1}X_{i_2}\cdots X_{i_\ell}\big) \;\leq\; C_\ell\,\frac{1}{N} \to 0 \tag{229}$$

Hence asymptotically (as $N \to \infty$)

$$\begin{aligned}
\mathbb{E}({m_N}^{\ell}) &\approx \frac{1}{N^{\ell}} \sum_{\substack{i_1,\cdots,i_\ell=1 \\ \text{no index repeated}}}^{N} \mathbb{E}(X_{i_1}\cdots X_{i_\ell}) \\
&= \frac{1}{N^{\ell}} \sum_{\substack{i_1,\cdots,i_\ell=1 \\ \text{no index repeated}}}^{N} \mathbb{E}(X_1 X_2 \cdots X_\ell) \\
&\approx \frac{1}{N^{\ell}} \quad N^{\ell}\, \mathbb{E}(X_1 X_2 \cdots X_\ell) \\
&\approx \mathbb{E}(X_1 X_2 \cdots X_\ell) \qquad (230)
\end{aligned}$$

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

Remark 5.22. All we needed for part 1 of Theorem 5.20 was that for all ℓ the expectations satify that $\mathbb{E}(X_1 X_2 \cdots X_\ell) \to 0$ (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,

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

converges to a normal distribution as long as $\beta < 1$.
Although $\mathbb{V}(X_i) = \mathbb{E}(X_i^2) = 1$ (in fact $X_i^2 = 1$) the limit distribution of Σ_N is not $\mathcal{N}(0,1)$ but $\mathcal{N}(0, \frac{1}{1-\beta})$.

In fact,

$$\begin{aligned}\mathbb{E}(\Sigma_N^2) &= \frac{1}{N}\sum_{i,j=1}^{N}\mathbb{E}(X_i\,X_j)\\ &= \frac{1}{N}\,\big(N+\sum_{i\neq j}\mathbb{E}(X_i\,X_j)\big)\\ &\approx\; 1+\frac{N(N-1)}{N}\,\frac{\beta}{1-\beta}\,\frac{1}{N}\end{aligned}$$

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

$$\mathbb{V}(\Sigma_N)=\mathbb{E}(\Sigma_N^2)\approx\frac{1}{1-\beta}. \tag{232}$$

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

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

$$\begin{aligned}\mathbb{E}\Big(\big(\frac{1}{N^\alpha}\sum X_i\big)^2\Big) &= \frac{1}{N^{2\,\alpha}}\sum_{i,j=1}^{N}\mathbb{E}(X_i\,X_j)\\ &= \frac{1}{N^{2\,\alpha}}\,\big(N+\sum_{i\neq j}\mathbb{E}(X_i\,X_j)\big)\\ &\approx\; N^{1-2\alpha}+\frac{N(N-1)}{N^{2\,\alpha}}\;c_2\,\frac{1}{N^{1/2}}\end{aligned}$$

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.

$$\Sigma_N^{'}=\frac{1}{N^{\frac{3}{4}}}\sum_{I=1}^{N}X_i \tag{233}$$

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

$$\mathbb{E}(X_{i_1}\cdots X_{i_L}) \tag{234}$$

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 $\underline{i} = (i_1, \cdots, i_L)$ the quantities $\rho_\ell(\underline{i})$ count the number of indices $i_1, \cdots, i_L$ that occur exactly ℓ-times in $\underline{i}$.

Since Curie-Weiss random variables are exchangeable it is clear that

$$\mathbb{E}(X_{i_1} X_{i_2} \cdots X_{i_L}) \tag{235}$$

depends only on $\underline{\boldsymbol{r}} = \underline{\boldsymbol{\rho}}(\underline{i}) = (\rho_1(\underline{i}), \cdots, \rho_L(\underline{i}))$.

We set

$$\mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \;=\; \mathbb{E}(X_{i_1} X_{i_2} \cdots X_{i_L}) \tag{236}$$

if $\underline{\boldsymbol{r}} = \underline{\boldsymbol{\rho}}(\underline{i})$.

