State Space Emergence: A New Formalism

Roozbeh Hajijafargholikhani Pazuki

Submitted for the degree of *Doctor of Philosophy in Mathematics*
and the *Diploma of Imperial College London*

Department of Mathematics

Imperial College London

December 2021

Declaration

I hereby certify that the work presented in this dissertation is the result of my own investigations during the *Doctor of Philosophy in Mathematics* project. Text and results obtained from other sources are referenced and properly acknowledged.

Roozbeh Hajijafargholikhani Pazuki
December 2021

To my *parents*,
to *Elham*,
and in memory of my *cats*

Copyright Declaration

Simplex Sigillum Veri
(Simplicity is the seal of truth)

There was the Door to which I found no Key;

There was the Veil through which I might not see:

Some little talk awhile of Me and Thee

There was — and then no more of Thee and Me.

Edward Fitzgerald (1809 - 1883)
Omar Khayyam (1048 - 1123)

Acknowledgements

I want to express my deep gratitude to my supervisor Henrik Jeldtoft Jensen for his continued support, insight, guidance, and from time to time, conversation and coffee at his warm and welcoming home.

I also want to thank Gunnar Pruessner, who advised and supervised me initially.

A special thanks to my external examiners, Emma McCoy and Keith Hopcraft, for their patient reading of, and helpful comments on, the manuscript, especially on pointing out the immigration process publication and introducing the usage of the Hermite and Laguerre polynomials.

This dissertation is based in part on the previously published articles listed below:

1. Henrik Jeldtoft Jensen et al 2018 *J. Phys. A: Math. Theor.* 51 375002. DOI: 10.1088/1751-8121/aad57b

2. Roozbeh H Pazuki and Henrik Jeldtoft Jensen 2021 *J. Phys. Commun.* 5 095002. DOI: 10.1088/2399-6528/ac1f74

Abstract

This thesis focuses on redefining the notion of emergence to a mathematically tractable concept: emergence in state spaces.

In doing that, we will study the probabilistic measures of state spaces with emergence, how to control their volume growth and the differences between them and typical state spaces. This study will introduce two stylistic models, both intuitively simple and practically helpful.

To provide statistical tools for modelling randomness with similar emerging properties, we will introduce different probability distributions from the first principle and derive their preliminary properties. At the same time, we will see that these results are expressible in closed form, by which we can analytically study the emergence in states. Also, for practical reasons, statistical inference will be revisited for distributions' parameter estimation.

Next, we briefly study systems with emerging properties in state spaces by using information-theoretic measures. Alongside that and inspired by the ideas from this discussion, we will propose a pairing time series that combines certainty and uncertainty. In addition, we prove that the Shannon entropy and the rate entropy are well-defined in various circumstances for infinite pairing time series.

And finally, we show that standard statistical mechanics methods fail to yield thermodynamical quantities for some simplistic models with emerging states. We will propose a mathematical tool rooted in the geometry of emergence states spaces from the first part of the thesis to resolve this problem.

Table of Contents

List of Figures

List of Tables

Glossary

$Ber(X;\rho)$	Bernoulli probability distribution for a binary random variable X with success rate ρ.
$Bin(n_h;n,\rho)$	Binomial probability distribution for observing n_h success amonge n trials with success rate ρ.
C_n	Average aggregated cost of the joint ventures.
$H(\mathcal{X})$	Rate entropy of paring time series.
$H[P]$	Shannon entropy for the probability distribution P.
H_B	Shannon entropy for the Balls model.
H_C	Shannon entropy for the Coins model.
H_L	Shannon entropy for a paring time series with lenght L.
$H_q[P]$	Tsallis entropy for parameter $q > 0$ and $q \neq 1$.
$H_\alpha[P]$	Rényi's entropy for parameter $\alpha > 0$ and $\alpha \neq 1$.
$I(m_n, s_n)$	Large deviation rate function for the C-model.
$I(m_n)$	Large deviation rate function for the B-model.
$I_1(m_n;\epsilon)$	Large deviation rate function for the limiting distribution of the B-model.
$I_2(m_n, s_n;\epsilon)$	Large deviation rate function for the limiting distribution of the C-model.
$I_n(S_1, S_1)$	Mutual information between two systems, each with one element.
$I_n(S_{n-1}, S_1)$	Mutual information between two systems with one and $n-1$ elements.
$I_n(X_1, X_2, \ldots, X_k)$	Interaction information of a system with n elements that partitions into k subsystems.
$I_\rho(x)$	Bernoulli rate function. $I_\rho(x) \equiv x\ln(\frac{x}{\rho}) + (1-x)\ln(\frac{1-x}{1-\rho})$.

$O(f)$	Big-O function. $g = O(f)$, implies $\lim_{x\to\infty} \frac{f(x)}{g(x)} < \infty$.
$W_N^{(s)}$	State space of a system with N elements and s stand-alone states.
$\Lambda_n^{(2)}(r, \rho \vert \alpha_1, \beta_1, \alpha_2, \beta_2)$	Conjugate prior distribution for the C-model.
$\Lambda_n(\alpha, \beta)$	Conjugate prior distribution for the B-model.
Λ_n	State space of addmissible configurations with lenght n.
$\Omega_s(N)$	State space cardinality for stand-alone elements with s states.
ϵ	Limiting distribution parameters. $\epsilon \equiv \lim_{n\to\infty} r/n$.
$\equiv$	Defenition sign.
$\hat{m}_n$	Scaled empirical mean of number of pairs.
$\hat{n}_h$	Empirical mean of number of heads.
$\hat{n}_p$	Empirical mean of number of pairs.
$\lambda_n(\alpha, \beta)$	Normalisation constant of the conjugate prior distribution for the B-model.
$\langle \cdot \rangle$	Expectation of a random variable.
$\lfloor \cdot \rfloor$	Floor function.
$\mathcal{A}_L$	Alphabet set $\{0, 1, 2, \ldots, L\}$ for a pairing time series.
$\mathcal{H}$	Hamiltonian of a system.
$\mathcal{S}'_n$	The set of all possible, distinct configurations for a C-model with n elements.
$\mathcal{S}_n$	The set of all possible, distinct configurations for a B-model with n elements.
$\phi_s(x)$	Extenstion of $\Omega_s(N)$ from $\mathbb{N}$ to $\mathbb{R}^+$, $\phi_s(x) : \mathbb{R}^+ \to \mathbb{R}^+$.
$\tilde{H}_r(p)$	Tilted Shannon entropy for parameter r.
$c_n(r)$	Normalisation constant for a system of n elements and $r \in [0, \infty)$.
f	Specific free energy.
m_n	Scaled number of pairs, $m_n \equiv 2n_p/n$.
n_h	The number of observed heads in an experiment with a system of pairing coins.
n_p	The number of observed pairs in an experiment with a system of paring coins or balls.
p_i	Probability of observing i balls in a pair state.
p_{ij}	Probability of observing coins such that i are in a pair state and j stand-alones are in a head state.

r	Ratio of the probability of observing no pairs to a pair.
s	Specific entropy.
s_n	Scaled number of heads, $s_n \equiv n_h/(n - 2n_p)$.
u	Specific free energy.
LDP	Large Deviation Principle.

Chapter

ONE

Introduction

The concept of *emergence* has been debated for around 150 years [16, 25], and more often than not, it is characterised ambiguously as "the whole is greater than the sum of the parts"[24]. Nevertheless, it seems baffling us in the modern-day without any coherent agreement about its definition. One can sceptically assume that the lack of consensus about the meaning of the emergence is rooted in its subjectivity. Therefore, there is no practical reason to utilise it in our mathematical modelling, let alone to take it as more than philosophical speculation.

However, a resolution of such a dilemma is mathematical clarity in the definition of the emergence. Moreover, the price of it would compromise the generality of the description. In this thesis, we shall pursue this goal by restricting the extent of emergent phenomena that one may categorise as such.

After a brief review of different definitions and perspectives, especially in complex systems science, we shall narrow the meaning of emergence to a concrete mathematical notion. Although our account may not be all-encompassing regarding the vast and diverse conceptions of emergence, its clear mathematical definition enables us to apply it to tractable models of complex systems.

We shall see that this intentional restriction is fruitful. We construct stylistic models that are intuitively straightforward, applicable and have practical consequences on our understanding of the distinction of complex systems from other phenomena. Besides discerning the distinction, we shall construct novel mathematical artefacts from first principles, such as probability distributions. Along the way, we will also discuss the geometry of state spaces with emergence property.

Common state spaces grow exponentially for the number of constituents elements of systems of interest. *i.e.*, for a system with N individual components, $\Omega(N)$ denotes the number of available states accessible to a fully interacting system. Asymptotically, such exponential state space grows as $\Omega(N) \sim O(k^N)$ for a real, positive constant k. Accordingly, deviation from exponential growth is sometimes a signature of complex systems [34, 62, 68].

For example, suppose the interdependence between the components can freeze some of the states and make them inaccessible. In that case, $\Omega(N)$ may grow slower than exponentially – some examples of this case are described in [26]. On the contrary, if new collective states become possible due to the inter-component interaction, $\Omega(N)$ will grow faster than exponentially. The faster than exponential growth as an indicator of complexity was initially proposed and reported in [33].

Some generic simplified models, known as pairing models [33, 53], construct the behaviour of faster than exponential state spaces and have been studied in relating the N dependence of $\Omega(N)$ to generalised entropies in several complexity publications – see [3, 26, 31, 33, 34, 38, 39, 40, 62, 68, 70]. We will revisit the pairing models in detail here.

The direct consequence of an exponential state space manifests itself as the additivity property of macroscopic quantities. We will see how the exponentially growing spaces result in additive quantities such as free energy in disciplines like statistical mechanics. In contrast, non-extensivity of the same quantities is inevitable for faster than exponential spaces. Consequently, the standard statistical mechanics fails to find well-defined quantities like specific free energy, and based on that, applying standard statistical mechanics for complex systems requires attention, or one might say, new techniques.

1.1 Emergence

The concept of *emergence* was first coined by British philosopher George H. Lewes [16, 25] around 150 years ago. In the early 20 century, Emergent Evolutionist philosophers and scientists [43] were discussing:

"*the emergence in terms of a sudden arising of new 'collocations' or 'integrations' with new properties arising on a new 'higher' emergent level out of 'lower' level components.*" [25].

The idea continued in contemporary research, specifically complexity science, including biology, cellular biology, evolutionary biology [1, 13], solid-state physics and statistical mechanics [1, 2, 18, 5], etc.

Ironically, the most cryptic and mystical definition of the emergence is the most well-known statement of it: the *whole* is greater than the *sum* of the *parts*[24]. The ambiguity in this definition is what we try to exclude in this study, as David Chalmers clearly refers to here:

"*The term 'emergence' often causes confusion in science and philosophy, as it is used to express at least two quite different concepts.*" [11].

Or, as John Holland humbly cautions us:

"*It is unlikely that a topic as complicated as emergence will submit meekly to a concise definition, and I have no such definition to offer*"[29].

He again tries to give us a flavour of what he believes in as:

"*The behaviour of the whole is much more complex than the behaviour of the parts*"[29].

Loosely speaking, emergence refers to the properties of an entity that is not observed in or owned by its parts. Perhaps one of the carefully articulated definitions of emergence finds in [52] Timothy O'Connor:

"Property P is an emergent property of an object O iff

(1) P supervenes on properties of the parts of O,

(2) P is not had by any of the object's parts,

(3) P is distinct from any structural property of O, and

(4) P has a direct ("downward") determinative influence on the pattern of behaviour

involving O's parts" [6].

Nevertheless, *reductionism*, as an opposition position to emergence, believes that understanding the fundamental laws accounts for the detailed knowledge of the behaviour of nature. To put it merely as what reductionism conveys, for example, elementary particle physics entails the explanation of solid-state physics, chemistry explains molecular biology, molecular biology maps the details of cell biology, and so on [2]. However, reductionism is not entirely immune from the same ambiguity and hypothesises a debatable claim. Following an argument by Phil Anderson in his celebrated article "More is different", he said:

"*The reductionist hypothesis does not by any means imply a constructionist one: The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe*"[2]. One can argue that the upward explanation of reductionism is a self-imposed assumption that might not be justifiable based on its merits.

Similarly, Mark Bedau pushes the definition further into more fine-grained categories of emergence. He tries to make a sharp distinction between two types of emergence: *weak* and *strong*. In his view, a strong emergence is a form of causal influence that is not irreducible to the micro-properties of its part [6]. For instance, Chalmers suggests *consciousness* is a strongly emergent phenomenon [11] in the light of Bedau's definition. And similarly, as Bedau explains, for a system S composed of micro-states/micro-levels and various macros-states/macro-levels, weak emergence defines such that "macrostate P of S with micro-dynamic D is weakly emergent iff P can be derived from D and S's external conditions but only by simulation"[6]. Here, it is assumed that the number and identity of micro-levels might change in time by the micro-dynamics denoted by D, and simulation corresponds to numerical modelling of the system's evolution.

The ambiguity around the definition of emergence increases as one tries to reconcile ideas from different thinkers and philosophers, and as usual, the mentioned brief definitions are the favourites of the author of this thesis, which can be prone to biases. At some point, it might not be unreasonable to say the number of accounts for emergence is close to the number of thinkers in that matter. However, one thing is clear: a definition with practical purpose will be justified by its merits in its applicability, as we try to develop in this thesis.

Intentionally, we kept the emergence review very brief without details of philosophical debates around it. The reason for omitting them is more practical than philosophical. Accessing a simple but mathematically rigorous definition of a concept turns it into a tractable mathematical practice and a verifiable model for natural phenomena. If history is any guide, from Newton's to modern time, constructing a mathematical model provides a level playing field for scientists to evaluate their understanding of a concept, up to the extent of the boundary of their hypothesis, and leaves philosophical debates for philosophers of science.

Considering this point of view, later in the next section, we shall define what we believe is a precise definition of emergence, not tacitly but explicitly in a mathematical fashion. We do not claim the proposed definition is universal, although it is unambiguous in its stance.

1.2 Emergence in State Space

For experiments or observations whose results involve uncertainty, the set of all possible outcomes is called *sample space*, denoted by W, whereas its subsets are *event sets* [22]. In general, modelling the *randomness* in experiments is instrumental in defining a *probability space* [51, 57], denoted by $(W, \mathcal{F}, P)$, such that $\mathcal{F}$ is the σ-algebra of subsets of the sample space W, and $P : \mathcal{F} \to [0, 1]$ is a probability measure satisfies

(1) $P(W) = 1$,

(2) $P(A_1 \cap A_2) = P(A_1) + P(A_2)$ for disjoint subsets $A_1, A_2 \subseteq W$.

This thesis handles probability spaces defined over *discrete sample spaces*, so we do not use measure-theoretic treatment. For discrete sample spaces, the probability measure P defines as usual, whereas the *volume* of the sample space is the cardinality of the number of its elements – or its points. We will distinguish the volume from the sample space by $\Omega = |W|$.

The sample state set is general enough for modelling the uncertainty in observation outcomes of any abstract object. However, the systems we are interested in are composed of individual elements; each has pre-specified states, and the properties of elements directly depend on their states. Hence, the properties of members supervene in their statistical states.

In the same way, this is true for the properties of an aggregate of the same elements:

the aggregate properties occur as the result of its states, and further, the set of aggregate states, namely the system's *ensemble*, is formed of all possible distinct configurations in experiment outcomes. Consequently, to highlight this fact, we rename the sample space to *state space*[1].

To elaborate, consider a die. The outcome of throwing a die involves observing a number between one to six on its top side. Thus, one can say, the die has six possible states. Meanwhile, when we denote the state space of N dice with $\mathcal{D}^N$, throwing N dice together involves a combination of an N-tuple, say $(X_1, X_2, \ldots, X_N)$, for $X_i \in D = \{1, 2, 3, 4, 5, 6\}$. For such an aggregate of N dice, the possible configurations are all in $\mathcal{D}^N$. In other words, the independently combined states of dice are the points of the aggregate's state space.

For the case of N dice and many other similar aggregates, the state space is decomposable to the states of single entities such that the aggregate state space is a *Cartesian product* of its elements' states. Thus, for instance, the N dice state space is a Cartesian product of N individual sets

$$\mathcal{D}^N = \underbrace{D \times D \times \cdots \times D}_{N}. \tag{1.1}$$

Calling it a *Cartesian product space*, if each individual element has $k \in \mathbb{N}$ distinct states, an aggregate of identical entities has an exponential volume/cardinality equal to k^N. For instance, for N dice, we have

$$|\mathcal{D}^N| = 6^N. \tag{1.2}$$

One can observe that for Cartesian state spaces, the state of individuals is independent of each other. *i.e.* the accessible states to a single element remain unaffected irrespective of being part of an aggregate or the aggregate's size. *e.g.* the state of a die does not depend on the states of others. To emphasise this feature in Cartesian state spaces, we say element states are *aggregate independent*.

We have to stress that independence concerns the possible outcomes of observation and not their probabilities. For example, the probability of observing a six in throwing a die might depend on the others – statistical dependence – while the die states are always in D – aggregate independence. Consequently, since element

[1]State space or *phase space* is more common terminology in statistical mechanics literature. The name phase space usually implies there is a dynamic over the states of the system.

properties depend on their states, they are independent of other elements' states or their properties too.

Despite the commonality of aggregate independence assumption, we argue it is not always the case. As we will see in some examples later, one can easily envisage objects that access more states whenever they are part of a group. Two or more elements together can access more states than any single one for such systems. Nevertheless, we shall see that new states are accessible only to compound elements, and stand-alone elements have their own state space.

Through such a mechanism, the new states emerge and directly result from being part of the aggregate. Therefore, we call them *emergent states*. As a result, properties that depend on these new states are *emergent properties*. To elaborate further, we start by giving two examples and next introduce a prototype model to define the emergent states more rigorously.

1.2.1 Example One: Larvae

The larvae of *Perreyia Flavipes Konow* has been documented in some part of South America since 1899 [60]. The larvae form small, closely packed masses on the ground – see figure (1.1) – and from June to September, these masses of larvae are found crawling over the grass, forming an orderly column approximately 15 cm long and 8 cm wide [60].

Looking at the rolling swarm of larvae[2], one can see the group is composed of moving layers with different speeds such that larvae in each layer move over the others in the underlying layer. For example, when a single layer of larvae moves at speed V – see figure (1.2) – the second one moves at $2V$ relative to the ground. Thus, on average, the speed of each larva is $3V/2$ when considering a complete circle.

Now, let us look at the speed state space of an aggregate of two larvae, as it is plotted in panel (c), figure (1.2). The Cartesian state space of two larvae is the $V \times V$ square. However, when we include the aggregation effect of layers, the $3V/2 \times 3V/2$ square becomes accessible to both larvae together. It is important to emphasise that the emergent state is accessible to the compound elements, not the stand-alone ones. Hence, although a system of two larvae moving alone has the same size, its states are aggregate independent, while the compound larvae states are aggregate dependent.

[2] A short video of a rolling swarm: https://twitter.com/rpazuki/status/1461692222215180294

Figure 1.1: Perreyia Flavipes Konow Larvae.
Image reproduced with permission of the rights holder, Elsevier.