Therefore, we can write

$$\mathbb{E}\Big(\Big(\frac{1}{N^{1/2}} \sum_{i=1}^{N} X_i\Big)^L\Big) \;=\; \frac{1}{N^{L/2}} \sum_{\underline{\boldsymbol{r}} \in \Pi} w_L(\underline{\boldsymbol{r}})\, \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \tag{237}$$

Here, as in section 3.2 Π denotes the set of all L-profiles and $w_L(\underline{\boldsymbol{r}})$ counts the number of multi-indices $\underline{i}$ with $\underline{\boldsymbol{\rho}}(\underline{i}) = \underline{\boldsymbol{r}}$ (see Definitions 3.9 and 3.10 for details).

From these definitions we also recall that Π_k^0 denotes the set of profiles $\underline{\boldsymbol{r}}$ with $r_1 = k$ and $r_\ell = 0$ for all $\ell > 2$. Also, Π_k^+ denotes the set of profiles $\underline{\boldsymbol{r}}$ 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 Π_k^0 with $k = 0, 1, \ldots, L$ and the sets Π_k^+.

We saw already in Section 3.3 (cf. Remark 3.21) that the sum over Π_0^+ is negligible in the limit. Independence was not used in the corresponding estimate.

Again as in the independent case, the contribution from the set Π_0^0 is easily seen to be 1.

For independent random variables (with expectation 0) we have $\mathbb{E}(X(\underline{\boldsymbol{r}})) = 0$ if $r_1 > 0$, hence the sets Π_k^0 and Π_k^+ for $k > 0$ do not contribute at all. This is different in the current context. It turns out that the sets Π_k^+ do not contribute to the limit, but the sum over the sets Π_k^0 does. It is here where the correlation estimates of Theorem 5.17 will play a crucial role.

Results and proofs

Theorem 5.23. *Let $X_1, \cdots, X_N$ be $\mathbb{P}_\beta^N$-distributed Curie-Weiss random variable with $\beta < 1$ then*

$$\Sigma_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \frac{1}{1-\beta}) \tag{238}$$

Proof: As above we write:

$$\begin{aligned} \mathbb{E}\Big(\big(\frac{1}{N^{1/2}} \sum_{i=1}^{N} X_i\big)^L \Big) &= \frac{1}{N^{L/2}} \sum_{\underline{\boldsymbol{r}} \in \Pi} w_L(\underline{\boldsymbol{r}})\ \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \\ &= \frac{1}{N^{L/2}} \sum_{k=0}^{L} \sum_{\underline{\boldsymbol{r}} \in \Pi_k^0} w_L(\underline{\boldsymbol{r}})\ \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \qquad (239) \\ &\quad + \frac{1}{N^{L/2}} \sum_{k=0}^{L} \sum_{\underline{\boldsymbol{r}} \in \Pi_k^+} w_L(\underline{\boldsymbol{r}})\ \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \qquad (240) \end{aligned}$$

To discuss the expressions $w_L(\underline{\boldsymbol{r}})\ \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big)$ we start with the case $\underline{\boldsymbol{r}} \in \Pi_k^+$:

Lemma 5.24. *Let $X_1, \cdots, X_N$ be $\mathbb{P}_\beta^N$-distributed Curie-Weiss random variable with $\beta < 1$. Suppose the $L-$ profile $\underline{\boldsymbol{r}}$ belongs to Π_k^+. Then*

1. *There is a constant C_1 independent of N such that*

$$\Big|\mathbb{E}\big(X(\underline{\boldsymbol{r}})\big)\Big| \ \leq\ C_1 \frac{1}{N^{k/2}}\,. \tag{241}$$

2. *The number $w_L(\underline{\boldsymbol{r}})$ of index tuples $\underline{i}$ with profile $\underline{\boldsymbol{r}}$ satisfies*

$$w_L(\underline{\boldsymbol{r}}) \ \leq\ L!\ N^{(L+k-1)/2} \tag{242}$$

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 $\underline{\boldsymbol{r}} \in \Pi_k^+$, we obtain

$$\mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \ = \ \mathbb{E}\big(X_1\, X_2\, \ldots\, X_\ell\big)$$

for some $\ell \geq k$.

From this Theorem 5.17 gives the claim 1.