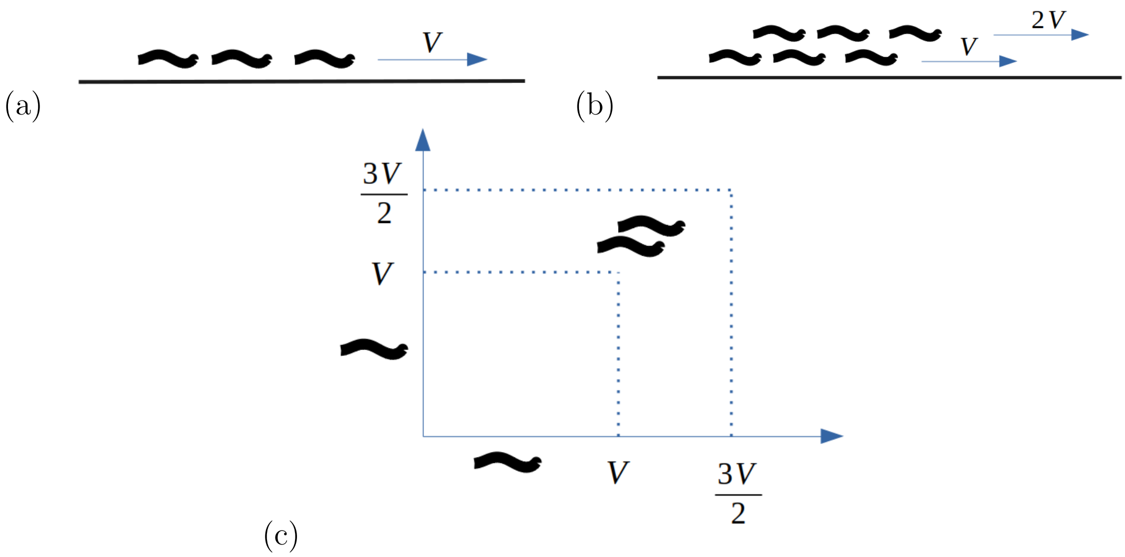


Figure 1.2: (a): A layer of Larvae that moves with speed V. (b): The second layer has a relative speed of $2V$, and following a single larva around one loop, on average, its speed is $3V/2$. (c) The state space of stand-alone and compound systems.

1.2.2 Example Two: Delivery Joint Venture

As a second example, let us assume two delivery companies that established their warehouses in two different locations, say A and B, separated by L. The first company delivers goods from A to B, while the second carries from B to A. Since each delivery vehicle must return to its original warehouse, one trip amounts to driving $2L$. Therefore, if the delivery per L values $C/2$, each trip must cost C, or the cost of two companies that simultaneously operate is $2C$ – see figure (1.3).

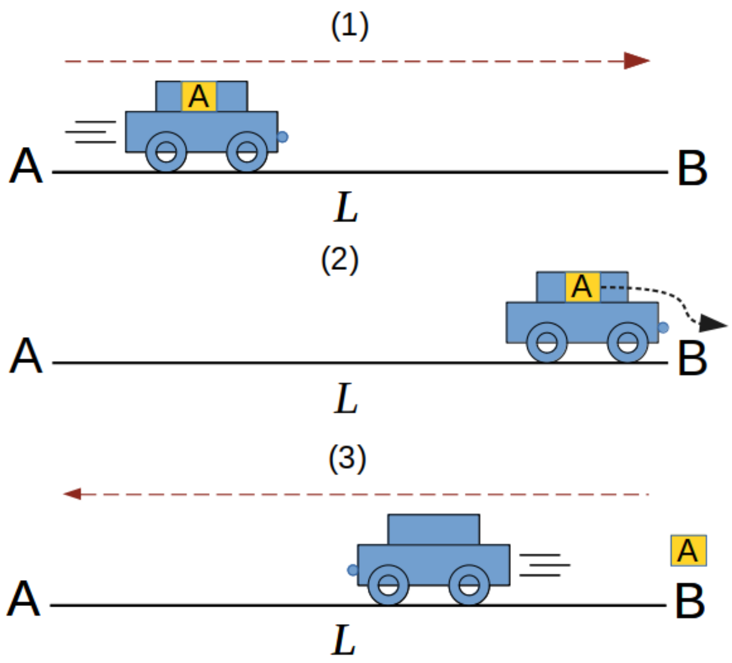


Figure 1.3: (1) Delivery company carries the parcel from A to B, (2) delivers, (3) and returns to its warehouse at A.

However, when both companies agree to set up a joint venture, they can reduce the cost by half through a new emergent state. For example, imagine each delivery vehicle drives $L/2$ of the AB road and exchanges its goods with another vehicle from the other warehouse – see figure (1.4). Thereupon this arrangement, each vehicle drives L, and the total cost of the joint venture reduces to C. Consequently, halving the cost is accessible only to the joint venture and is impossible for stand-alone companies.

The joint venture and its emerging states are cooperative. However, the joint venture can have a spectrum of cooperative/competitive states in a more general setting. So, denoting α as the factor of cost reduction, $\alpha = 1$ amounts to no change in cost, or say, companies act stand-alone, while $\alpha < 1$ corresponds to cooperation, especially $\alpha = 1/2$ is halving the cost like the midway exchange strategy. Similarly, for $\alpha > 1$, the system is in a competitive state, or say, the cost of delivery increases when another delivery company is in the neighbourhood.

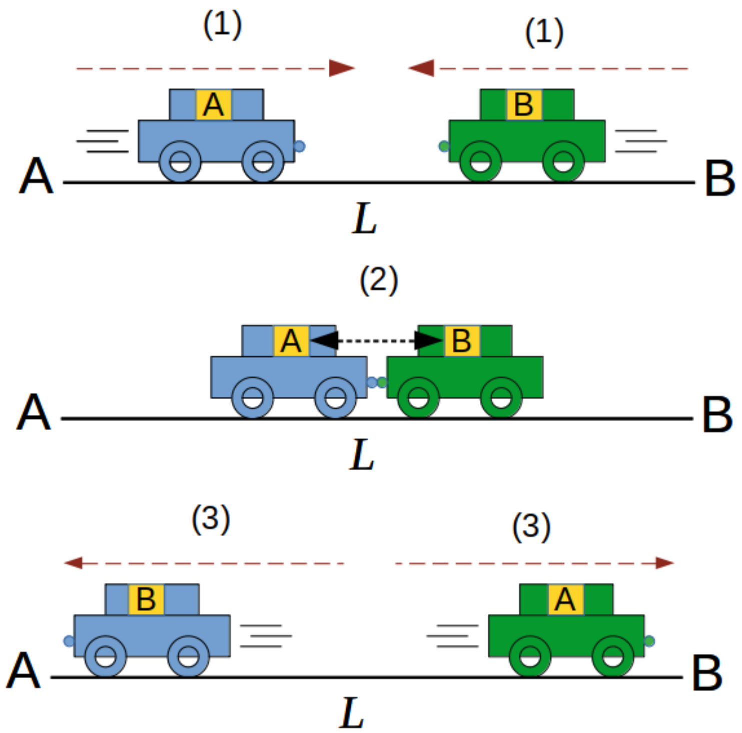


Figure 1.4: (1) Each company's vehicle drives halfway through AB (2) exchanges the parcels, (3) and returns to its warehouse of origin.

Finally, for a system of N companies, let us say at each moment $2n_p$ of them are working together and $n_s = N - 2n_p$ are stand-alone. On average, the cost of the system is

$$C_N = \alpha C \langle 2n_p \rangle + C \langle n_s \rangle \implies C_N = C \left[N + (\alpha - 1) \langle 2n_p \rangle \right], \tag{1.3}$$

where the expectation $\langle . \rangle$ is taken over the ensemble of states – see figure (1.5). In section (5.3), we will derive C_N analytically.

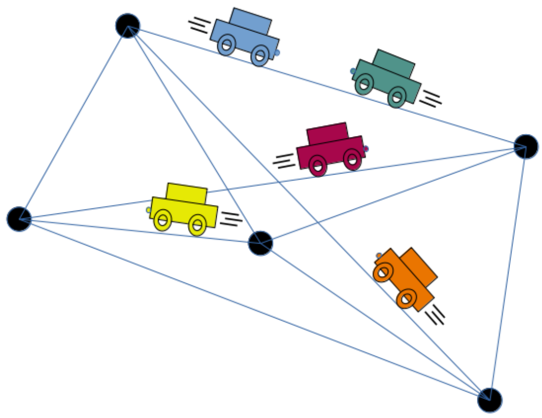


Figure 1.5: A system of N delivery companies, some simultaneously working as a joint venture.

1.2.3 Pairing Models: Introduction

Returning to the discussion of emergence, we will propose a generic model that systematically generates emerging states. Recall that coins are an instrumental tool to form binary variables in statistical modelling. Inspired by them, we propose paring coins that every two can attach and stand upright. One can assume they are magnetised on sides and stick together. The upright state, or the *pair state*, is an emergent one and is not accessible to a single pairing coin or two in stand-alone mode – see figure (1.6).

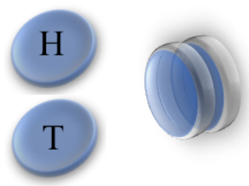


Figure 1.6: Pairing coins in single and pair states.

This model will be discussed in detail in the next chapter, but it is enough to mention that the faster than exponential growth of their state space – due to emergent states – is imposed by a recursive relation. This relation is one way to specify the state space's geometry accurately. We will find it through a combinatorial argument, making the calculation mathematically tractable.

In conclusion, to study the emergence in general, we will restrict our definition to the emergence in states and, further, construct generic models that precisely control the rate of emerging states. In the end, quantities in Cartesian state space have a corresponding counterpart in state spaces that grow faster than exponentially, and the differences between them are due to the emergence of states. For instance, the entropy of binary random variables for ordinary coins has a corresponding pairing entropy. We remark here that studying the pairing entropy and related quantities is the main aim of this research program.

1.3 Additivity and Cartesian Product Spaces

As mentioned in the previous section, we restrict the definition of emergence to state-space emergence to systematically study it. Furthermore, after studying its effect on the volume growth rate, we will investigate the corresponding quantities in faster than exponentially growing state spaces. Logically, the difference between exponential and faster than exponential effects is due to emerging states.

In particular, perhaps the first effect of the emerging states is on the additivity of some properties in that state space. In short, the additivity principle violates in faster than exponentially growing state spaces. To clarify, the aggregate value of an additive quantity is equal to the sum of its parts. Mathematically, denoting the aggregate quantity and its parts by Q_N and Q_i, respectively, it is

$$Q_N = \sum_{i=1}^{N} Q_i, \tag{1.4}$$

and for N identical parts such that $Q_i = Q$, it simply is

$$Q_N = NQ. \tag{1.5}$$

For instance, quantities like volume, energy and number of molecules are *extensive* [10, 23], and the same additive relation governs them. In this brief introduction, to differentiate between additivity and weaker conditions, we need to introduce two more related concepts, namely *extensivity* and *asymptotic extensivity*, since sometimes the distinction between them can confuse, as we read in [35]:

"*Many sources, tacitly if not explicitly, equate additivity and extensivity, often leading to trouble*".

One can say extensivity is a weaker condition, and additivity is sufficient and not necessary for extensivity. Note that extensivity is a general property of a continuous function $Q : \mathbb{R}^k \to \mathbb{R}$. Mathematically, it is defined as [10]

$$Q(\lambda X_1, \ldots, \lambda X_k) = \lambda Q(X_1, \ldots, X_k), \tag{1.6}$$

where $\lambda \in \mathbb{R}^+$ and X_is are intensive parameters of the function[3] Q. Broadly speaking, if a function is extensive, it is additive too, but not vice versa [35, 10]

$$Q_N = Q(NX_1, \ldots, NX_k) = NQ(X_1, \ldots, X_k) = NQ \implies Q_N = NQ. \tag{1.7}$$

In some contexts, a more relaxed version of the extensivity introduces in its asymptotic form, and we call it *asymptotic extensivity* to prevent confusion

$$\lim_{N\to\infty} \frac{Q_N}{N} < \infty. \tag{1.8}$$

Asymptotic extensivity is more or less the same as additivity, although for macroscopic systems only – by macroscopic here, we mean $1 \ll N$. Mind that an extensive function is asymptotic extensive, but the inverse is not necessarily true. In short, for an extensive function Q, we have

$$Q(NX_1, \ldots, NX_k) = NQ(X_1, \ldots, X_k) \implies$$

$$\lim_{N\to\infty} \frac{Q(NX_1, \ldots, NX_k)}{N} = Q(X_1, \ldots, X_k) < \infty. \tag{1.9}$$

Specifically, in the following, we look at the additivity of some quantities in statistical mechanics for exponential state spaces. Recall that the Boltzmann distribution finds the probability of a configuration, say c, with its Hamiltonian $\mathcal{H}(c)$ and inverse temperature β as

$$P(c) = \frac{\mathrm{e}^{-\beta\mathcal{H}(c)}}{Z_N}, \tag{1.10}$$

where Z_N is the normalisation constant or partition function – as it is called in

[3]Or more precisely, Q is a homogenous function degree one.

statistical mechanics literature [49] – and it is defined as

$$Z_N = \sum_{i=1}^{\Omega(N)} \mathrm{e}^{-\beta \mathcal{H}(c_i)}. \tag{1.11}$$

Here again, $\Omega(N)$ denotes the state space volume or the number of distinct configurations. Observe that for $\beta \to 0$, we find

$$\lim_{\beta \to 0} Z_N = \Omega(N). \tag{1.12}$$

So, properties like multiplicity and exponential growth that govern the partition function must also comply with the state space volume. Thus, we will investigate the property of both the emerging state space volume and the normalisation constant of the probability distributions that define over them.

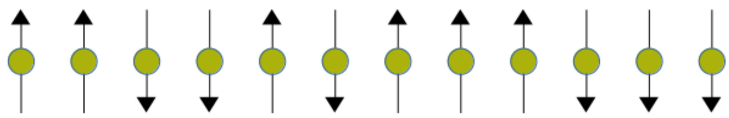


Figure 1.7: One dimensional Ising model, which is composed of up and down spins.

For the purpose of this introduction, we use the Ising model as a prototype of a system with a Cartesian state space [27, 28, 37]. In this model, each element has a spin that orients in an up or down direction, denoting by $s_j = 1$ and $s_j = -1$ respectively. So, a single spin has a binary state space – similar to an ordinary coin – see figure (1.7).

The Hamiltonian of a configuration depends on the orientation of all its spins and the interactions between them

$$\mathcal{H}(c) = J \sum_i \sum_j s_i s_j + B \sum_j s_j, \tag{1.13}$$

for B as external magnetic field and J as neighbour's interaction strength coefficient. Note that, depending on the dimensionality of the model ($1D$, $2D$, $3D$, etc.), the sums run over the lattice that spins are located.

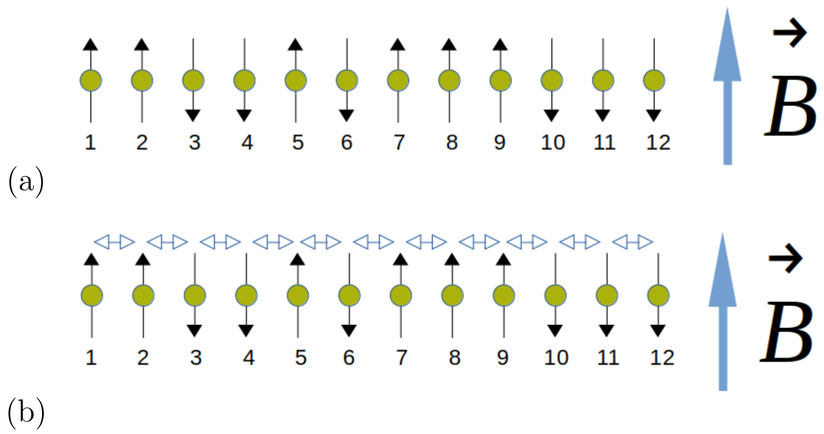


Figure 1.8: One dimensional Ising model (a): spins only interact with an external magnetic field. (b): spins only interact with the external magnetic field and their nearest neighbours.

One can construct different variations of the above Hamiltonian. *e.g.*, for $J = 0$, the neighbour's interaction is zero, and the Hamiltonian only depends on the interaction of spins with the external magnetic field – figure (1.8), panel (a) –

$$\mathcal{H}(c) = B \sum_j s_j, \tag{1.14}$$

Or, including the nearest neighbour interactions – figure (1.8), panel (b) – the Hamiltonian is

$$\mathcal{H}(c) = J \sum_{\langle i,j \rangle} s_i s_j + B \sum_j s_j, \tag{1.15}$$

where the sum is over nearest neighbours, $\langle i, j \rangle$. In particular, including all interactions, as it is in $\sum_i \sum_j$, is known as Curie–Weiss model [8, 14, 65].

By listing the Hamiltonians here, we want to highlight that one always finds the state space is Cartesian. In other words, this is an intrinsic feature of the spins that their states are aggregate independent, regardless of the Hamiltonian. For instance, for N spins, we see $\Omega(N) = 2^N$ for all the mentioned Hamiltonians, and this fact is valid for all d-dimensional Ising models for $d \in \mathbb{N}$.

At the same time, the system's free energy, say F, is proportional to the logarithm of the partition function. As a result, the free energy is always additive in Cartesian state spaces. To see that, since the partition function decomposes as a multiplicative quantity

$$Z_N = Z_1 Z_{N-1}, \tag{1.16}$$

or equivalently its state space volume

$$\Omega(N) = 2^N = \Omega(N-1)\Omega(1), \tag{1.17}$$

therefore, the logarithm of partition function must be additive

$$F_N \equiv \ln Z_N = \ln Z_1 + \ln Z_{N-1} \implies$$

$$F_N = N F_1. \tag{1.18}$$

Later, we will propose probability distributions for systems with emerging states, and these objects live in spaces that have faster than exponential growth. Also, we will construct an emerging Ising model using pairing coins in chapter (5). Further, we will show that the normalisation constants of these probability distributions are not multiplicative, nor is the logarithm of their partition functions additive.

1.4 Parts of the Thesis

In what follows, chapter (2) introduces two pairing models: *the pairing coins and balls*. After that, we will discuss the state space volume and its asymptotic leading term.

Chapter (3) begins with constructing probability distributions over these models from the first principle. Next, the normalisation constant of the distribution and its properties are derived in the same section. After that, we will review the large deviation probabilities of the mentioned distributions, plus two limiting distributions which are resulted from parameter scaling.

Some of the statistics of the mentioned distributions are expressible in closed form. First, section (3.3) addresses the statistics and their asymptotic leading terms in detail, and following that, marginal and joint probability distributions will be discussed in sections (3.4) and (3.5).

It is sometimes necessary to infer the distribution parameters from one or more observed values. We derive the maximum likelihood estimations of the parameters in section (3.6). Finally, section (3.7) makes the statistical inference in a Bayesian setting.

Chapter (4) is about information theory measures. In its first part, we will review the ensemble entropy of distributions from chapter (3). Next, the joint entropy for two or more elements in a system with size n is briefly discussed, and we will see how it relates to the additivity property of the entropy in emerging state space. After that, we shall derive the mutual information of a single element and the rest of the system. Then, it follows the mutual information between two single elements or the interaction entropy of three or more elements in a more general setting. Finally, we shall see two non-extensive entropies derived for the pairing models.

Part two of the chapter (4) introduces a pairing time series. After devising a technique to enumerate their state space, we show that the entropy of the infinitely long pairing time series is well defined under some assumptions.

Chapter (5) is about the application of the pairing model. Its first part introduces a pairing Ising model and shows that free energy is diverging using standard statistical mechanics. The second part briefly reviews the analytic solution of the delivery joint venture introduced in the section (1.2.2).

Finally, the last chapter outlines the open questions and future works that this research program can pursue. We have to remark that the details of the calculations are reported in appendices to separate the results from calculations anywhere that it was possible.

Chapter

TWO

Pairing Models

After restricting the general definition of emergence to states, it demands a concrete mathematical model to investigate the concept further. This section introduces models within which the generic state-space grows faster than exponentially. In what follows, we shall briefly explain how immediately after a new element adds to the system, the model's emergent states accumulate in its state space by design. Furthermore, as the arguments follow, we shall see that states emergence causes faster than exponential growth of the state space volume. Additionally, the mechanism details of the emerging state directly control the growth rate.

We shall introduce two intuitively simple but mathematically rigorous models, namely *pairing balls* and *coins*, each of which constitutes elements with emerging properties. Alongside their simplicity, pairing models are the first step toward modelling state emergence. As we will describe later, both models include pairwise combinations. The emerging properties result from double-element compounding, even though one can easily generalise this machinery to include more than two coalescence elements. However, this report contains the pairing mechanism for mathematical convenience, although we must stress that every step is reproducible for the trio, quartet, quintet mechanisms and more.

Meanwhile, a combinatorial argument finds a *recursive relation* by which the complete description of the state space geometry is provided. Using that recursive equation inside a generating function technique lets us see the state space volume regarding the number of constituent elements. From there, we find the asymptotic leading term of the volume for large system sizes. Hence, it is not unreasonable to say the recursive equation is the core of this chapter.

The importance of the combinatorial argument and its recursive relation becomes evident when one generalises the model to more than two-element compounding: There is a corresponding combinatorial structure and its recursive relation for every number of elements in a compounding mechanism. Nevertheless, we touch on this fact briefly in appendix (E.1) and postpone the matter to future work.

Finally, we will see the asymptotic form of the state space volume grows faster than exponentially. We have to remark that this is the first manifestation of emerging states. Consequently, the logarithm of the state space volume is no more additive.

2.1 Pairing Coins: A Simple Model

Customarily, coins are a suitable prototype to envisage a Bernoulli random variable and its binary states – see figure (2.1). Denoting the configuration of a single coin by c_1 and its state space by W_1, the outcome of throwing a coin would be head (H) or tail (T) such that

$$c_1 \in W_1 = \{H, T\}. \tag{2.1}$$

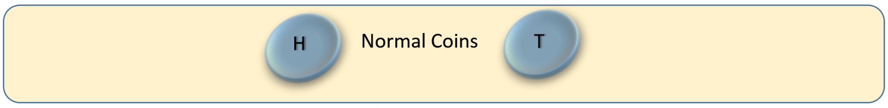


Figure 2.1: A normal coin in two different states.

Again, denoting a configuration of N coins by c_N and its state space by W_N, we must have

$$c_N \in W_N = \{(x_1, x_2, ..., x_N) : x_i = H, T\}, \tag{2.2}$$

whereby the state space is a Cartesian product of single coin state spaces

$$W_N = \{H, T\} \times \cdots \times \{H, T\}. \tag{2.3}$$

Next, suppose a mechanism for aggregating two coins with two possibilities: The first is that a coin behaves as an isolated one and takes either head or tail state. The second possibility is that a coin enters into a pair state with another one as if they were sticking to each other, such that two coins can stand upright. This new state is solely the result of coins aggregation.

Figure (2.2) shows the case for two coins schematically. We see that the first four configurations are the usual combinations for ordinary coins, while the last one is the emergent state.

Of course, one might allow the pair to have internal states such as head against the head or head against tail etc., but for simplicity, we assume the pair state of being structureless and unique.

When the pair state is structureless and unique, the corresponding state space for two pairing coins, denoted by $W_2^{(2)}$, includes the set W_2 and a pair state, P, such that

$$W_2^{(2)} = W_2 \cup \{P\}. \tag{2.4}$$

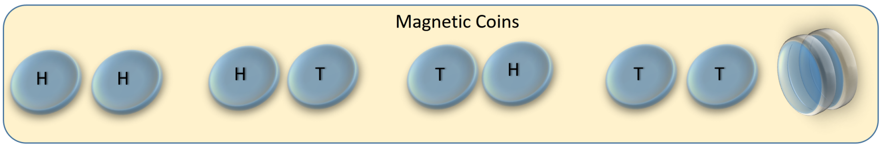
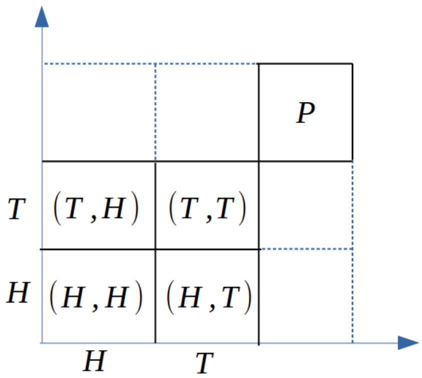
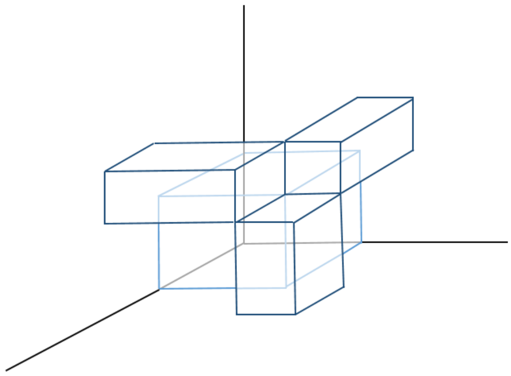


Figure 2.2: Two magnetic coins and their state space.

Let us first elaborate on the notation. Remember that a coin has two possibilities in its stand-alone state. So, in general, for elements with k possible stand-alone states ($k \in \mathbb{N}$), we denote the state space of N pairing coins by $W_N^{(k)}$. Therefore, $W_N^{(2)}$ denotes the state space of N pairing coins.



(a)



(b)

Figure 2.3: Pairing coins state space: (a) The fifth square in the top right is an emergent state. The overall state space is larger than the usual two by two squares of the Cartesian product. (b) For three coins, the inner blue cube is a Cartesian product space. The three rectangular prisms are emergent states that the pairing construct.

Notice that including the mechanism of pairing introduces emergent states compared to standard sets constructed by Cartesian products. Figure (2.3) plots state spaces for systems of two and three pairing coins, respectively, to elaborate on this statement. In panel (a), the first four distinct configurations construct as $W_1 \times W_1$, or the Cartesian product of the set $W_1 = \{H, T\}$ by itself. And the fifth state is an emergent one, resulting from the pairing mechanism. Panel (b) plots the state space for three coins. The inner blue cube is a Cartesian product, or $W_1 \times W_1 \times W_1$. And the three rectangular prisms are emergent states of the pairing mechanism.

Indeed, compared to the subsets constructed by the Cartesian product, the state space volume is generally larger than its inner hypercube. So, in conclusion, the state space grows faster than exponential.

2.1.1 The Pairing Coins State Space Volume

For an aggregate of N pairing coins, $\Omega_2(N)$ denotes the cardinality or volume of the set $W_N^{(2)}$

$$\Omega_2(N) = |W_N^{(2)}|. \tag{2.5}$$

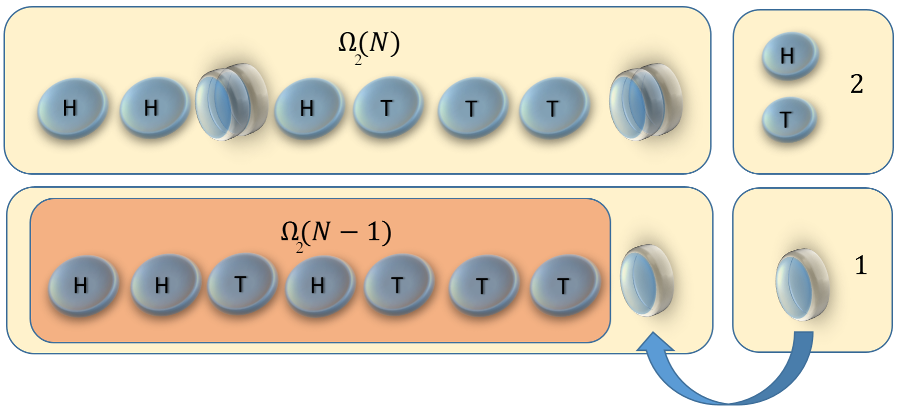


Figure 2.4: The mechanism of adding a new coin to the system.

To determine $\Omega_2(N)$ through a combinatorial argument, we observe that adding a new coin to an aggregate of N existing coins introduces two possibilities:

1. The new coin stays in a stand-alone (non-pair) state. So, it must be in one of the head or tail states. Since the existing N coins have $\Omega_2(N)$ distinct states, there are $2\Omega_2(N)$ available states for the whole system of $N + 1$ coins. For example, see the top row in figure (2.4).

2. The newly added coin pairs with another one in the aggregate. However, the

remaining coins make an aggregate of $N-1$ elements with $\Omega_2(N-1)$ distinct configurations. Moreover, for the new coin, there are N possible choices to make pair. Consequently, there are $N\Omega_2(N-1)$ available states for the whole system of $N+1$ coins. See the bottom row in figure (2.4).

Combining both cases, $\Omega_2(N+1)$ must be the sum of distinct configurations for each possibility

$$\Omega_2(N+1) = 2\Omega_2(N) + N\Omega_2(N-1). \tag{2.6}$$

Equation (2.6) is a recursive relation amongst volumes of state spaces with different sizes. In fact, starting from initial values $\Omega_2(0) = 1$ and $\Omega_2(1) = 2$, it finds volumes of every sizes iteratively. For instance, figure (2.5) shows the first four values schematically. We shall find a closed form for $\Omega_2(N)$ in the following section.

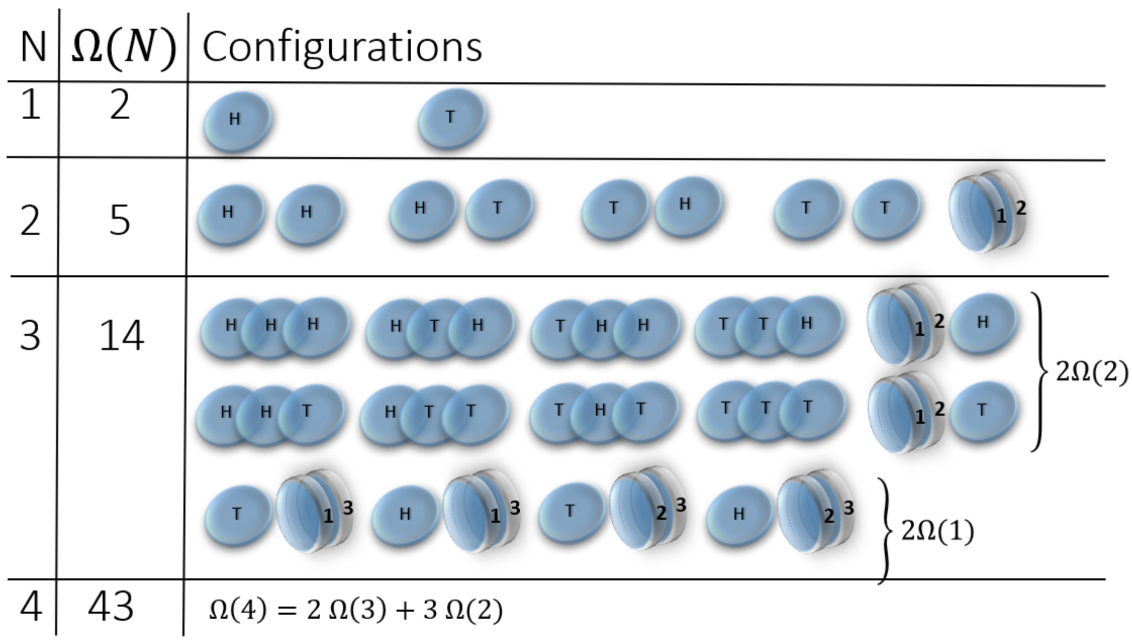


Figure 2.5: Magnetic coins and their state-space.
Image reproduced with permission of the rights holder, IOP Publishing Ltd.

2.2 Pairing Balls Model

The paring coins model (*C-model*) assumes two distinct states for a stand-alone coin – panel (a) in figure 2.6. We introduce a second model whose constituent elements are stateless, like balls in a stand-alone state, to simplify the matter further. We call it the *paring balls model*, or in short, B-model. It is schematically plotted in figure (2.6), panel (b). The B-model has an emerging state similar to the C-model. Nevertheless, it is mathematically convenient to start from the B-model and later move to the C-model.

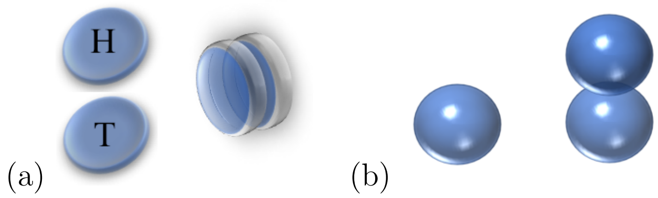


Figure 2.6: (a) Coins in single and pair state (b) Balls in pair and stand-alone state.

Following the same notation, balls have a single state as a stand-alone element or $k = 1$, and therefore, $W_N^{(1)}$ denotes the state space of N pairing balls. Similarly, $\Omega_1(N)$ denotes its state space volume

$$\Omega_1(N) = |W_N^{(1)}|. \tag{2.7}$$

Adding a new ball introduces two possibilities:

1. The newly added ball stays in a stand-alone (non-pair) state. Since the ball is in a single state and the existing N coins have $\Omega_1(N)$ distinct states, in total, there are $\Omega_1(N)$ available states for the whole system of $N + 1$ balls.

2. The new ball makes a pair with another one in the aggregate, and in total, there are $N\Omega_1(N - 1)$ available states for the whole system of $N + 1$ balls.

Therefore, the total number of distinct configurations for $N+1$ pairing balls obtains as

$$\Omega_1(N + 1) = \Omega_1(N) + N\Omega_1(N - 1). \tag{2.8}$$

Notice that the difference between equations (2.6) and (2.8) is in the factor that corresponds to the number of stand-alone element states. Specifically, for $s \in \mathbb{N}$ and $p \in \mathbb{N}$ as the number of states for stand-alone and pair elements, respectively, the general state-space volume complies with the following recursive equation

$$\Omega_s(N+1) = s\Omega_s(N) + pN\Omega_s(N-1). \tag{2.9}$$

For instance, the C-model corresponds to $s = 2$ and $p = 1$. In the next section, we will derive the general form of $\Omega_s(N)$ by using the following initial values

$$\Omega_s(0) = 1, \qquad \Omega_s(1) = s, \tag{2.10}$$

and after that, find $\Omega_1(N)$ and $\Omega_2(N)$.

2.3 Finding $\Omega_s(N)$

By employing a power series generating function [71], we will show how to derive $\Omega_s(n)$. Denoting $G(z)$ as a converging power series like

$$G(z) = \sum_{n\geq 0} a_n \frac{z^n}{n!}, \tag{2.11}$$

for $z \in A \subseteq \mathbb{R}$, we assume $\Omega_s(n)$ is the coefficient of the term $\frac{z^n}{n!}$ in the power series expansion of $G(z)$, or

$$a_n \equiv \Omega_s(n). \tag{2.12}$$

The details of obtaining $G(z)$ are explained in appendix (A.1). We see in equations (A.16) and (A.17) the coefficients of even and odd powers are different, and $G(z)$ derives as

$$G(z) = \sum_{n\geq 0} \left[\mathbf{1}_{\text{odd}}(n) \sum_{k=0}^{\frac{n}{2}} n! \frac{s^{2k}(\frac{p}{2})^{\frac{n}{2}-k}}{2k!(\frac{n}{2}-k)!} + \mathbf{1}_{\text{even}}(n) \sum_{k=0}^{\frac{n-1}{2}} n! \frac{s^{2k+1}(\frac{p}{2})^{\frac{n-1}{2}-k}}{(2k+1)!(\frac{n-1}{2}-k)!} \right] \frac{z^n}{n!}, \tag{2.13}$$

whereas $\mathbf{1}_{\text{odd}}(n)$ and $\mathbf{1}_{\text{even}}(n)$ are indicator functions for odd and even numbers. In equation (A.26), the last result writes $\Omega_s(N)$ as

$$\Omega_s(N) = \sum_{n_p=0}^{\lfloor N/2 \rfloor} \binom{N}{2n_p} (2n_p - 1)!! s^{N-2n_p} p^{n_p}, \tag{2.14}$$

where $\lfloor . \rfloor$ denotes the floor function. Observe that $\binom{N}{2n_p}(2n_p - 1)!!$ is the degeneracy corresponds to n_p pairs among N elements. To elaborate, choosing $2n_p$ elements from N is equal to $\binom{N}{2n_p}$ distinct combinations, and there are $(2n_p - 1)!!$ different *distinguishable* pairs that $2n_p$ selected elements can make. In total, $\binom{N}{2n_p}(2n_p - 1)!!$ is the number of distinct configurations without considering the states of single or pair elements.

For pairs with p state, p^{n_p} enumerates distinct combinations of a given n_p pairs. Similarly, s^{N-2n_p} enumerates combinations for stand-alone elements with s states. Overall, in total, for n_p pairs

$$W_N(n_p) = \binom{N}{2n_p}(2n_p - 1)!!s^{N-2n_p}p^{n_p}, \tag{2.15}$$

is the degeneracy of the set of N elements with n_p pairs.