From Corollary 3.18 we learn

$$\begin{aligned} w_L(\underline{\boldsymbol{r}}) &\leq L!\; N^{\sum_{\ell=1}^L r_\ell} \\ &\leq L!\; N^{k+\frac{1}{2}\sum_{\ell=2}^L \ell\, r_\ell - \frac{1}{2}} \qquad \text{(we used that } r_\ell > 0 \text{ for some } \ell > 2) \\ &\leq L!\; N^{\frac{k}{2}+\frac{1}{2}\sum_{\ell=1}^L \ell\, r_\ell - \frac{1}{2}} \\ &\leq L!\; N^{\frac{1}{2}(L+k)-\frac{1}{2}} \end{aligned}$$

□

Continuing the proof of Theorem 5.23 we observe that for $\underline{\boldsymbol{r}} \in \Pi_k^+$

$$w_L(\underline{\boldsymbol{r}})\, \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \;\leq\; L!\; N^{\frac{1}{2}(L+k)-\frac{1}{2}}\, C_1\, \frac{1}{N^{k/2}} \;\leq\; C_2\, N^{\frac{L}{2}-\frac{1}{2}}$$

Thus

$$\frac{1}{N^{L/2}} \sum_{k=0}^{L} \sum_{\underline{\boldsymbol{r}} \in \Pi_k^+} w_L(\underline{\boldsymbol{r}})\, \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big)$$

converges to zero as $N \to \infty$ (and L is fixed).

We turn to the case $\underline{\boldsymbol{r}} \in \Pi_k^0$. 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

$$\mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \;=\; \mathbb{E}\big(X_1\, X_2\, \ldots\, X_k\big) \;= 0 \tag{243}$$

by Theorem 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 $\underline{\mathbf{r}}$ belongs*

to Π_k^0. Then

1. *For large N*

$$\mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \;\approx\; (k-1)!!\,\big(\frac{\beta}{1-\beta}\big)^{k/2}\, N^{-k/2}. \tag{244}$$

2. *The number $w_L(\underline{\boldsymbol{r}})$ of index tuples $\underline{i}$ with profile $\underline{\mathbf{r}}$ satisfies*

$$w_L(\underline{\boldsymbol{r}}) \;\approx\; \frac{L!}{k!\,(\frac{L-k}{2})!\,2^{(L-k)/2}}\; N^{(L+k)/2} \tag{245}$$

Proof (Lemma 5.25)**:** Since for $\underline{\boldsymbol{r}} \in \Pi_k^0$ we have $\mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \;=\; \mathbb{E}\big(X_1\, X_2\, \dots\, X_k\big)$ 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

$$\begin{aligned}
&\mathbb{E}\Big(\big(\frac{1}{N^{1/2}}\,\sum_{i=1}^{N} X_i\big)^L\Big)\\
&\approx \frac{1}{N^{L/2}}\sum_{\ell=0}^{L/2}\sum_{\underline{\boldsymbol{r}}\in\Pi_{2\ell}^0} w_L(\underline{\boldsymbol{r}})\;\mathbb{E}\big(X(\underline{\boldsymbol{r}})\big)\\
&\approx \frac{1}{N^{L/2}}\sum_{\ell=0}^{L/2}\frac{L!}{(2\ell)!\,(\frac{L}{2}-\ell)!\,2^{L/2-\ell}}\;N^{L/2+\ell}\;(2\ell-1)!!\,\big(\frac{\beta}{1-\beta}\big)^{\ell}\,N^{-\ell}\\
&= \sum_{\ell=0}^{L/2}\frac{L!}{(2\ell)!\,(\frac{L}{2}-\ell)!\,2^{L/2-\ell}}\;(2\ell-1)!!\,\big(\frac{\beta}{1-\beta}\big)^{\ell}
\end{aligned}$$

we use Lemma 2.46 to express $(2\ell-1)!!$:

$$\begin{aligned}
&= \frac{L!}{(L/2)!\,2^{L/2}}\sum_{\ell=0}^{L/2}\frac{(L/2)!}{(L/2-\ell)!\,\ell!}\;\big(\frac{\beta}{1-\beta}\big)^{\ell}\\
&= (L-1)!!\;\big(\frac{2}{1-\beta}\big)^{L/2}
\end{aligned} \tag{246}$$

Thus we have proved that the moments of Σ_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 $X_1,\cdots,X_N$ be $\mathbb{P}_\beta^N$-distributed Curie-Weiss random variable with $\beta = 1$ and denote by μ_1 the measure on $\mathbb{R}$ with density*

$$\phi_1(x) \;=\; 2\,(12)^{-1/4}\,\Gamma(1/4)^{-1}\;e^{-\frac{1}{12}x^4}$$

then

$$\Sigma'_N = \frac{1}{N^{3/4}}\sum_{i=1}^{N} X_i \overset{\mathcal{D}}{\Longrightarrow} \mu_1\,. \tag{247}$$