In appendix (A.2), we used a second method to derive $\Omega_s(N)$ in terms of the renowned generalised Laguerre polynomials, denoted by $L_n^{(\alpha)}(x)$, for degree n and $\alpha = \{-\frac{1}{2}, \frac{1}{2}\}$. Equations (A.36) and (A.37) write $\Omega_s(N)$ as

$$\Omega_s(2N) = N!(2p)^N L_N^{(-\frac{1}{2})}(\frac{-s^2}{2p}), \tag{2.16}$$

$$\Omega_s(2N+1) = sN!(2p)^N L_N^{(\frac{1}{2})}(\frac{-s^2}{2p}). \tag{2.17}$$

2.3.1 B-model's $\Omega_1(N)$

For B-model, using $s = p = 1$ equation (2.14) obtains

$$\Omega_1(N) = \sum_{n_p=0}^{\lfloor N/2 \rfloor} \binom{N}{2n_p}(2n_p - 1)!!\,. \tag{2.18}$$

2.3.2 C-model's $\Omega_2(N)$

For C-model, $s = 2$ and $p = 1$ equation (2.14) yields

$$\Omega_2(N) = \sum_{n_p=0}^{\lfloor N/2 \rfloor} \binom{N}{2n_p}(2n_p - 1)!!\,2^{N-2n_p}. \tag{2.19}$$

2.4 Asymptotic Leading Terms

Generally, macroscopic systems have many constituent elements, namely, $1 \ll N$. Accordingly, for any practical purposes, the leading term of the asymptotic expansion of state-space volume contains all the relevant information. On account of this fact, in this section, we shall derive the asymptotic leading terms of equations (2.18) and (2.19).

We start with a different form of equation (2.14), shown in appendix (A.1), equation (A.23)

$$\Omega_s(N) = N!(\frac{p}{2})^{\lfloor N/2 \rfloor} \sum_{k=0}^{\lfloor N/2 \rfloor} \frac{(\frac{2s^2}{p})^k}{(2k)!(\lfloor \frac{N}{2} \rfloor - k)!}. \tag{2.20}$$

Appendix (A.3) explains the details of the steps. Here, we go through the main results in order.

First, note that we use the exact form for even numbers and start by numerically investigating the properties of the summand in equation (2.20). Since different choices of s and p do not affect the general feature of the summand, the numerical calculation is done for $s = 2$ and $p = 1$. This choice corresponds to $\Omega_2(2N)$ parameters.

Numerical calculation shows that the summand in the following sum has a maximum at $k^* = \lfloor \sqrt{2N} \rfloor$ (or in general $k^* = \lfloor \sqrt{s^2 N/2p} \rfloor$) for even numbers

$$\sum_{0 \le k \le N} \frac{2^{3k}}{(2k)!(N-k)!} = \sum_{0 \le k \le N} t_N(k). \tag{2.21}$$

Next, we obtain the limit of the ratio $t_N(2\sqrt{2N})/t_N(\sqrt{2N})$, and finds its limit approaching zero exponentially fast

$$\lim_{n \to \infty} \frac{t_N(\sqrt{2N})}{t_N(2\sqrt{2N})} = \lim_{n \to \infty} \left(\frac{e}{4}\right)^{-2\sqrt{2N}} e^4 \to 0. \tag{2.22}$$

In other words, other terms around the maximum decrease exponentially fast, and in that sum, the bulk of contributions are from the terms around the maximum.

Using these findings, we can divide the range of the sum into

$$\Omega_s(2N) = \frac{(2N)!}{(2/p)^N} \left(\frac{1}{N!} + \sum_{1 \leq k \leq 2\sqrt{s^2N/2p}} \frac{(\frac{2s^2}{p})^k}{(2k)!(N-k)!} + \Delta \right), \tag{2.23}$$

where $\frac{1}{N!}$ and Δ are exponentially small in comparison to the sum in the middle. We have to highlight the term corresponds to $k = 0$ is $1/N!$. After that, using the Stirling approximation [7, 17]

$$N! = \sqrt{2\pi N}(\frac{N}{\mathrm{e}})^N(1 + O(\frac{1}{N})), \tag{2.24}$$

equation (2.23) simplifies to

$$\Omega_s(N) = \frac{p^N}{\sqrt{2\pi}\mathrm{e}^{\frac{s^2}{4p}}(\frac{s^2N}{2p})^{\frac{1}{4}}} \left(\frac{2N}{\mathrm{e}}\right)^N \sum_{1 \leq k \leq 2\sqrt{s^2N/2p}} \left(\frac{\sqrt{\frac{s^2N}{2p}}\mathrm{e}}{k} \right)^{2k} \left[1 + O(\frac{1}{\sqrt{N}})\right]. \tag{2.25}$$

And finally, the Euler-Maclaurin summation formula [59] derives the leading term for the summation as

$$\Omega_s(2N) = \frac{p^N}{\sqrt{2}\mathrm{e}^{\frac{s^2}{4p}}} \left(\frac{2N}{\mathrm{e}}\right)^N \mathrm{e}^{2\sqrt{s^2N/2p}} \left(1 + O(\frac{1}{\sqrt{N}})\right). \tag{2.26}$$

Equation (2.26) gets the asymptotic leading term for both $\Omega_1(2N)$ and $\Omega_2(2N)$. For the former, $s = p = 1$ and

$$\Omega_1(2N) = \frac{1}{\sqrt{2}\mathrm{e}^{\frac{1}{4}}} \left(\frac{2N}{\mathrm{e}}\right)^N \mathrm{e}^{2\sqrt{N/2}} \left(1 + O(\frac{1}{\sqrt{N}})\right), \tag{2.27}$$

while for the latter, $s = 2$ and $p = 1$

$$\Omega_2(2N) = \frac{1}{\sqrt{2}\mathrm{e}} \left(\frac{2N}{\mathrm{e}}\right)^N \mathrm{e}^{2\sqrt{2N}} \left(1 + O(\frac{1}{\sqrt{N}})\right). \tag{2.28}$$

It is apparent from equation (2.26) that the dependence of state-space volume on N grows faster than exponentially. For instance, keeping the most significant term written as

$$\Omega(N) \sim \left(\frac{N}{\mathrm{e}}\right)^{\frac{\mathrm{e}}{2}\times\frac{N}{\mathrm{e}}} = \tilde{N}^{\gamma\tilde{N}}, \tag{2.29}$$

for $1 \ll N$ and $\tilde{N} = N/\mathrm{e}$, with e as Euler constant and $\gamma = \frac{\mathrm{e}}{2}$ [33].

However, for a typical state space where a single element has k distinct states, the Cartesian multiplications of N sets have a volume equal to k^N. In comparison, the state space volume equal to $N^{\gamma N}$ grows asymptotically faster than exponentially for positive γ. That is to say, $N^{\gamma N}$ is faster than the exponential function, namely k^N, and similar to N^N for factorial state spaces[1].

Indeed, state spaces with faster than factorial growth also exist. For example, the state space growth rate for directed networks with self-loop is 2^{N^2} for an ensemble of networks with N nodes.

2.5 Conclusion

In section (1.3), we reviewed the definition of additivity and its relation to multiplicative quantities such as partition function or the volume of exponential state spaces. Recall that for elements with k states, the volumes of exponential state space with different sizes write as

$$\Omega(N+1) = \Omega(1)\Omega(N). \tag{2.30}$$

Since $\Omega(N)$ is a strictly increasing function and therefore has an inverse function, there exist two continuous conjugate functions as

$$\Omega(x) = k^x \Leftrightarrow \Omega^{-1}(x) = \ln_k x, \tag{2.31}$$

for a positive, real number x. Consequently, after taking the logarithm of the multiplicative relation, it transforms into an additive quantity

$$\Omega(N+1) = \Omega(1)\Omega(N) \implies \ln_k \Omega(N+1) = \ln_k \Omega(1) + \ln_k \Omega(N). \tag{2.32}$$

For example, we observed that when the free energy is proportional to the logarithm of the multiplicative partition function, say Z_N, it complies with the additivity

[1]Note that the factorial state spaces correspond to N^N, where the logarithm of the state space volume is $N \ln N \sim \ln N!$ when we use Stirling's approximation.

relation as

$$Z_{N+1} = Z_1 Z_N, \quad F_N = \ln_k Z_N \implies F_{N+1} = F_1 + F_N. \tag{2.33}$$

On the other hand, for pairing models, the volumes with different sizes govern by equation (2.9) written as

$$\Omega_s(N+1) = \Omega_s(1)\Omega_s(N) + pN\Omega_s(N-1). \tag{2.34}$$

Note that, since $\Omega_s(1) = s$, we rewrite the last equation in terms of $\Omega_s(1)$ to make the first term on the right-hand side similar to the multiplicative form.

Following the same idea, we observe $\Omega_s(N)$ is a strictly increasing function, and it must have an inverse function. Extending the domain of $\Omega_s(N)$ from $\mathbb{N}$ to $\mathbb{R}^+$ and assuming $\Omega_s(N)$ is continuous in $\mathbb{R}^+$, the pairing conjugate functions written as

$$\Omega_s \equiv \phi_s(x) : \mathbb{R}^+ \to \mathbb{R}^+, \quad \Omega_s^{-1} \equiv \phi_s^{-1}(x) : \mathbb{R}^+ \to \mathbb{R}^+, \tag{2.35}$$

and the relation equivalent to additivity obtains as

$$\Omega_s(N+1) = \Omega_s(1)\Omega_s(N) + pN\Omega_s(N-1) \implies$$

$$\phi_s^{-1}\left(\Omega_s(N+1)\right) = \phi_s^{-1}\left(\Omega_s(1)\Omega_s(N) + pN\Omega_s(N-1)\right). \tag{2.36}$$

Although we found the exact form of $\Omega_s(N)$ for discrete values in equation (2.14), its continuous extension, $\phi_s(x)$, and its inverse , $\phi_s^{-1}(x)$, are not expressible in terms of elementary functions. Later, we will return to this idea and find the asymptotic form of $\phi_s(x)$.

Generalising pairing models to more than two compounding elements is straightforward. Evidently, the recursive relation in equation (2.9) describes the geometry of the pairing mechanism. For brevity, without proof, we report the recursive relation for the *n-tet mechanism*. *e.g.* for the trio, quartet and quintet mechanisms, n is equal to three, four and five, respectively. In general, adding a new element amounts to a stand-alone configuration in s distinct states for the newly added element or creating an n-tet in p_n states with $\binom{N}{n}$ distinct combinations. In total, the recursive relation is

$$\Omega_s(N+1) = s\Omega_s(N) + p_n \binom{N}{n} \Omega_s(N-n). \tag{2.37}$$

If all n-tet mechanisms are permitted simultaneously, each with p_i states, we get

$$\Omega_s(N+1) = s\Omega_s(N) + \sum_{i=1}^{n} p_i \binom{N}{i} \Omega_s(N-i). \tag{2.38}$$

We do not pursue this program any further here, but it is possible to reapply the arguments of this chapter and the next for the generalised cases. However, we report some insights in appendix (E.1) containing future work conjectures.

Chapter

THREE

Probability Distributions of Pairing Models

The paring models introduced in the previous chapter are combinatorial objects. In other words, they only enumerate the sets of distinct pairing configurations. However, a probability distribution over pairing configurations models the randomness in the outcome of an observation. So, to find statistical emergence properties, it is natural to construct such distribution from the first principle.

This section will find Binomial-like probability distributions for B and C models. It is analytically and practically interesting that both probabilities are expressible in closed form, and consequently, most of their statistics, such as their mean, standard deviation, and other moments, are in closed form. Also, we find the asymptotic leading terms of the same statistics for large system sizes.

Furthermore, distributions of B and C models satisfy the Large Deviation Principle (LDP) [18, 65, 66]. As a result, we will derive their corresponding large deviation distributions and their rate functions. In addition, as obtained in the previous chapter, the logarithm of state-space volume is proportional to $N \ln N$. Due to this fact, we shall find the logarithm of LDP distribution has $N \ln N$ speed, which is the direct consequence of faster than exponential state space for emerging states. We also find the marginal and joint probability distributions for one or more elements in closed form. These distributions will be helpful in finding some information theory measures in the later chapter.

The normalisation constant of distributions complies with a recursive relation corresponding to the state space recursive equation. We will see the geometry of the state space encodes as the normalisation constant recursive relation.

To reduce the notation clutter and convenience in reading them, we use lower case n in place of capital N to denote the number of elements and system sizes, although the latter is common in physics publications. We will follow the same convention in this and the following chapter but return to N in chapter (5).

In introductory statistics in the limit $n \to \infty$, the Poisson distribution obtains from the Binomial distribution when $n\rho = \lambda$ is kept constant, while ρ denotes the probability of success and λ is the Poisson rate parameter [22]. On the same footing, similar limiting distributions exist for B and C models. The limiting distributions are obtained by scaling the B and C models' parameters with system size such that the ratio is kept constant – similar to the success parameter in the Binomial distribution and rate parameter in the Poisson distribution.

Together, we shall see that the models' averages can be considered as an order parameter such that in the thermodynamic limit ($n \to \infty$): the average of LDP distributions is zero, whereas the limiting distributions have non-zero averages. Thus, the average as an order parameter corresponds to a second-order phase transition in physics [8]. One can envisage that the phase transition must occur whenever a secondary model utilises one of B or C models as its building blocks and control the size dependency of the free parameter through some internal mechanisms.

To use the resulting distributions in statistical modelling, one needs to infer the free parameters from one or more observed values of realisations of random variables. This chapter finds the maximum likelihood estimations for the free parameters in both B and C distributions.

Lastly, we demonstrate the parameter inference of the distributions in a Bayesian setting. Interestingly, we will derive a conjugate prior for the parameters that are expressible in closed form. Meanwhile, we show that the normalisation constant of the conjugate prior is well-defined, converging and expressible in closed form. Finally, we remark that expressing the results in closed form for Bayesian statistics or others is not only a mathematical convenience. It also provides an opportunity to apply and employ them for further analytical investigation or modelling.

3.1 Binomial-like Distribution

This section examines random variables that are functions of emergent states of pairing models, introduced in chapter (2). Before delving into the *pairing random variables*, let us review the properties of an ordinary binary random variable. It is common practice [9, 48] to model a binary random variable, say X_i, by the Bernoulli distribution

$$Ber(X_i; \rho) = \rho^{X_i}(1-\rho)^{(1-X_i)}, \qquad 0 \leq \rho \leq 1. \tag{3.1}$$

For instance, the random variable X_i is a function over the state of a single coin showing either head or tail

$$X_i : \{\text{tail}, \text{head}\} \to \{0, 1\}, \tag{3.2}$$

and the parameter ρ defines the probability of observing the head state.

Let us consider a model consisting of n distinct coins such that the random variable n_h denotes the number of coins showing head. For the random variable X_i, n_h defines as

$$n_h = \sum_{i=1}^{n} X_i, \tag{3.3}$$

and its probability is determined by the Binomial distribution [9, 48]

$$Bin(n_h; n, \rho) = \binom{n}{n_h} \rho^{n_h}(1-\rho)^{(n-n_h)}. \tag{3.4}$$

Note that because of the Binomial expansion identity, the normalisation condition is always satisfied

$$\sum_{n_u=0}^{n} \binom{n}{n_u} \rho^{n_u}(1-\rho)^{(n-n_u)} = (\rho + 1 - \rho)^n = 1. \tag{3.5}$$

However, to model the pairing mechanism in B and C models, one needs to construct similar distributions that satisfy the normalisation condition for their non-negative probabilities. In these cases, the domain of random variables is the elements' states, resulting from an underlying pairing mechanism. Therefore, this section defines random variables relevant to the pairing mechanism and their state spaces and subsequently introduces their probability distributions.

3.1.1 B-Model's Probability Distribution

Let $\mathcal{S}_n$ denotes the set of all possible configurations for the B-Model, or as it is called in statistical mechanics [64, 28], the ensemble of configurations. And also, let n_p denotes the number of pairs in an observed configuration, say c_{n_p}, such that $n_p \in \{0, 1, \ldots, \lfloor \frac{n}{2} \rfloor\}$. Notice that since the number of pairs is a whole number, the maximum of n_p is equal to $\lfloor n/2 \rfloor$.

From the onset, for the purpose of this chapter, we assume equal probability among configurations that consist of the same number of pairs, and therefore, n_p provides the means to partition the ensemble accordingly. To use the number of pairs for partitioning $\mathcal{S}_n$, we define the subset of configurations with i pairs, denoted by S_i, as

$$S_i \equiv \{c_i \in \mathcal{S}_n : c_i \text{ has } i \text{ pairs}\}, \tag{3.6}$$

such that

$$\mathcal{S}_n = \bigcup_{i=0}^{\lfloor \frac{n}{2} \rfloor} S_i, \qquad S_i \bigcap S_j = \emptyset \quad \text{for} \quad i \neq j. \tag{3.7}$$

For instance, figure (3.1) represents partitions of $\mathcal{S}_4$, and as we can see for $n = 4$, the number of pairs are restricted to $n_p \in \{0, 1, 2\}$.

Next, the probability of an event set S_i, denoted by p_i, defines as

$$p_i \equiv P_n(S_i) = P_n(n_p = i), \qquad 0 \leq p_i \leq 1, \tag{3.8}$$

and the normalization condition necessitates

$$\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} p_i = 1. \tag{3.9}$$

Although we assumed equal probability among configurations with the same number of pairs, the cardinality of each subset is different from the others. Naturally, the cardinality of the subset S_i is the right measure to define p_i. From the discussion in the previous chapter, we know that the cardinality of the subset S_i is equal to the number of distinct $2i$ pairs choices among n elements times $(2i-1)!!$ distinguishable pairs

$$|S_i| = \binom{n}{2i}(2i-1)!!. \tag{3.10}$$

In the case of a uniform probability over $\mathcal{S}_n$, we must have

$$p_i = P_n(S_i) = \frac{|S_i|}{|\mathcal{S}_n|} = \frac{\binom{n}{2i}(2i-1)!!}{\Omega_1(n)}. \tag{3.11}$$

However, while the event set S_i can contain one or more configurations with i pairs, the probability p_i corresponds to the event of observing any of them. In addition, to have a probability for each and every $c_i \in S_i$, we define a fine-grained probability

$$q_i \equiv P_n(c_i). \tag{3.12}$$

Using the assumption of equal probability among configurations with the same number of pairs, p_i must be equal to q_i times the cardinality of the subset S_i. Thus

$$P_n(S_i) = |S_i|\; P_n(c_i) \Rightarrow p_i = \binom{n}{2i}(2i-1)!!\; q_i. \tag{3.13}$$

Accordingly, the normalisation condition (3.9) is

$$\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2i}(2i-1)!!\; q_i = 1. \tag{3.14}$$

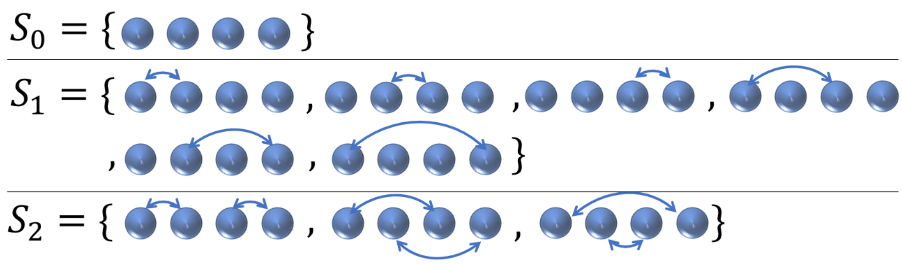


Figure 3.1: Partitioning $\mathcal{S}_4$ to three disjoint subsets. The arrows show the pairing between balls.
Adapted from "New probability distribution describing emergence in state space," Pazuki, Roozbeh H. and Jensen, Henrik J., 2021, *Journal of Physics Communications*, 5(9), p. 095002. DOI: 10.1088/2399-6528/ac1f74. Copyright 2021 under the terms of the Creative Commons Attribution 4.0 licence.

Two examples are provided here for such distributions over $\mathcal{S}_4$. Table 3.1 shows a uniform distribution, by which all configurations have the same probability: $q_i = 1/10$. Whereas, table 3.2 shows the equal probability for all event sets: $p_i = 1/3$.

Table 3.1: Uniform distribution.

p_i	q_i
$p_0 = \frac{1}{10}$	$q_0 = \frac{1}{10}$
$p_1 = \frac{6}{10}$	$q_1 = \frac{1}{10}$
$p_2 = \frac{3}{10}$	$q_2 = \frac{1}{10}$

Table 3.2: $\mathcal{S}_4$ distribution.

p_i	q_i
$p_0 = \frac{1}{3}$	$q_0 = \frac{1}{3}$
$p_1 = \frac{1}{3}$	$q_1 = \frac{1}{3\times 6}$
$p_2 = \frac{1}{3}$	$q_2 = \frac{1}{3\times 3}$

3.1.2 C-Model's Probability Distribution

For the C-model, we use a similar notation as in the previous section to define the set of configurations and the corresponding probability distribution. Let us denote by $\mathcal{S'}_n$ the ensemble of the C-model's configurations. As we did for the B-model, $\mathcal{S'}_n$ partitions to disjoint subsets, namely S_i. Recall that configurations in S_i have the same number of pairs, and consequently, each configuration has $(n-2i)$ coins that are in a non-pair state. If n_h denotes the number of coins in the head state, then, we must have

$$n_h \in \{0, 1, \ldots, (n-2i)\}. \tag{3.15}$$

We assert an equal probability assumption for the C-model, such that the probability of observing configurations that have the same number of pairs, n_p, and heads, n_h, are equal. Furthermore, S_i partitions to disjoint subsets, denoted by S_{ij}, that have j coins in head states. Thus

$$S_{ij} \equiv \{c_{ij} \in S_i : c_{ij} \text{ has } j \text{ coins in head state}\}, \tag{3.16}$$

and

$$S_i = \bigcup_{j=0}^{n-2i} S_{ij}, \qquad S_{ij} \bigcap S_{ik} = \emptyset \quad \text{for} \quad j \neq k. \tag{3.17}$$

Given S_{ij} is a subset of S_i, we define the conditional probability of event $S_{ij}|S_i$, or the probability of observing configurations with j heads, given i pairs

$$P_n(S_{ij}|S_i) \equiv P_n(n_h = j|n_p = i) = p_{j|i}, \qquad \forall i: \quad 0 \leq p_{j|i} \leq 1. \tag{3.18}$$

There are $\binom{n-2i}{j}$ distinct configurations for selecting j heads among $n-2i$ non-pair coins. Therefore, for the $n-2i$ non-pair coins with j heads, the Binomial distribution finds the probability $p_{j|i}$ as

$$p_{j|i} = \binom{n-2i}{j}\rho^j(1-\rho)^{n-2i-j}, \qquad 0 \leq \rho \leq 1, \tag{3.19}$$

where ρ defines the probability of getting head in throwing a single coin. Note that the conditional distribution admits the normalisation condition

$$\sum_{j=0}^{n-2i} p_{j|i} = (\rho + 1 - \rho)^{n-2i} = 1. \tag{3.20}$$

Finally, the probability of event set S_{ij}, or observing configurations with i pairs and j heads, is

$$P_n(S_{ij}) \equiv P_n(n_h = j, n_p = i) = p_{ij}, \tag{3.21}$$

and according to the probability chain rule [48], it can be written as

$$p_{ij} = p_i \times p_{j|i} = \binom{n}{2i}(2i-1)!!\ q_i \left[\binom{n-2i}{j}\rho^j(1-\rho)^{n-2i-j}\right]. \tag{3.22}$$

Apart from that, we observe the cardinality of S_{ij} is equal to $\binom{n}{2i}(2i-1)!!\ \binom{n-2i}{j}$. At the same time, because all configurations in S_{ij} have the same probability and are similar to the B-model, one concludes the probability of observing a specific configuration $c_{ij} \in S_{ij}$ must be

$$P_n(c_{ij}) = q_i\rho^j(1-\rho)^{n-2i-j}. \tag{3.23}$$

This argument is schematically represented in figure (3.2) as a probability tree.

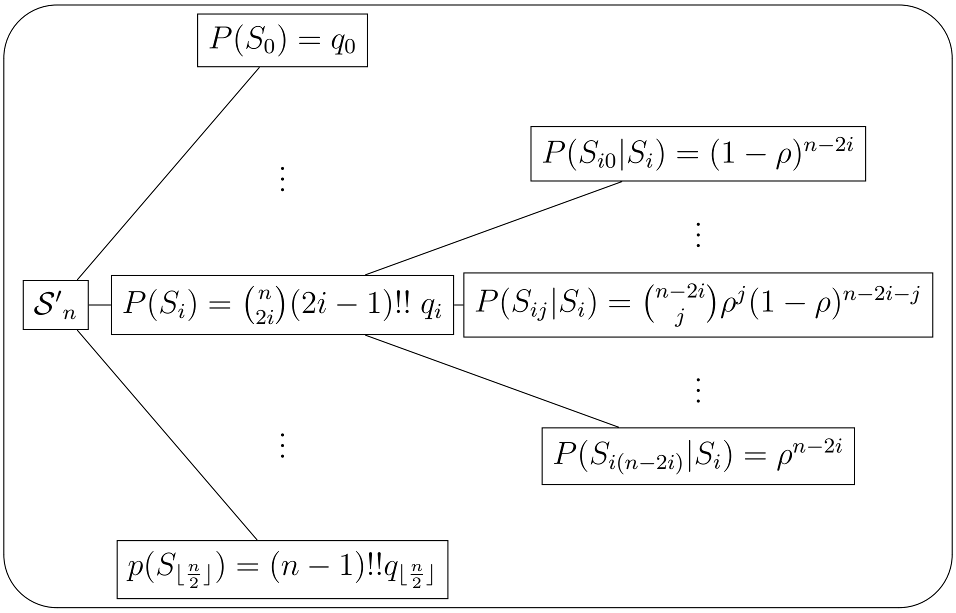


Figure 3.2: The pairing coins probability tree.

Adapted from "New probability distribution describing emergence in state space," Pazuki, Roozbeh H. and Jensen, Henrik J., 2021, *Journal of Physics Communications*, 5(9), p. 095002. DOI: 10.1088/2399-6528/ac1f74. Copyright 2021 under the terms of the Creative Commons Attribution 4.0 licence.

To check the normalisation condition for p_{ij}, we write

$$\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor}\sum_{j=0}^{n-2i} p_{ij} = \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor}\sum_{j=0}^{n-2i} \binom{n}{2i}(2i-1)!!\ q_i \left[\binom{n-2i}{j}\rho^j(1-\rho)^{n-2i-j}\right]$$

$$= \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2i}(2i-1)!!\ q_i \sum_{j=0}^{n-2i}\binom{n-2i}{j}\rho^j(1-\rho)^{n-2i-j}$$

$$= \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2i}(2i-1)!!\ q_i(\rho+1-\rho)^{n-2i} = \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2i}(2i-1)!!\ q_i = 1. \tag{3.24}$$

3.1.3 Finding p_i and p_{ij} in Closed Form

One of the appealing aspects of well-known statistical probability distributions is their handful of free parameters, through which the shape, rate, scale or other properties are controlled. To put it simply, a small degree of freedom is enough to specify the details. For example, in a statistical modelling setting, one needs to estimate a small number of parameters from data.

In this section, we will redefine $P_n(n_p = i)$ as a one parameter probability distribution, and $P_n(n_p = i, n_h = j)$ as a two parameters one, including ρ, for the C and B models, respectively. Let us focus on the C-model for a moment. For two balls, as depicted in figure (3.3), there are two configurations in the set $\mathcal{S}_2$, and the normalisation condition is written as

$$q_0 + q_1 = 1. \tag{3.25}$$

Denoting the ratio of these two probabilities by r for

$$r = \frac{q_0}{q_1}, \qquad r \in [0, \infty), \tag{3.26}$$

we can write q_0 and q_1 in terms of r as

$$q_0 = \frac{r}{r+1}, \qquad q_1 = \frac{1}{r+1}. \tag{3.27}$$

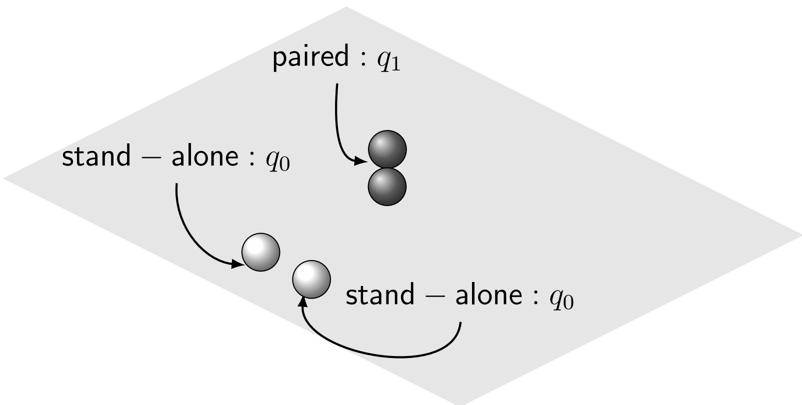


Figure 3.3: Two balls in paired or stand-alone state.
Adapted from "New probability distribution describing emergence in state space," Pazuki, Roozbeh H. and Jensen, Henrik J., 2021, *Journal of Physics Communications*, 5(9), p. 095002. DOI: 10.1088/2399-6528/ac1f74. Copyright 2021 under the terms of the Creative Commons Attribution 4.0 licence.

Remember that for n balls, $(\lfloor \frac{n}{2} \rfloor + 1)$ distinct probabilities are required to completely define the distribution of the C-model, namely q_i. To find q_i in terms of a single parameter, we are pursuing a rule that relates q_i for $n-1$ to q'_i for n. Therefore, to stress the dependence of q_i on the number of balls, we replace q_i by $q_n(i)$, and establish a relation between $q_n(i)$ and $q_{n-1}(i)$.

For $n = 3$, the number of possible pairs are the same as the two balls case, in other words, $n_p \in \{0, 1\}$. Indeed, as it is shown in figure (3.4), the difference between $n = 2$ and $n = 3$ is in the degenerate configurations for $n_p = 1$: there are three distinguishable configurations for $(n = 3, n_p = 1)$, when it compares to figure (3.3).

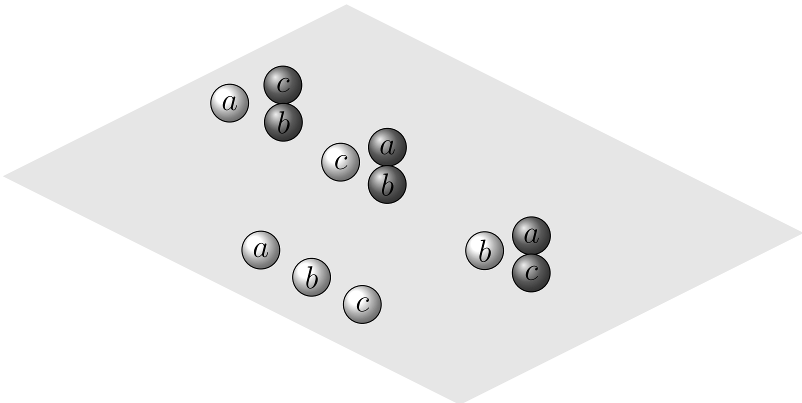


Figure 3.4: Three balls in paired or stand-alone states.

This is true for all consecutive even and odd numbers, $2n$ and $2n+1$, since $n = \lfloor \frac{2n}{2} \rfloor = \lfloor \frac{2n+1}{2} \rfloor$. But the number of degeneracies is different

$$P_{2n}(n_p = i) = \binom{2n}{2i} (2i-1)!!\; q_{2n}(i),$$

$$P_{2n+1}(n_p = i) = \binom{2n+1}{2i} (2i-1)!!\; q_{2n+1}(i). \tag{3.28}$$

Therefore, it is reasonable to assume the differences between probabilities of observing the n_p pairs for even and odd numbers are due to the differences in degeneracies, while the probability of making n_p pairs remains the same.

For example for $n = 3$, the ratio of $q_3(0)$ and $q_3(1)$ must be equal to r, while for $n = 2$ the ratio of $q_2(0)$ and $q_2(1)$ is defined as r

$$\frac{q_3(0)}{q_3(1)} = \frac{q_2(0)}{q_2(1)} = r. \tag{3.29}$$

So, writing the normalisation relation implies

$$q_3(0) + 3q_3(1) = 1 \implies$$

$$q_3(0) = \frac{r}{3+r}, \qquad q_3(1) = \frac{1}{3+r}. \tag{3.30}$$

To check the last result, consider the case of uniform distribution. In that case for $n = 2$, when the probability of a pairing state is equal to a stand-alone state ($q_2(0) = q_2(1) = \frac{1}{2}$), or $r = 1$, for $n = 3$ the probability distribution is also uniform

$$q_3(0) = \frac{1}{4}, \qquad q_3(1) = \frac{1}{4}. \tag{3.31}$$

Thus, the equal probability for the $n = 2$ propagates to $n = 3$.

Furthermore, for $n = 4$, the number of pairs are $n_p \in \{0, 1, 2\}$, and the normalisation is

$$q_4(0) + 6q_4(1) + 3q_4(2) = 1. \tag{3.32}$$

And since in a configuration for $n_p = 2$ there are twice more pairs in comparison to $n_p = 1$, we expect $r \times r$ as the ratio between $q_4(0)$ and $q_4(2)$, and r as the ratio between $q_4(0)$ and $q_4(1)$

$$\frac{q_4(0)}{q_4(2)} = r^2, \qquad \frac{q_4(0)}{q_4(1)} = r, \tag{3.33}$$

which implies

$$r^2 q_4(2) + 6rq_4(2) + 3q_4(2) = 1 \implies$$

$$q_4(0) = \frac{r^2}{r^2+6r+3}, \quad q_4(1) = \frac{r}{r^2+6r+3}, \quad q_4(2) = \frac{1}{r^2+6r+3}. \tag{3.34}$$

Let us again check the case of the uniform distribution. For $r = 1$,

$$q_4(0) = \frac{1}{10}, \quad q_4(1) = \frac{1}{10}, \quad q_4(2) = \frac{1}{10}, \tag{3.35}$$

and once more it obtains the uniform distribution. Hence, again, the equal probability propagates to $n = 4$.

Using induction steps, while considering the alternating conditions between odds and even numbers, we find

$$q_n(i) = \frac{r^{\lfloor \frac{n}{2} \rfloor - i}}{\sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2k} (2k-1)!!\, r^{\lfloor \frac{n}{2} \rfloor - k}}, \tag{3.36}$$

and the probability distribution derives as

$$P_n(n_p = i) = \frac{\binom{n}{2i}(2i-1)!!\ r^{\lfloor \frac{n}{2} \rfloor - i}}{\sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2k}(2k-1)!!\ r^{\lfloor \frac{n}{2} \rfloor - k}}. \tag{3.37}$$

Defining the normalisation constant $c_n(r)$ as

$$c_n(r) = \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2k}(2k-1)!!\ r^{\lfloor \frac{n}{2} \rfloor - k}, \tag{3.38}$$

the C-model's probability distribution writes as

$$P_n(n_p = i) = \frac{1}{c_n(r)} \binom{n}{2i}(2i-1)!!\ r^{\lfloor \frac{n}{2} \rfloor - i}. \tag{3.39}$$

The q_i parameter in the B-model and C-model was defined based on the same partitioning procedures, and therefore, they are the same. Hence, the B-model's probability distribution must be

$$P_n(n_p = i, n_h = j) = \frac{1}{c_n(r)} \binom{n}{2i}(2i-1)!!\ r^{\lfloor \frac{n}{2} \rfloor - i} \binom{n-2i}{j} \rho^j (1-\rho)^{n-2i-j}. \tag{3.40}$$

3.1.3.1 Example

For $n = 4$, when $r = \frac{1}{2}$, or the ratio of probability of being in stand-alone state is half of making a pair, $c_4(1/2)$ obtains as

$$c_4(\frac{1}{2}) = \left(\frac{1}{2}\right)^2 + 6\left(\frac{1}{2}\right) + 3 = \frac{25}{4},$$

and the distribution is

$P_4(n_p = 0)$	$\frac{1}{25}$
$P_4(n_p = 1)$	$\frac{12}{25}$
$P_4(n_p = 2)$	$\frac{12}{25}$

Table 3.3: The probability distribution for $n = 4$ and $r = \frac{1}{2}$.

3.1.3.2 Finding $c_n(r)$

The first eight iterations of $c_n(r)$ is shown in table (3.4). These are very similar to Hermite polynomials as a function of r, though all their coefficients are positive. Therefore, we expect to find properties similar to Hermite polynomials for $c_n(r)$.

n	$c_n(r)$
2	$r+1$
3	$r+3$
4	r^2+6r+3
5	$r^2+10r+15$
6	$r^3+15r^2+45r+15$
7	$r^3+21r^2+105r+105$
8	$r^4+28r^3+210r^2+420r+105$
9	$r^4+36r^3+378r^2+1260r+945$

Table 3.4: Normalisation constant for first eight ns.

We start with even numbers and write equation (3.38) for $2n$

$$c_{2n}(r) = r^n \sum_{i=0}^{n} \binom{2n}{2i} (2i-1)!!\ r^{-i}. \tag{3.41}$$

If we define a power series, say $f_{2n}(x)$, as

$$f_{2n}(x) \equiv \sum_{i=0}^{n} \binom{2n}{2i} (2i-1)!!\ x^{i}, \tag{3.42}$$

then, the normalisation constant in terms of $f_{2n}(x)$ writes as

$$c_{2n}(r) = r^n f_{2n}(\frac{1}{r}). \tag{3.43}$$

Similarly, for odd numbers, we define

$$f_{2n+1}(x) \equiv \sum_{i=0}^{n} \binom{2n+1}{2i} (2i-1)!!\ x^{i}, \tag{3.44}$$

and the normalisation constant writes in terms of $f_{2n+1}(x)$ as

$$c_{2n+1}(r) = r^n f_{2n+1}(\frac{1}{r}). \tag{3.45}$$

Using these power series in appendix (B.1), we find recursive relations between normalisation constants as follows

$$\begin{cases} c_{2n}(r) = rc_{2n-1}(r) + (2n-1)c_{2n-2}(r) \\ c_{2n+1}(r) = c_{2n}(r) + 2nc_{2n-1}(r). \end{cases} \tag{3.46}$$

This result resembles the recursive relation for the state space volume in equation (2.9). For even numbers, it is equivalent to $s = r$ and $p = 1$. And for odd numbers, it is $s = 1$ and $p = 1$.