Remark 5.27. The measure μ_1 is indeed a probability measure and its moments $m_k(\mu_1)$ are given by

$$m_k(\mu_1) = \begin{cases} 12^{\frac{k}{4}} \frac{\Gamma(\frac{k+1}{4})}{\Gamma(\frac{1}{4})}, & \text{for even } k; \\ 0, & \text{for odd } k. \end{cases} \tag{248}$$

Proof (Remark)**:** For even k we have

$$\begin{aligned} \int_{-\infty}^{+\infty} x^k \, e^{-\frac{1}{12}x^4} \, dx &= \frac{1}{2} \, (12)^{\frac{k+1}{4}} \int_0^{\infty} t^{\frac{k-3}{4}} \, e^{-t} \, dt \\ &= \frac{1}{2} \, (12)^{\frac{k+1}{4}} \, \Gamma(\frac{k+1}{4}) \end{aligned}$$

by a change of variable $t = \frac{1}{12}x^4$. □

Proof (Theorem 5.26)**:** For $\underline{\boldsymbol{r}} \in \Pi_k$ we have by Theorem 5.17:

$$\Big| \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \Big| \; \leq \; C_1 \, \frac{1}{N^{k/4}} \, .$$

and

$$w_L(\underline{\boldsymbol{r}}) \; \leq \; C \, N^{\sum_{\ell=1}^{L} r_\ell} \; \leq \; C \, N^{\frac{1}{2}(k+L)}$$

since

$$\begin{aligned} \sum_{\ell=1}^{L} r_\ell &= k + \sum_{\ell=2}^{L} r_\ell \; \leq \; k + \frac{1}{2} \sum_{\ell=2}^{L} 2r_\ell \\ &\leq \frac{1}{2} k + \frac{1}{2} \sum_{\ell=1}^{L} 2r_\ell = \frac{1}{2}(k+L) \end{aligned}$$

Thus

$$\frac{1}{N^{\frac{3}{4}L}} \, w_L(\underline{\boldsymbol{r}}) \, \big| \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \big| \; \leq C' N^{-\frac{1}{4}(L-k)} \, .$$

This term goes to zero unless $k = L$.

We conclude

$$\begin{aligned} & \mathbb{E}\Big(\Big(\frac{1}{N^{3/4}} \, \sum_{i=1}^{N} X_i \Big)^L \Big) \\ &= \frac{1}{N^{\frac{3}{4}L}} \, \sum_{k=0}^{L-1} \sum_{\underline{\boldsymbol{r}} \in \Pi_k} w_L(\underline{\boldsymbol{r}}) \, \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \; + \; \frac{1}{N^{\frac{3}{4}L}} \, \sum_{\underline{\boldsymbol{r}} \in \Pi_L} w_L(\underline{\boldsymbol{r}}) \, \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \\ &\approx \frac{1}{N^{\frac{3}{4}L}} \, \sum_{\underline{\boldsymbol{r}} \in \Pi_L} w_L(\underline{\boldsymbol{r}}) \, \mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \end{aligned}$$

For $\underline{\boldsymbol{r}} \in \Pi_L$ we have

$$w_L(\underline{\boldsymbol{r}}) \;\approx\; N^L \qquad \text{(by Theorem 3.14)}$$

and

$$\mathbb{E}\big(X(\underline{\boldsymbol{r}})\big) \;\approx\; \begin{cases} 12^{\frac{L}{4}} \frac{\Gamma(\frac{L+1}{4})}{\Gamma(\frac{1}{4})} \, \frac{1}{N^{L/4}}, & \text{for even } L; \\ 0, & \text{for odd } L. \end{cases} \qquad \text{(by Theorem 5.17)}$$

From this the theorem follows. □

5.7 Fluctuation of the magnetization for large β

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

$$\tanh(\beta m) = m \tag{249}$$

$m(\beta) > 0$ if $\beta > 1$ ($m(\beta) = 0$ if $\beta \leq 1$).