However, we must emphasise the generality of these relations in comparison to the state space volumes in equation (2.9), since they are satisfied by these polynomials of an arbitrary non-negative r. It seems that the state space structure is not only encoded in a single recursive relation but also in normalisation constants of the probability distributions that are defined over it.

One can even see them as the composition law that decomposes the normalisation constant in terms of smaller system ones. Recall that in statistical mechanics jargon, the normalisation constant is called *partition function*, and for this specific space, we know its decomposition law [33].

Furthermore, it is important to mention that initial conditions for recursive relation in equation (3.46) are

$$c_1(r) = 1, \qquad c_2(r) = r + 1, \tag{3.47}$$

whereas $\Omega_1(n)$ and $\Omega_2(n)$ had different initial conditions

$$\Omega_1(1) = 1, \qquad \Omega_1(2) = 2, \tag{3.48}$$

and

$$\Omega_2(1) = 2, \qquad \Omega_2(2) = 5. \tag{3.49}$$

Note that for $r = 4$,

$$c_{2n}(4) = \Omega_2(2n). \tag{3.50}$$

The last observation tells us that the same machinery that derived the asymptotic leading terms of $\Omega_2(2n)$ can also give us the asymptotic form of $c_{2n}(r)$. To achieve that, first, observe the rearrangement of the terms in $c_{2n}(r)$ sum

$$c_{2n}(r) = \sum_{i=0}^{n} \binom{2n}{2i} (2i-1)!!\, r^{n-i} = (2n)! \sum_{i=0}^{n} \frac{r^{n-i} 2^{-i}}{i!(2n-2i)!}$$

$$= (2n)! \sum_{k=0}^{n} \frac{r^k 2^{k-n}}{(n-k)!(2k)!}, \tag{3.51}$$

where we used $k = n - i$ in the last step. The resulting sum is the same expansion we had in equation (A.38) for $\Omega_s(2n)$, whenever $s = \sqrt{r}$ and $p = 1$. Thus, using equation (A.73), one finds

$$c_{2n}(r) = \frac{1}{\sqrt{2\pi} e^{\frac{r}{4}} (\frac{rn}{2})^{\frac{1}{4}}} \left(\frac{2n}{e}\right)^n \sum_{1 \le k \le 2\sqrt{rn/2}} \left(\frac{\sqrt{\frac{rn}{2}} e}{k}\right)^{2k} \left[1 + O(\frac{1}{\sqrt{n}})\right]. \tag{3.52}$$

However, when r is kept constant and $1 \ll n$, all the assumptions and steps in finding the asymptotic leading term of $\Omega_s(2n)$ in equation (A.99) are valid and give

$$c_{2n}(r) = \frac{e^{-\frac{r}{4}}}{\sqrt{2}} \left(\frac{2n}{e}\right)^n e^{\sqrt{2rn}} \left(1 + O(\frac{1}{\sqrt{n}})\right). \tag{3.53}$$

We have to point out here that the above leading term is valid for $r \le n$, which is good enough when r is kept constant and n is large. This is the case for large deviation estimates in the next section.

3.2 Large Deviation Estimates

A probability distribution satisfies the Large Deviation Principle (LDP) [19, 65, 66] if the limit

$$\lim_{n\to\infty} -\frac{1}{a_n} \log P_n(X_n), \tag{3.54}$$

exists for a random variable X_n and a sequence of distributions P_n, and a sequence of positive numbers a_n, called *speed*, for $n \in \{1, 2, \dots\}$ that tends to ∞. In this section, we will check the existence of the limit for both $P_n(n_p)$ and $P_n(n_p, n_u)$.

For a configuration with size n, let the random variable X_i denote the states of a single element. Therefore, the number of pairs is defined as

$$n_p = \frac{1}{2}\sum_{i=0}^{n} \delta_{X_i,p}, \tag{3.55}$$

whereas $\delta_{X_i,p}$ is the Kronecker delta if X_i is in a pair state. Moreover, the number of elements in head state is defined as

$$n_h = \sum_{i=0}^{n} \delta_{X_i,h}, \tag{3.56}$$

whereas $\delta_{X_i,h}$ is the Kronecker delta if X_i is in a head state.

To study the large deviation property of the distributions, we use the following transformation to define a normalised random variable for the ratio of the number of pairs as follows

$$m_n \equiv \frac{2n_p}{n} = \frac{1}{n}\sum_{i=0}^{n} \delta_{X_i,p}, \qquad 0 \le m \le 1, \tag{3.57}$$

and similarly, for the number of heads in a random configuration as

$$s_n \equiv \frac{n_h}{n-2n_p} = \frac{1}{n-2n_p}\sum_{i=0}^{n} \delta_{X_i,h}, \qquad 0 \le s_n \le 1. \tag{3.58}$$

In the continuum limit, or $n \to \infty$, both variables are in $\mathbb{R}$ [65].

3.2.1 Large Deviation Principle Satisfied by $P_n(n_p)$

In Appendix (B.2), equation (B.21), and by using Sterling's approximation for $\log n!$, we obtain

$$\ln P_n(m_n) = -(\frac{1-m_n}{2})n\ln n - \frac{n}{2}\left[m_n \ln m_n + (1-m_n)\ln\frac{(1-m_n)^2}{er}\right] + O(\sqrt{n}). \tag{3.59}$$

The terms inside the square bracket in the last equation resemble the Shannon entropy [15]. We remind the reader, for $0 \le p \le 1$, Shannon entropy is defined as

$$H(p) = -p\ln p - (1-p)\ln(1-p), \tag{3.60}$$

and along the same line, we define $\tilde{H}_r(p)$ for different r as

$$\tilde{H}_r(p) \equiv -p\ln p - (1-p)\ln\frac{(1-p)^2}{\mathrm{e}r}, \tag{3.61}$$

in which, e is Euler constant. So, $\ln P_n(m_n)$ rewrites as

$$\ln P_n(m_n) = -(\frac{1-m_n}{2})n\ln n + \frac{n}{2}\tilde{H}_r(m_n) + O(\sqrt{n}). \tag{3.62}$$

To find the LDP limit, the speed is not linear but is of order $O(n\ln n)$. By dividing $\ln P_n(m_n)$ with $n\ln n$ and taking the limit with respect to n, we find that the limit exists and results in

$$\lim_{n\to\infty} -\frac{1}{n\ln n}\ln P_n(m_n) = \frac{1-m_n}{2}. \tag{3.63}$$

In other words, the large deviation principle is satisfied, and the *rate* function [19] for the B-model distribution is

$$I(m_n) = \frac{1-m_n}{2}. \tag{3.64}$$

In practice, when the value of $n\ln n$ is not appreciably larger than n, one requires to include the correction due to terms in $O(n)$ order as

$$P_n(m_n) \asymp \mathrm{e}^{-n\ln n I(m_n) + \frac{n}{2}\tilde{H}_r(m_n)}. \tag{3.65}$$

In figure (3.5), left panel, we see the shape of $\tilde{H}_r(m_n)$ and its dependence on r. As it is shown, in spite of the right leg being fixed on zero, the left leg of the curve is r dependent. Note that $\tilde{H}_r(m_n)$ can be negative, which may seems problematic for a rate function[1]. This is not a striking result, since $\tilde{H}_r(m_n)$ is a $O(\frac{1}{\ln n})$ correction to the rate function.

Although the large deviation rate function is $(1-m_n)/2$, for any practical purpose the $O(\frac{1}{\ln n})$ correction must be included. For instance, let say the number of elements is of order of Avogadro's number, $n = 10^{23}$. Then $n\ln n = 53\times10^{23}$ is 53 times larger than n, and $\tilde{H}_r(m_n)$ can have a comparable significant contribution in comparison to $(1-m_n)/2$.

Also, in figure (3.5), right panel, we plotted the error in estimating m_n in comparison to the directly calculated value at the point of maximum probability. The estimated

[1]Remember that rate functions must be non-negative [19].

value is evaluated at the maximum of corrected rate function, say $n \ln n I(m_n) - (n/2)\tilde{H}_r(m_n)$, and we see that the error decreases for increasing n.

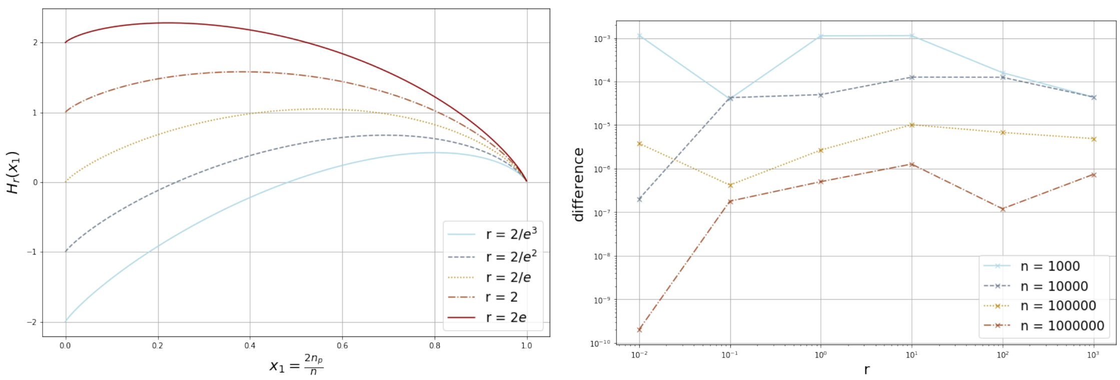


Figure 3.5: Left panel shows $\tilde{H}_r(m_n)$ for different r. Similar to the Shannon entropy, the $\tilde{H}_r(m_n)$ is zero at $m_n = 1$, and the value at $m_n = 0$ depends on r. **Right panel** plots the difference between the estimated m_n at the maximum of the distribution from the directly calculated value versus r. Each plot is for a different n, and the error reduces for increasing n.

3.2.2 Large Deviation Principle Satisfied by $P_n(n_p, n_u)$

In appendix (B.3), equation (B.24) finds $\ln P_n(m_n, s_n)$ as

$$\ln P_n(m_n, s_n) = -(\frac{1-m_n}{2})\, n \ln n + \frac{n}{2}\left[\tilde{H}_r(m_n) - 2(1-m_n)\left(s_n \ln(\frac{s_n}{\rho}) + (1-s_n)\ln(\frac{1-s_n}{1-\rho})\right)\right] + O(\sqrt{n}). \tag{3.66}$$

Remember, ρ is the probability of getting head in the case of a single pairing coin, and defining the Bernoulli rate function as [65, 66]

$$I_\rho(x_2) = x_2 \ln(\frac{x_2}{\rho}) + (1-x_2)\ln(\frac{1-x_2}{1-\rho}), \tag{3.67}$$

$\ln P_n(m_n, s_n)$ becomes

$$\ln P_n(m_n, s_n) = -(\frac{1-m_n}{2})\, n \ln n + \frac{n}{2}\left[\tilde{H}_r(m_n) - 2(1-m_n) I_\rho(s_n)\right] + O(\sqrt{n}). \tag{3.68}$$

Finally, the LDP limit exists and is equal to

$$\lim_{n\to\infty} -\frac{1}{n\ln n}\log P_n(m_n, s_n) = \frac{1-m_n}{2}. \tag{3.69}$$

Therefore, $P_n(m_n, s_n)$ satisfies the large deviation principle with the rate function

$$I(m_n, s_n) = \frac{1-m_n}{2}. \tag{3.70}$$

As we explained in the previous section, in practice, the correction terms of the rate function are significant. So, including the correction terms in order $O(n)$ obtains

$$P_n(m_n, s_n) \asymp \mathrm{e}^{-n\ln n I(m_n,s_n)+\frac{n}{2}\tilde{H}_r(m_n)-n(1-m_n)I_\rho(s_n)}. \tag{3.71}$$

3.2.3 The Limiting Case $r/n \to \epsilon$

In the thermodynamic limit, or when $n \to \infty$, for B and C models, we can find

$$\lim_{n\to\infty} P_n(m_n = 1) \to 1, \tag{3.72}$$

and

$$\lim_{n\to\infty} P_n(m_n = 1, s_n = 0) \to 1. \tag{3.73}$$

In other words, in both cases, all elements are in the pair state.

Recall that r is the ratio of abundance of stand-alone to pair states, and in finding the above limits r is kept constant. To elaborate, $m_n = 1$ in thermodynamic limit is the consequence of the fast growth rate of the term $\binom{n}{2n_p}(2n_p-1)!!$ that for a given n_p counts the degenerate states. So, in thermodynamic limit for any finite value of r, the volume of the subset $S_{\lfloor \frac{n}{2}\rfloor} \in \mathcal{S}$ is so large in comparison to the remaining ones that we find all elements in the pair state.

Despite this fact, one can envisage a non-constant r, especially as an increasing and size-dependent function, and consequently study the large deviation property of the distributions if the following limit exists

$$0 < \lim_{n\to\infty} \frac{r}{n} = \epsilon < \infty. \tag{3.74}$$

Note that in finding the asymptotic leading term of $c_n(r)$ in equation (A.99), we assumed r is constant. However, in appendix (B.4), equation (B.55), for r/n is non-zero and bounded above, we find the leading term for $c_n(\epsilon)$ as

$$c_n(\epsilon) = \begin{cases} \frac{g(\epsilon)^{-\frac{n}{2}g(\epsilon)} \mathrm{e}^{n\sqrt{\frac{\epsilon}{\mathrm{e}}}} (\frac{n}{\mathrm{e}})^{n/2}}{\sqrt{\sqrt{\mathrm{e}\epsilon} f(\epsilon) g(\epsilon)}} & \epsilon \leq \mathrm{e} \\ \frac{g(\epsilon)^{-\frac{n}{2}g(\epsilon)} (\epsilon n)^{n/2}}{\sqrt{n\pi\epsilon f(\epsilon) g(\epsilon)}} & \epsilon > \mathrm{e} \end{cases} \tag{3.75}$$

where e is Euler constant, $f(\epsilon)$ is defined in equation (B.31) as

$$f(\epsilon) = \sqrt{1 + \frac{4}{\epsilon}} - 1, \tag{3.76}$$

and $g(\epsilon)$ is defined in equation (B.35) as

$$g(\epsilon) = (1 - \frac{\epsilon f(\epsilon)}{2}). \tag{3.77}$$

Taking the logarithm of $\ln c_n(\epsilon)$ results in

$$\ln c_n(\epsilon) = \begin{cases} -\frac{n}{2} g(\epsilon) \ln g(\epsilon) + n\sqrt{\frac{\epsilon}{\mathrm{e}}} + \frac{n}{2}\ln n - \frac{n}{2} + O(1) & \epsilon \leq \mathrm{e} \\ -\frac{n}{2} g(\epsilon) \ln g(\epsilon) + \frac{n}{2}\ln n + \frac{n}{2}\ln(\epsilon) + O(\ln n) & \epsilon > \mathrm{e} \end{cases}, \tag{3.78}$$

then, by replacing $\ln c_n(r)$ by its asymptotic leading term and r by ϵn, equation (B.19) rewrite as follows:

- For $\epsilon \leq \mathrm{e}$:

$$\begin{aligned} \ln P_\epsilon(m_n) &= -\frac{n}{2}\left[2\sqrt{\frac{\epsilon}{\mathrm{e}}} - g(\epsilon)\ln g(\epsilon) + m_n \ln m_n + (1 - m_n) \ln \frac{(1-m_n)^2}{\mathrm{e}\epsilon}\right] \\ &= -\frac{n}{2}\left[2\sqrt{\frac{\epsilon}{\mathrm{e}}} - g(\epsilon)\ln g(\epsilon) - \tilde{H}_\epsilon(m_n)\right]. \end{aligned} \tag{3.79}$$

- For $\epsilon > \mathrm{e}$:

$$\begin{aligned} \ln P_\epsilon(m_n) &= -\frac{n}{2}\left[\ln(\epsilon \mathrm{e}) - g(\epsilon)\ln g(\epsilon) + m_n \ln m_n + (1 - m_n) \ln \frac{(1-m_n)^2}{e\epsilon}\right] \\ &= -\frac{n}{2}\left[\ln(\epsilon \mathrm{e}) - g(\epsilon)\ln g(\epsilon) - \tilde{H}_\epsilon(m_n)\right]. \end{aligned} \tag{3.80}$$

So,

$$P_\epsilon(m_n) \asymp \mathrm{e}^{-nI_1(m_n;\epsilon)}, \tag{3.81}$$

where

$$I_1(m_n;\epsilon) = \frac{1}{2}\begin{cases} 2\sqrt{\frac{\epsilon}{\mathrm{e}}} - g(\epsilon)\ln g(\epsilon) - \tilde{H}_\epsilon(m_n) & 0 < \epsilon \le \mathrm{e} \\ \ln(\epsilon \mathrm{e}) - g(\epsilon)\ln g(\epsilon) - \tilde{H}_\epsilon(m_n) & \epsilon > \mathrm{e} \end{cases}, \tag{3.82}$$

Similarly, for $P_\epsilon(m_n, s_n)$, all the terms are the same, and we have

$$P_\epsilon(m_n, s_n) \asymp \mathrm{e}^{-nI_2(m_n, s_n;\epsilon)}, \tag{3.83}$$

for

$$I_2(m_n, s_n;\epsilon) = 2(1-m_n)I_\rho(s_n) + I_1(m_n;\epsilon). \tag{3.84}$$

Notice that the logarithm of these two probability distributions are in order $O(n)$.

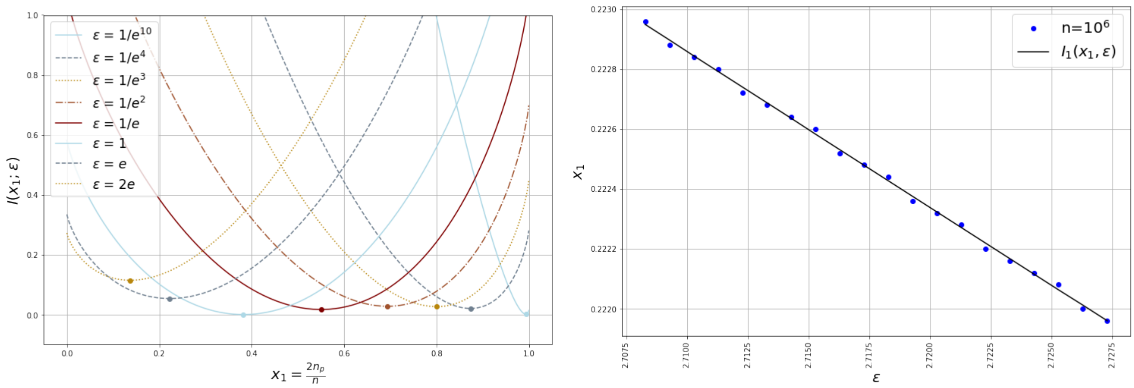


Figure 3.6: Left panel: $I_1(x_1;\epsilon)$ against $x_1 \in [0,1]$ for different ϵs. **Right panel:** The maximum of the distribution versus ϵ. Blue circles are the maxima of directly calculated distributions for $n = 10^6$. And the black line is the minimum of $I_1(x_1;\epsilon)$. Because of discreteness of the distribution, we see step-like changes in the position of maxima.