Consequently for $\beta > 1$, the random variables

$$\frac{1}{\sqrt{N}} \sum (X_i - c) \tag{250}$$

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 $\Sigma_N = \frac{1}{\sqrt{N}} \sum\limits_{i=1}^{N} \big(X_i - m(\beta)\big)$ under the condition that $\frac{1}{N}\sum\limits_{i=1}^{N} X_i > 0$.

This conditional probability is defined as

$$P_\beta\Big(\Sigma_N \in A \;\Big|\; \frac{1}{N}\sum_{i=1}^{N} X_i > 0\Big) = \frac{P_\beta\Big(\Sigma_N \in A \wedge \frac{1}{N}\sum\limits_{i=1}^{N} X_i > 0\Big)}{P_\beta(\frac{1}{N}\sum X_i > 0)} \tag{251}$$

We know that $P_\beta(\frac{1}{N}\sum X_i > 0)$ is strictly positive and, in fact, converges to $\frac{1}{2}$.

We remark that we could as well consider

$$P_\beta\Big(\Sigma_N \in A \mid |\frac{1}{N}\sum_{i=1}^{N} X_i - m(\beta)| < \alpha\Big) \tag{252}$$

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

$$\Sigma_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i - m(\beta)) \chi_{\{\frac{1}{N}\sum\limits_{i=1}^{N} X_i > 0\}} \tag{253}$$

To shorten notation we set

$$P^{+}_{\Sigma_N}(A) = P_\beta\Big(\Sigma_N \in A \mid \frac{1}{N}\sum_{i=1}^{N} X_i > 0\Big) \tag{254}$$

and

$$\chi^{+}_{N} = \chi_{\{\frac{1}{N}\sum\limits_{i=1}^{N} X_i > 0\}} . \tag{255}$$

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 $\Sigma_N = \frac{1}{\sqrt{N}} \sum (X_i - m(\beta))$ conditioned on $\frac{1}{N}\sum\limits_{i=1}^{N} X_i > 0$ converges to a normal distribution $\mathcal{N}(0, \sigma^2)$ with*

$$\sigma^2 = \frac{1 - m(\beta)}{1 - \beta(1 - m(\beta))} \tag{256}$$

Proof: We set

$$\zeta_N(L) = \int_{-1}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1 - t^2}\, E_t\Big(\sum_{N}^{L} \chi^{+}_{N}\Big)\, dt \tag{257}$$

where as usual $F_\beta(t) = \frac{1}{\beta}(\frac{1}{2}\ln(\frac{1-t}{1+t}))^2 + \ln(1 - t^2)$ since

$$\mathbb{E}_\beta\Big({\Sigma_N}^L \mid \frac{1}{N}\sum X_i > 0\Big) = \frac{\zeta_N(L)}{\zeta_N(0)} \tag{258}$$

we have to analyze the large-N-behaviour of $\zeta_N(L)$. For $\beta > 1$ the function F_β 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 rule due to the condition $\frac{1}{N}\sum X_i > 0$. To simplify notation we will write m instead of $m(\beta)$. We will also assume that $F_\beta(m) = F_\beta(-m) = 0$.

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 $E_t\Big(\Sigma_N{}^L \chi_N^+\Big)$.

Our strategy will be to evaluate

$$\xi_N(L) = \int\limits_{\frac{m}{2}}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\; E_t\Big(\Sigma_N{}^L\Big)\; dt \tag{259}$$

instead and prove that the difference to $\zeta_N(L)$ is asymptotically negligible.

First, we decompose $\zeta_N(L)$ in three parts

$$\zeta_N(L) = \int\limits_{-1}^{-\frac{m}{2}} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\; E_t\Big(\Sigma_N{}^L \chi_N^+\Big) \;\; (= \zeta_N^-(L)) \tag{260}$$

$$+ \int\limits_{-\frac{m}{2}}^{\frac{m}{2}} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\; E_t\Big(\Sigma_N{}^L \chi_N^+\Big) \;\; (= \zeta_N^0(L)) \tag{261}$$

$$+ \int\limits_{\frac{m}{2}}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\; E_t\Big(\Sigma_N{}^L \chi_N^+\Big) \;\; (= \zeta_N^+(L)) \tag{262}$$