In figure (3.6), the left panel plots $I_1(m_n;\epsilon)$ and its minimum for different values of ϵ. Also, the right panel compares the estimated maximum of the probability distribution and directly calculated maximum with respect to ϵ in the vicinity of $\epsilon = e$. In the direct case, we took $n = 10^6$ and used $P_n(m_n, r = \epsilon n)$. The estimated value is the minimum of $I_1(m_n;\epsilon)$.

3.3 Statistical Properties of Distributions

Having probability distributions in closed form for pointed out statistical models enables us to derive their statistical properties and quantities. Apart from common statistics such as mean and standard deviation, we also obtain the asymptotic form of the same quantity for $1 \ll n$.

3.3.1 The closed form of $P_n(n_p)$ moments

As we shall see, different moments of the probability distribution $P_n(n_p)$ are expressible in closed form. Let us start with its first moment. Taking the expectation of n_p with respect to $P_n(n_p)$ finds

$$\langle n_p \rangle_n = \sum_{i=0}^{\lfloor n/2 \rfloor} i P_n(n_p = i) = \frac{1}{c_n(r)} \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{2i} (2i-1)!!\; i\; r^{\lfloor n/2 \rfloor - i}. \tag{3.85}$$

In appendix (B.5), we show that for the generating function $f_n(x)$, defined in equation (B.1), the first moment is equal to

$$\langle n_p \rangle_n = \frac{r^{\lfloor n/2 \rfloor}}{c_n(r)} \left[x \frac{d}{dx} \right] f_n(x) \Big|_{x=\frac{1}{r}}. \tag{3.86}$$

However equation(B.61) obtains

$$\frac{df_n(x)}{dx} = \frac{n(n-1)}{2} f_{n-2}(x). \tag{3.87}$$

Recall that equations (3.43) and (3.45) writes the normalisation constant in terms of $f_n(x)$, and consequently, in equation (B.64), the first moment is derived as

$$\langle n_p \rangle_n = \frac{n(n-1)}{2} \frac{c_{n-2}(r)}{c_n(r)}. \tag{3.88}$$

And also from $m_n = 2n_p/n$, last result asserts

$$\langle m_n \rangle_r = (n-1) \frac{c_{n-2}(r)}{c_n(r)}. \tag{3.89}$$

Remember that $c_n(r)$ is a polynomial, evaluated at r. Indeed, the first moment is the ratio of two polynomials times a constant.

3.3.2 The Asymptotic Form of the First Moment

In this part, we find the asymptotic form of $\langle m_n \rangle_n$ for $1 \ll n$. Equation (3.53) finds the asymptotic leading term of $c_n(r)$, so it writes

$$\frac{c_{n-2}(r)}{c_n(r)} \sim \frac{\frac{\mathrm{e}^{-\frac{r}{4}}}{\sqrt{2}} \left(\frac{n-2}{\mathrm{e}}\right)^{n/2-1} \mathrm{e}^{\sqrt{rn-2r}}}{\frac{\mathrm{e}^{-\frac{r}{4}}}{\sqrt{2}} \left(\frac{n}{\mathrm{e}}\right)^{n/2} \mathrm{e}^{\sqrt{rn}}} \sim \frac{\mathrm{e}^{-\sqrt{\frac{r}{n}}}}{n}, \tag{3.90}$$

and therefore, using equation (3.89), we get

$$\langle m_n \rangle_r \sim \mathrm{e}^{-\sqrt{\frac{r}{n}}}. \tag{3.91}$$

This last result is reasonable, since for $n \to \infty$, the expectation of m_n is one. This is the same result that we found for the large deviation distribution.

For $P_\epsilon(m_n)$, when we replace r/n by its limit, namely ϵ, we get

$$\langle m_n \rangle_\epsilon \sim \mathrm{e}^{-\sqrt{\epsilon}}. \tag{3.92}$$

The substitution of r/n by ϵ in the asymptotic leading term is justifiable from the fact that the function $f(x) = \mathrm{e}^{-\sqrt{x}}$ is analytic everywhere for non-zero x [7].

For $1 \ll n$, equation (3.53) provides accurate asymptotic expansion, as long as r is constant. But for the scaling case $\epsilon = r/n$, we use $c_n(\epsilon)$ asymptotic leading term in equation (3.75) that we rewrite it here

$$c_n(\epsilon) = \begin{cases} \frac{g(\epsilon)^{-\frac{n}{2}g(\epsilon)} \mathrm{e}^{n\sqrt{\frac{\epsilon}{\mathrm{e}}}} (\frac{n}{\mathrm{e}})^{n/2}}{\sqrt{\sqrt{\mathrm{e}\epsilon} f(\epsilon) g(\epsilon)}} & \epsilon \leq \mathrm{e} \\ \frac{g(\epsilon)^{-\frac{n}{2}g(\epsilon)} (\epsilon n)^{n/2}}{\sqrt{n\pi\epsilon f(\epsilon) g(\epsilon)}} & \epsilon > \mathrm{e} \end{cases}. \tag{3.93}$$

The assumption $1 \ll n$ asserts $\epsilon = r/n \sim r/(n-2)$, therefore, the ratio of the normalisation constants writes

$$\frac{c_{n-2}(\epsilon)}{c_n(\epsilon)} \sim \begin{cases} \frac{g(\epsilon)^{g(\epsilon)} \mathrm{e}^{-2\sqrt{\frac{\epsilon}{\mathrm{e}}}}}{n} & \epsilon \leq \mathrm{e} \\ \frac{g(\epsilon)^{g(\epsilon)}}{\epsilon n} & \epsilon > \mathrm{e} \end{cases}. \tag{3.94}$$

Hence, combining it with equation (3.89), we find

$$\langle m_n \rangle_\epsilon \sim \begin{cases} g(\epsilon)^{g(\epsilon)} \mathrm{e}^{-2\sqrt{\frac{\epsilon}{\mathrm{e}}}} & \epsilon \leq \mathrm{e} \\ \frac{g(\epsilon)^{g(\epsilon)}}{\epsilon} & \epsilon > \mathrm{e} \end{cases}. \tag{3.95}$$

So, we have three asymptotic cases and one solution of the minimum of $I_1(m_n;\epsilon)$:

1. For $\epsilon \to 0$, the first moment obtains as

$$\langle m_n \rangle_\epsilon \sim \mathrm{e}^{-\sqrt{\epsilon}}. \tag{3.96}$$

2. For $0 < \epsilon \leq \mathrm{e}$, the first moment writes as

$$\langle m_n \rangle_\epsilon \sim g(\epsilon)^{g(\epsilon)} \mathrm{e}^{-2\sqrt{\frac{\epsilon}{\mathrm{e}}}}. \tag{3.97}$$

3. For $\epsilon > \mathrm{e}$, it writes as

$$\langle m_n \rangle_\epsilon \sim \frac{g(\epsilon)^{g(\epsilon)}}{\epsilon}. \tag{3.98}$$

4. The minimum of $I_1(m_n;\epsilon)$ obtains $\langle m_n \rangle_\epsilon$ as

$$\frac{dI_1(m_n;\epsilon)}{dm_n} = 0 \implies \frac{d\tilde{H}_\epsilon(m_n)}{dm_n} = \ln m_n - 2\ln(1-m_n) + \ln\epsilon = 0 \implies$$

$$\langle m_n \rangle_\epsilon \sim g(\epsilon) = 1 - \frac{\epsilon f(\epsilon)}{2} \tag{3.99}$$

as $f(\epsilon)$ defined in equation (B.31).

Figure (3.7) plots the numerical comparison of the first case for $n = 10^6$. It is accurate for $\epsilon \ll 1$. Figure (3.8) plots the second and third cases, which are accurate for $\epsilon \ll 1$ and $\epsilon \gg 1$. And finally, figure (3.9) plots the fourth case or the minimum of $I_1(m_n;\epsilon)$. The last estimate is accurate for all values of ϵ.

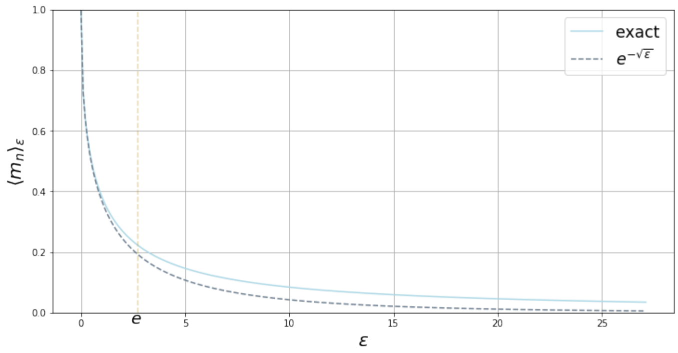


Figure 3.7: The comparison of exact and estimated $\langle m_n \rangle_\epsilon \sim \mathrm{e}^{-\sqrt{\epsilon}}$ for $\epsilon \in (0, 30]$, given $n = 10^6$.

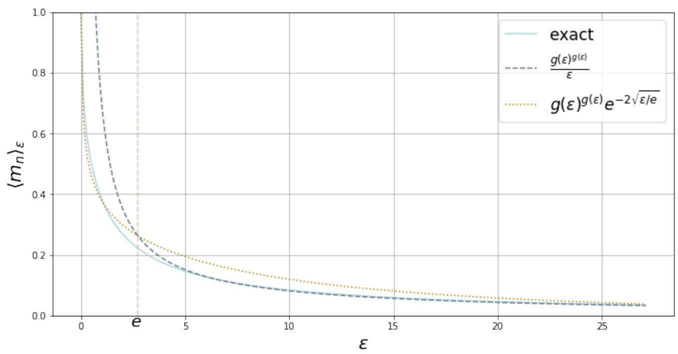


Figure 3.8: The comparison of exact and estimated $\langle m_n \rangle_\epsilon \sim g(\epsilon)^{g(\epsilon)}\mathrm{e}^{-2\sqrt{\frac{\epsilon}{\mathrm{e}}}}$ and $\langle m_n \rangle_\epsilon \sim \frac{g(\epsilon)^{g(\epsilon)}}{\epsilon}$ for $\epsilon \in (0, 30]$, given $n = 10^6$.

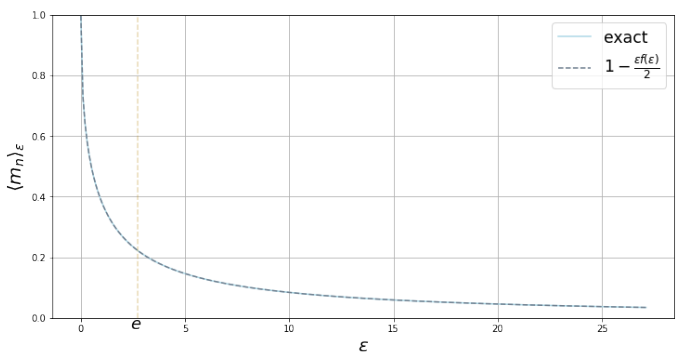


Figure 3.9: The comparison of exact and estimated $\langle m_n \rangle_\epsilon \sim 1 - \frac{\epsilon f(\epsilon)}{2}$ for $\epsilon \in (0, 30]$, given $n = 10^6$.

Putting all together, there are three choices that exist that one can select based on the situation. One asymptotic case is written as

$$\langle m_n \rangle_\epsilon \sim \begin{cases} \mathrm{e}^{-\sqrt{\epsilon}} & \epsilon \ll 1 \\ \frac{g(\epsilon)^{g(\epsilon)}}{\epsilon} & \epsilon \gg 1 \end{cases}, \tag{3.100}$$

or the second choice is

$$\langle m_n \rangle_\epsilon \sim \begin{cases} g(\epsilon)^{g(\epsilon)} \mathrm{e}^{-2\sqrt{\frac{\epsilon}{\mathrm{e}}}} & \epsilon \ll 1 \\ \frac{g(\epsilon)^{g(\epsilon)}}{\epsilon} & \epsilon \gg 1 \end{cases}. \tag{3.101}$$

Finally, the accurate asymptotic leading term for all values is written as

$$\langle m_n \rangle_\epsilon \sim 1 - \frac{\epsilon f(\epsilon)}{2} = 1 + \frac{\epsilon}{2} - \frac{\epsilon}{2}\sqrt{1 + \frac{4}{\epsilon}}. \tag{3.102}$$

3.3.3 Other Moments

The kth moment defines as the expectation of the kth power of n_p with respect to $P_n(n_p)$. In the previous part, we showed that applying the operator $\left[x\frac{d}{dx}\right]$ once on $f_n(x)$ (defined in equation (B.1)) finds the first moment. And applying $\left[x\frac{d}{dx}\right]$ k-times results in

$$\left[x\frac{d}{dx}\right]^k f_n(x) = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{2i} (2i-1)!!\; i^k\; x^i. \tag{3.103}$$

Evaluating the last result at $1/r$, the kth moment $\langle n_p^k \rangle_n$ must be

$$\langle n_p^k \rangle_n = \frac{r^{\lfloor n/2 \rfloor}}{c_n(r)} \left[x\frac{d}{dx}\right]^k f_n(x)\Bigg|_{x=\frac{1}{r}}. \tag{3.104}$$

In appendix (B.6), equation (B.68), we show that

$$\left[x\frac{d}{dx}\right]^k f_n(x)\Bigg|_{x=\frac{1}{r}} = \sum_{i=1}^{k} \frac{b_i^{(k)}}{2^i} \frac{c_{n-2i}(r)}{r^{\lfloor n/2 \rfloor}}, \tag{3.105}$$

where

$$b_i^{(k)} \equiv \frac{n!}{(n-2i)!} a_i^{(k)}, \tag{3.106}$$

and

$$a_i^{(k)} = a_{i-1}^{(k-1)} + i a_i^{(k-1)}, \qquad a_1^{(k)} = a_k^{(k)} = 1. \tag{3.107}$$

Hence, the kth moment writes as

$$\langle n_p^k \rangle_n = \frac{r^{\lfloor n/2 \rfloor}}{c_n(r)} \left[x \frac{d}{dx} \right]^k f_n(x) \Bigg|_{x=\frac{1}{r}} = \sum_{i=1}^{k} \frac{b_i^{(k)}}{2^i} \frac{c_{n-2i}(r)}{c_n(r)}. \tag{3.108}$$

3.3.4 The Asymptotic Form of the kth Moment

Appendix (B.9), equation (B.82) finds the asymptotic expansion of $\langle n_p^k \rangle_n$ as

$$\langle n_p^k \rangle_n \sim \frac{n^k \mathrm{e}^{-k\sqrt{r/n}}}{2^k} \left(1 + \frac{k(k-1)}{n} (\mathrm{e}^{\sqrt{r/n}} - 1) \right) + O(n^{k-2}), \tag{3.109}$$

and $\langle n_s^k \rangle_n = n - 2\langle n_p^k \rangle_n$ in equation (B.86) as

$$\langle n_s^k \rangle \sim n^k \left(1 - \mathrm{e}^{-\sqrt{r/n}} \right)^k \left[1 + \frac{k(k-1)\mathrm{e}^{-\sqrt{r/n}}}{n \left(1 - \mathrm{e}^{-\sqrt{r/n}} \right)} \right] + O(n^{k-2}). \tag{3.110}$$

3.3.5 A Relation Among First Moments with Different Sizes

There is an interesting property among the moments of systems in different sizes. In appendix (B.8), equation (B.79) shows that the kth moment for a system size n written in terms of smaller systems sizes as

$$\langle n_p^k \rangle_n = \sum_{i=1}^{k} a_i^{(k)} \langle n_p \rangle_n \langle n_p \rangle_{(n-2)} \ldots \langle n_p \rangle_{(n-2i+2)}. \tag{3.111}$$

For instance, we can use the last result to write $\mathrm{Var}[n_p]$ in terms of the expectation of smaller system sizes. The second moment writes as

$$\begin{aligned} \langle n_p^2 \rangle_n &= a_1^{(2)} \langle n_p \rangle_n + a_2^{(2)} \langle n_p \rangle_n \langle n_p \rangle_{(n-2)} \\ &= \langle n_p \rangle_n \left[1 + \langle n_p \rangle_{(n-2)} \right]. \end{aligned} \tag{3.112}$$

Hence

$$\mathrm{Var}[n_p]_n = \langle n_p^2 \rangle_n - \langle n_p \rangle_n^2$$

$$= \langle n_p \rangle_n \left[1 + \langle n_p \rangle_{(n-2)} - \langle n_p \rangle_n \right]. \tag{3.113}$$

3.3.6 Probability Generating Function of $P_n(n_p)$

Appendix (B.10) finds the probability generating function for random variables n_p and n_h. Recall that the normalization constant $c_n(r)$ is a polynomial degree n, evaluates at r. For n_p, the probability generating function is

$$G_n(s) = s^{\frac{n}{2}} \frac{c_n(\frac{r}{s})}{c_n(r)}, \tag{3.114}$$

where $c_n(\frac{r}{s})$ is the normalization constant, evaluates at $\frac{r}{s}$.

Similarly, for the probability generating function for random variables n_p and n_h is

$$G_n(s, u) = s^{\frac{n}{2}} \frac{c_n\left(\frac{r(\rho u + 1 - \rho)^2}{s}\right)}{c_n(r)}. \tag{3.115}$$

The normalization constant in the numerator evaluates at $\left(\frac{r(\rho u+1-\rho)^2}{s}\right)$.

3.4 Marginal Distributions

For a system with size n, a single random variable, say X_l, is a function of states that projects the state of an element at index l to its domain. For example, for the B-Model, X_l is defined as

$$X_l : \mathcal{S}_n \to \{0, 1\}, \tag{3.116}$$

where $\mathcal{S}_n$ is the set of all configurations with length n. Here, we use the following convention: $X_l = 0$ represents a pair state and $X_l = 1$ represents a stand-alone state. In contrast, X_l for the B-model defines over the set of configurations, namely $\mathcal{S'}_n$, as

$$X_l : \mathcal{S'}_n \to \{-1, 0, 1\}, \tag{3.117}$$

such that $X_l = -1$ represents a tail state, $X_l = 0$ a pair state and $X_l = 1$ a head state.

To find the marginal distribution, clearly, it suffices to sum all probabilities of the states for which X_l is the same

$$P_n(X_l = s) = \sum_{c \in \mathcal{S}_n : X_l(c) = s} P_n(c). \tag{3.118}$$

Therefore, partitioning the state space $\mathcal{S}_n$ according to the state of an element at index l is the key to finding the marginal probability. Intuitively, we understand that the marginal is invariant with respect to l, so the index l is arbitrary.

3.4.1 B-model's marginal

Let us start with the B-model. Recall that in the previous section, we defined S_i as a subset of $\mathcal{S}_n$ that its member configurations have i pairs. We partition the elements of S_i like

$$S_i = S_i^{(1)} \bigcup S_i^{(2)}, \qquad S_i^{(1)} \bigcap S_i^{(2)} = \emptyset, \tag{3.119}$$

such that $S_i^{(1)}$ contains only the configurations that do not have a paired link with the element at index l – figure (3.10) – whereas $S_i^{(2)}$ contains those configurations that have one – figure (3.11). In other words, for every configuration in $S_i^{(1)}$, the ball at l is in stand-alone state, whilst for configurations belong to $S_i^{(2)}$, the ball at l is in pair state.