Now, we prove that (260) and (260) do not contribute asymptotically. □

Lemma 5.29. *For each $M \in \mathbb{N}$ there is a constant $C_M < \infty$ such that*

$$\zeta_N^-(L) = \int\limits_{-1}^{-\frac{1}{2}m} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\; E_t\Big(\Sigma_N{}^L \chi_N^+\Big)\; dt \tag{263}$$

$$\leq \frac{C_M}{N^M} \tag{264}$$

Proof: (Lemma):

By the Cauchy-Schwarz inequality

$$E_t\Big(\Sigma_N{}^L \chi_{\{\frac{1}{N}\sum X_i > 0\}}\Big) \leq E_t\Big(\Sigma_N{}^{2L}\Big)^{\frac{1}{2}}\; P_t\big(\frac{1}{N}\sum X_i > 0\big)^{\frac{1}{2}} \tag{265}$$

Due to Theorem 3.24 we have

$$P_t(\frac{1}{N}\sum X_i > 0) \leq \frac{C_M}{N^{2M}} \tag{266}$$

for all $t \in]-1, \frac{1}{2}m[$.

Thus

$$\zeta_N^-(L) \leq \frac{C}{N^M}\int_{-1}^{0} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\, E_t\Big({\Sigma_N}^{2L}\Big)\, dt \tag{267}$$

$$\leq \frac{C^l}{N^M} \tag{268}$$

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 ζ_N between $\frac{-m(\beta)}{2}$ and $\frac{+m(\beta)}{2}$.

Lemma 5.30. *There is a constant C_L and a $\delta > 0$ such that*

$$\zeta_N^0(L) = \int_{-\frac{m(\beta)}{2}}^{+\frac{m(\beta)}{2}} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\, E_t\Big({\Sigma_N}^{L}\chi_N^+\Big)\, dt \tag{269}$$

satisfies

$$|\zeta_N^0(L)| \leq C_L e^{-N\delta} \tag{270}$$

Proof: Since $F_\beta(t)$ has unique minima at $\pm m(\beta)$ there exists a $\delta > 0$ such that $F_\beta(t) > 2\delta$ for all $t \in [-\frac{m(\beta)}{2}, \frac{m(\beta)}{2}]$.

Consequently

$$\zeta_N^0(L) \leq e^{-N\delta}\int_{-\frac{m(\beta)}{2}}^{\frac{m(\beta)}{2}} \frac{1}{1-t^2}\, dt \sup_{t,N} E_t\Big({\Sigma_N}^{2L}\Big)^{\frac{1}{2}} \tag{271}$$

□

It remains to estimate

$$\zeta_N^+(L) = \int_{\frac{m(\beta)}{2}}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\, E_t\Big({\Sigma_N}^{L}\chi_N^+)\Big)\, dt \tag{272}$$

We write

$$\zeta_N^+(L) = \int_{\frac{m(\beta)}{2}}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\, E_t\Big({\Sigma_N}^L\Big)\, dt \tag{273}$$

$$- \int_{\frac{m(\beta)}{2}}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\, E_t\Big({\Sigma_N}^L(1-\chi_N^+)\Big)\, dt \tag{274}$$

The final expression can be estimated in a similar way as $\zeta_N^-(L)$, since

$$E_t(1-\chi_N^+) = P_t(\{\frac{1}{N}\sum_{i=1}^{N} X_i \le 0\}) \tag{275}$$

which is very small for $t \ge \frac{m(\beta)}{2}$.

Thus it remains to estimate

$$\zeta_N^{++}(L) = \int_{\frac{m(\beta)}{2}}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\, E_t\Big({\Sigma_N}^L\Big)\, dt \tag{276}$$

We have

$$E_t\Big({\Sigma_N}^L\Big) = \frac{1}{N^{\frac{L}{2}}} E_t\Big(\big(\sum_{i=1}^{N}(X_i - m(\beta))\big)^L\Big) \tag{277}$$

setting $Y_i = X_i - m(\beta)$ we see that

$$E_t(Y_{i_1}\cdots Y_{i_L}) = E_t(Y_{j_1}\cdots Y_{j_L}) \tag{278}$$

whenever $\underline{r} = \rho(\underline{i}) = \rho(\underline{j})$ hence the above expressions depend only on $r = \rho(\underline{i})$ so we write

$$E_t\big(Y(\underline{r})\big) = E_t(Y_{i_1}\cdots Y_{i_L}) \tag{279}$$

for $\underline{r} = \rho(\underline{i})$.

Now, we can expand

$$E_t\Big({\Sigma_N}^L\Big) = \frac{1}{N^{\frac{L}{2}}} \sum_{\underline{r}\in\Pi} w_L(\underline{r}) E_t\big(Y(\underline{r})\big) \tag{280}$$

For $k = r_1$ we may write

$$E_t\big(Y(\underline{r})\big) = E_t(Y_1)^k E_t(Y(\underline{r}')) \tag{281}$$

with $r' = (0, r_2, r_3, \cdots, r_L)$.

Lemma 5.31. *For $\underline{r} \in \Pi_k$ we have for even k*

$$\Big| \int_{\frac{m}{2}}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\; E_t(Y(\underline{r}))\; dt \Big| \leq \frac{C_L}{N^{\frac{k}{2}}} \tag{282}$$

and for odd k

$$N^{\frac{k}{2}} \int_{\frac{m}{2}}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\; E_t(Y(\underline{r}))\; dt \to 0 \tag{283}$$

Proof: The Lemma follows from Corollary 5.12. □

So far we proved that

$$\zeta_N(L) \approx \zeta_N^{++}(L) = \int_{\frac{m}{2}}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\; E_t\Big(\Sigma_N{}^L\Big)\; dt \tag{284}$$

$$= \frac{1}{N^{\frac{L}{2}}} \sum_{\underline{r}\in\Pi} w_L(\underline{r}) \int_{\frac{m}{2}}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\; E_t\big(X(r)\big)\; dt \tag{285}$$

We set $\mathcal{E}(\underline{r}) = \int_{\frac{m}{2}}^{1} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2}\; E_t(X(r))\; dt$. If $\underline{r} \in \Pi_k^+$ we have by Lemma 5.31

$$|\mathcal{E}(\underline{r})| \leq C N^{-\frac{k}{2}} \tag{286}$$

and by Corollary 3.18 (see also (93))

$$w_L(\underline{r}) \leq C N^{\frac{k+L}{2}-\frac{1}{2}} \tag{287}$$

Consequently, the terms $\frac{1}{N^{\frac{L}{2}}} w_L(\underline{r})\mathcal{E}(\underline{r})$ go to zero for $\underline{r} \in \Pi_k^+$.

Thus we have

$$\zeta_N(L) \approx \frac{1}{N^{\frac{L}{2}}} \sum_{k=0}^{L} \sum_{\underline{r}\in\Pi_k^0} w_L(\underline{r})\mathcal{E}(\underline{r}) \tag{288}$$

For $\underline{r} \in \Pi_k^0$ we know $k + 2r_2 = L$, so either k and L are both odd or they are both even. If k is odd $N^{\frac{k}{2}}\mathcal{E}(\underline{r}) \to 0$, so $\zeta_N(L) \to 0$ if L is odd.

Finally, we arrive at

$$\zeta_N(2M) \approx \frac{1}{N^M} \sum_{k=0}^{M} \sum_{\underline{r}\in\Pi_{2k}^0} w_N(\underline{r})\mathcal{E}(\underline{r}) \tag{289}$$

Lemma 5.32. *For* $\underline{r} \in \Pi_{2k}^0$ *with* $L = 2M$

$$\frac{\mathcal{E}(\underline{r})}{\zeta_N(0)} \approx (1-m^2)^M \left(\frac{\beta(1-m^2)}{1-\beta(1-m^2)}\right)^k (2k-1)!! \, \frac{1}{N^k} \tag{290}$$

Proof: The function F_β has a unique minimum in $[\frac{m}{2}, 1]$, namely $t = m$ (see the computation in Proposition 5.16).

We computed

$$F_\beta''(m) = \frac{2}{1-m^2} \, \frac{1-\beta(1-m^2)}{\beta(1-m^2)} \;\; (>0) \tag{291}$$

To apply Corollary 5.12 we compute for $\underline{r} \in \Pi_{2k}^0$, so

$$r = (2k, M-k, 0, \cdots, 0) \tag{292}$$

$$E_t(X(\underline{r})) = E_t(X_1)^{2k} \, E_t(X_1^2)^{M-k} \tag{293}$$

$$= (t-m)^{2k}(1-2mt+m^2)^{M-k} \tag{294}$$

Applying Corollary 5.12 we get

$$\mathcal{E}(\underline{r}) \approx (1-m^2)^{M-k}(1-m^2)\Big(\frac{\beta(1-m^2)}{1-\beta(1-m^2)}\Big)^{\frac{2k+1}{2}} \cdot 2^{\frac{2k+1}{2}}\Gamma\Big(\frac{2k+1}{2}\Big) \cdot \frac{1}{N^{\frac{2k+1}{2}}} \tag{295}$$

Since

$$\zeta_N(0) = \mathcal{E}(\underline{0}) \quad (\underline{0} \text{ is the unique element in } \Pi \text{ for } L=0) \tag{296}$$

we obtain

$$\frac{\mathcal{E}(\underline{r})}{\zeta_N(0)} \approx (1-m^2)^M \left(\frac{\beta(1-m^2)}{1-\beta(1-m^2)}\right)^k 2^k \, \frac{\Gamma(\frac{2k+1}{2})}{\Gamma(\frac{1}{2})} \, \frac{1}{N^k} \tag{297}$$

$$= (1-m^2)^M \left(\frac{\beta(1-m^2)}{1-\beta(1-m^2)}\right)^k (2k-1)!! \cdot \frac{1}{N^k} \tag{298}$$

where we used Lemma 5.19. □

To finish the proof of Theorem 5.28 we sum over k:

$$\frac{1}{N^M}\sum_{k=0}^{M}\sum_{\underline{r}\in\Pi_{2k}^0} w_{2M}(\underline{r})\,\frac{\mathcal{E}(\underline{r})}{\mathcal{E}(\underline{0})} \tag{299}$$

$$\approx \frac{1}{N^M}\sum_{k=0}^{M} N^{M+k}\,\frac{(2M)!}{(2k)!(M-k)!\,2^{M-k}}\,(1-m^2)^M \tag{300}$$

$$\cdot\left(\frac{\beta(1-m^2)}{1-\beta(1-m^2)}\right)^k\cdot(2k-1)!!\cdot\frac{1}{N^k} \tag{301}$$

$$=(1-m^2)^M\sum_{k=0}^{M}\frac{M!}{k!(M-k)!}\,\frac{(2M)!\,k!}{M!\,(2k)!\,2^{M-k}} \tag{302}$$

$$\cdot\,(2k-1)!!\cdot\left(\frac{\beta(1-m^2)}{1-\beta(1-m^2)}\right)^k \tag{303}$$

$$=(1-m^2)^M(2M-1)!!\,\sum_{k=0}^{M}\binom{M}{k}\left(\frac{\beta(1-m^2)}{1-\beta(1-m^2)}\right)^k \tag{304}$$

$$=(1-m^2)^{\frac{L}{2}}(L-1)!!\left(\frac{1}{1-\beta(1-m^2)}\right)^{\frac{L}{2}} \tag{305}$$

Index

References

[1] D. Aldous: *Exchangeability and related topics*, pp. 1-198 in: Lecture Notes in Mathematics 117, Springer (1985).

[2] R. Ellis: Entropy, Large Deviations, and Statistical Mechanics, Springer (1985)

[3] B. de Finetti: *Funzione caratteristica di un fenomeno aleatorio*, Atti della R. Academia Nazionale dei Lincei, Serie 6. Memorie, Classe di Scienze Fisiche, Mathematice e Naturale, **4**, 251–299 (1931).

[4] Y. Katznelson: An introduction to Harmonic Analysis, Dover (1968)

[5] W. Kirsch: On Penrose's square root law and beyond, Homo Oeconomicus 24, 357–380, (2007)

[6] W. Kirsch, J. Langner: The fate of the square root law for correlated voting, *in:* Fara, Leech, Salles (Eds.): Voting Power and Procedures, Springer (2014)

[7] A. Klenke: Wahrscheinlichkeitstheorie, Springer (2013)

[8] S. Lang: Real Analysis, Addison-Wesley (1983)

[9] B. Simon: The Statistical Mechanics of Lattice Gases, Princeton University Press (1993)

[10] C. Thompson: Mathematical Statistical Mechanics, Macmillian Company (1972)

[11] W. Wu: Fourier Series and Fejér's Theorem, `https://www.ocf.berkeley.edu/~wwu/articles/fejer_theorem.pdf`