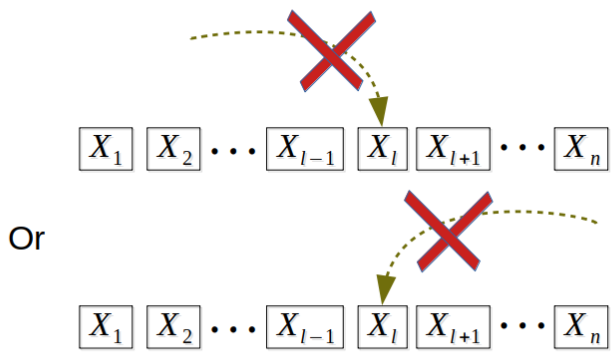


Figure 3.10: An example of configurations in subset $S_i^{(1)}$.

Next, we partition $S_i^{(2)}$ to disjoint subsets like

$$S_i^{(2)} = \bigcup_{\substack{k=1 \\ k \neq l}}^{n} S_{i,k}^{(2)}, \qquad S_{i,k}^{(2)} \bigcap S_{i,h}^{(2)} = \emptyset \quad \text{for} \quad k \neq h, \tag{3.120}$$

such that, $S_{i,k}^{(2)}$ contains the configurations that has a pair state between index k and l.

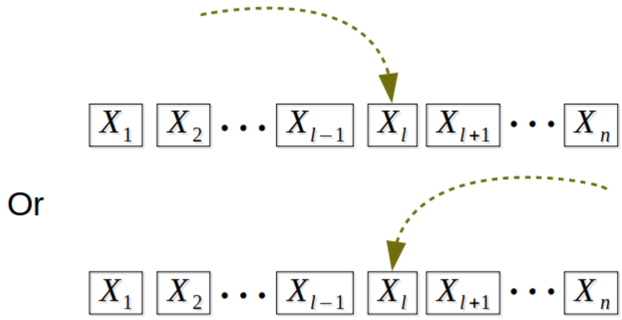


Figure 3.11: An example of configurations in subset $S_i^{(2)}$.

Finally, the partitioning of $\mathcal{S}_n$ enables us to rewrite the marginal sum as

$$P_n(\{\text{the index } l \text{ is in stand-alone state}\}) = P_n(X_l = 1) = \sum_{i=0}^{\lfloor n/2 \rfloor} P(S_i^{(1)}), \tag{3.121}$$

and

$$P_n(\{\text{the index } l \text{ is in pair state}\}) = P_n(X_l = 0) = \sum_{i=0}^{\lfloor n/2 \rfloor} \sum_{\substack{k=1 \\ k \neq l}}^{n} P(S_{i,k}^{(2)}). \tag{3.122}$$

We have already shown that the probability of each configuration in S_i is equal to

$$P_n(c \in S_i) = \frac{r^{\lfloor n/2 \rfloor - i}}{c_n(r)}, \tag{3.123}$$

subsequently, to find the marginal sums, the cardinalities of $S_i^{(1)}$ and $S_{i,k}^{(2)}$ are required. From its definition, $S_i^{(1)}$ contains i pairs, whilst the index l is in stand-alone state. So, there are $n-1$ ways to choose $2i$ pairs among all the balls except the one at l, and there are $(2i-1)!!$ permutations among i pairs – see figure (3.12). It concludes that the cardinality of $S_i^{(1)}$ must be

$$|S_i^{(1)}| = \binom{n-1}{2i}(2i-1)!!. \tag{3.124}$$

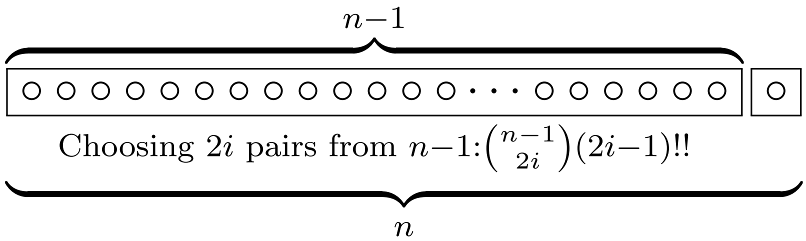


Figure 3.12: $|S_i^{(1)}|$ cardinality.

In finding the cardinality of $S_{i,k}^{(2)}$, we see that one of the pairs has already been selected, so there are $2(i-1)$ choices from $n-2$ candidates, and $(2i-3)!!$ permutations among them – see figure (3.13). It gets

$$|S_{i,k}^{(2)}| = \binom{n-2}{2i-2}(2i-3)!!. \tag{3.125}$$

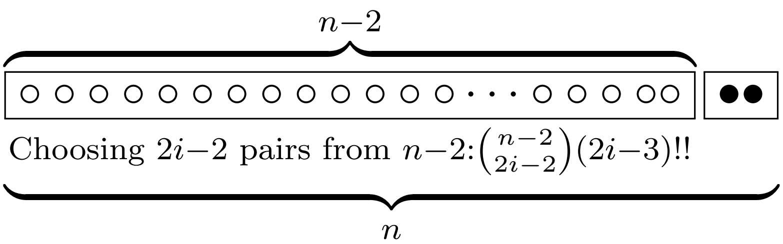


Figure 3.13: $|S_{i,k}^{(2)}|$ cardinality.

Eventually, the marginal for a stand-alone state writes as

$$P_n(X_l = 1) = \frac{1}{c_n(r)} \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n-1}{2i}(2i-1)!!\, r^{\lfloor n/2 \rfloor - i}$$

$$= \frac{1}{c_n(r)} \sum_{i=0}^{\lfloor n/2 \rfloor} \frac{n-2i}{n} \binom{n}{2i} (2i-1)!! \, r^{\lfloor n/2 \rfloor - i}$$
$$= 1 - \frac{2\langle n_p \rangle}{n}, \tag{3.126}$$

where we used the definition of the normalisation constant and the expectation of n_p in the last step. Similarly, for a pair state

$$P_n(X_l = 0) = \frac{1}{c_n(r)} \sum_{i=0}^{\lfloor n/2 \rfloor} \sum_{\substack{k=1 \\ k \neq l}}^{n} \binom{n-2}{2i-2} (2i-3)!! \, r^{\lfloor n/2 \rfloor - i}$$

$$= \frac{1}{c_n(r)} \sum_{i=0}^{\lfloor n/2 \rfloor} (n-1) \binom{n-2}{2i-2} (2i-3)!! \, r^{\lfloor n/2 \rfloor - i}$$
$$= \frac{1}{c_n(r)} \sum_{i=0}^{\lfloor n/2 \rfloor} \frac{2i}{n} \binom{n}{2i} (2i-1)!! \, r^{\lfloor n/2 \rfloor - i} = \frac{2\langle n_p \rangle}{n}. \tag{3.127}$$

Note that

$$P_n(X_l = 0) + P_n(X_l = 1) = 1, \tag{3.128}$$

by which, it assures the validity of the final result. In short, the marginal distribution is

$$P_n(X_l) = \begin{cases} \frac{2\langle n_p \rangle}{n} & , X_l = 0 \\ 1 - \frac{2\langle n_p \rangle}{n} & , X_l = 1 \end{cases}, \tag{3.129}$$

For $n \gg 1$, using equation (3.91) we get

$$P_n(X_l) = \begin{cases} e^{-\sqrt{\frac{r}{n}}} & , X_l = 0 \\ 1 - e^{-\sqrt{\frac{r}{n}}} & , X_l = 1 \end{cases}. \tag{3.130}$$

3.4.2 C-model's marginal

For the C-model, we start by partitioning S_{ij} to disjoint subsets. Recall that configurations in the set S_{ij} contain i pairs and j head coins. Indeed, similar to the previous section, S_{ij} partitions into two subsets

$$S_{ij} = S_{ij}^{(1)} \bigcup S_{ij}^{(2)}, \qquad S_{ij}^{(1)} \bigcap S_{ij}^{(2)} = \emptyset, \tag{3.131}$$

and further, $S_{ij}^{(2)}$ partitions to more subsets that, for each, there is a pair links between index i and k such that

$$S_{ij}^{(2)} = \bigcup_{\substack{k=1 \\ k\neq l}}^{n} S_{ij,k}^{(2)}, \qquad S_{ij,k}^{(2)} \bigcap S_{ij,h}^{(2)} = \emptyset \quad \text{for} \quad k \neq h. \tag{3.132}$$

As a matter of fact, $S_{ij}^{(1)}$, $S_{ij}^{(2)}$ and $S_{ij,h}^{(2)}$ are similar to the subsets in the previous section, except the j index includes the number of head states. Furthermore, for the B-model, we have to partitions $S_{ij}^{(1)}$ to two subsets

$$S_{ij}^{(1)} = S_{ij,h}^{(1)} \bigcup S_{ij,t}^{(1)}, \qquad S_{ij,h}^{(1)} \bigcap S_{ij,t}^{(1)} = \emptyset, \tag{3.133}$$

where $S_{ij,h}^{(1)}$ and $S_{ij,t}^{(1)}$ contain stand-alone element in head and tail state at index l, respectively.

Thus, one by one, we write the marginals of the element l in a different state. For a head state

$$P_n(\{\text{the site } l \text{ is in head state}\}) = P_n(X_l = 1) = \sum_{i=0}^{\lfloor n/2 \rfloor} \sum_{j=0}^{n-2i} P(S_{ij,h}^{(1)}), \tag{3.134}$$

for a tail state

$$P_n(\{\text{the site } l \text{ is in tail state}\}) = P_n(X_l = -1) = \sum_{i=0}^{\lfloor n/2 \rfloor} \sum_{j=0}^{n-2i} P(S_{ij,t}^{(1)}), \tag{3.135}$$

and finally for a pair state

$$P_n\{\text{the site } l \text{ is in pair state}\} = P_n(X_l = 0) = \sum_{i=0}^{\lfloor n/2 \rfloor} \sum_{j=0}^{n-2i} \sum_{\substack{k=1 \\ k\neq l}}^{n} P(S_{ij,k}^{(2)}). \tag{3.136}$$

The cardinality of these subsets are – see figures (3.14), (3.15) and (3.16) –

$$|S_{ij,h}^{(1)}| = \binom{n-1}{2i} (2i-1)!! \binom{n-2i-1}{j-1}, \tag{3.137}$$

$$|S_{ij,t}^{(1)}| = \binom{n-1}{2i} (2i-1)!! \binom{n-2i-1}{n-2i-j}, \tag{3.138}$$

and

$$|S^{(2)}_{ij,k}| = \binom{n-2}{2i-2}(2i-3)!! \binom{n-2i}{n-2i-j}. \tag{3.139}$$

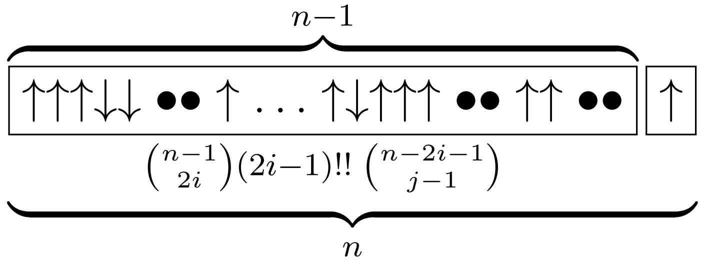


Figure 3.14: $|S^{(1)}_{ij,h}|$ cardinality.

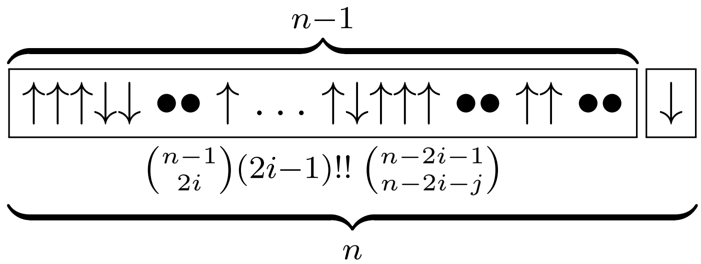


Figure 3.15: $|S^{(1)}_{ij,t}|$ cardinality.

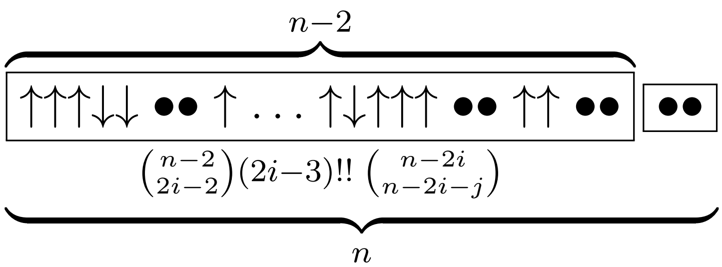


Figure 3.16: $|P(S^{(2)}_{ij,k})|$ cardinality.

Remember that the probability of observing a configuration in $S_{i,j}$ is

$$P_n(c \in S_{i,j}) = \frac{r^{\lfloor n/2 \rfloor - i} \rho^j (1-\rho)^{n-2i-j}}{c_n(r)}. \tag{3.140}$$

Hence, using the calculated cardinalities, we get

$$P_{2n}(X_l = 1) = \frac{1}{c_n(r)} \sum_{i=0}^{\lfloor n/2 \rfloor} \sum_{j=0}^{n-2i} \binom{n-1}{2i} (2i-1)!! \binom{n-2i-1}{j-1} r^{\lfloor n/2 \rfloor - i} \rho^j (1-\rho)^{n-2i-j}$$

$$= \frac{1}{c_n(r)} \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n-1}{2i} (2i-1)!! \frac{r^{\lfloor n/2 \rfloor - i}}{n-2i} (n-2i)\rho,$$

$$= \rho \left(1 - \frac{2\langle n_p \rangle_n}{n}\right), \tag{3.141}$$

where we used the Binomial distribution's mean value in the second step, and used the result from the previous case for $X_l = 1$. Moreover, $P_n(X_l = -1)$ finds similar result, and we get

$$P_n(X_l = -1) = (1-\rho) \left(1 - \frac{2\langle n_p \rangle}{n}\right). \tag{3.142}$$

Finally, for $P_n(X_l = 0)$

$$P_n(X_l = 0) = \sum_{i=0}^{\lfloor n/2 \rfloor} \sum_{j=0}^{n-2i} \sum_{\substack{k=1 \\ k \neq l}}^{n} P(S_{ij,k}^{(2)})$$

$$= \frac{1}{c_n(r)} \sum_{i=0}^{\lfloor n/2 \rfloor} \sum_{j=0}^{n-2i} \sum_{\substack{k=1 \\ k \neq l}}^{n} \binom{n-2}{2i-2} (2i-3)!! \binom{n-2i}{n-2i-j} r^{\lfloor n/2 \rfloor - i} \rho^j (1-\rho)^{n-2i-j}$$

$$= \frac{1}{c_n(r)} \sum_{i=0}^{\lfloor n/2 \rfloor} (n-1) \binom{n-2}{2i-2} (2i-3)!!\, r^{\lfloor n/2 \rfloor - i} \sum_{j=0}^{n-2i} \binom{n-2i}{n-2i-j} \rho^j (1-\rho)^{n-2i-j}$$

$$= \frac{1}{c_n(r)} \sum_{i=0}^{\lfloor n/2 \rfloor} (n-1) \binom{n-2}{2i-2} (2i-3)!!\, r^{\lfloor n/2 \rfloor - i}, \tag{3.143}$$

using the Binomial expansion in the last step. For this result, the sum has already been calculated for the stand-alone case, and we know

$$P_n(X_l = 0) = \frac{2\langle n_p \rangle}{n}. \tag{3.144}$$

Therefore, the marginal is

$$P_n(X_l) = \begin{cases} (1-\rho)\left(1 - \frac{2\langle n_p\rangle}{n}\right) & , X_l = -1 \\ \frac{2\langle n_p\rangle}{n} & , X_l = 0 \\ \rho\left(1 - \frac{2\langle n_p\rangle}{n}\right) & , X_l = 1 \end{cases} . \tag{3.145}$$

Also, for $n \gg 1$, the asymptotic result of the averages in equation (3.91) finds

$$P_n(X_l) = \begin{cases} (1-\rho)\left(1 - e^{-\sqrt{\frac{r}{n}}}\right) & , X_l = -1 \\ e^{-\sqrt{\frac{r}{n}}} & , X_l = 0 \\ \rho\left(1 - e^{-\sqrt{\frac{r}{n}}}\right) & , X_l = 1 \end{cases} . \tag{3.146}$$

3.5 Joint Probability Distributions

In section (3.4), we described in detail how one could find the marginal distributions of a single element for both B and C models. In this section, we will derive the joint probability distribution of two or more elements in a system with the size n. However, we shall focus on the B-model only and provide the combinatorial arguments in appendices.

To begin, let us start with the joint probability distribution of the state of a single element at an arbitrary index and the number of pairs in the whole system, denoted by $P_n(X_1, n_p)$. As usual, $X_1 = 0$ refers to a ball in pair state and $X_1 = 1$ to stand-alone state. It must be intuitively clear that the resulting distribution is the same for all elements, and X_1 can refer to an element at any index. Also, we suppose n_p counts the number of pairs, including the state of X_1.

Appendix (B.12.1) explains it in detail, and we do not reiterate the steps of the derivation of the joint, conditional and marginal distributions here. Altogether, it suffices to write them as

$$P_n(X_1, n_p) = \frac{r^{\lfloor n/2 \rfloor}(2n_p - 1)!!}{c_n(r)} \times \begin{cases} \binom{n-1}{2n_p-1} & , X_1 = 0 \\ \binom{n-1}{2n_p} & , X_1 = 1 \end{cases} , \tag{3.147}$$

$$P_n(X_1|n_p) = \frac{P_n(X_1, n_p)}{P_n(n_p)} = \begin{cases} \frac{2n_p}{n} & , X_1 = 0 \\ \frac{n-2n_p}{n} & , X_1 = 1 \end{cases} , \tag{3.148}$$

$$P_n(X_1) = \begin{cases} \langle\frac{2n_p}{n}\rangle & , X_1 = 0 \\ \langle\frac{n-2n_p}{n}\rangle & , X_1 = 1 \end{cases}, \tag{3.149}$$

to reveal the pattern that we can exploit in the following parts. This pattern becomes comprehensible when we derive the equivalent distributions for two elements, as we do in appendix (B.12.2)

$$P_n(X_1, X_2, n_p) = \frac{r^{\lfloor n/2\rfloor - n_p}(2n_p-1)!!}{c_n(r)} \times \begin{cases} \binom{n-2}{2n_p-2} & , X_l = X_k = 0 \\ \binom{n-2}{2n_p-1} & , X_l = 1, X_k = 0 \\ \binom{n-2}{2n_p-1} & , X_l = 0, X_k = 1 \\ \binom{n-2}{2n_p} & , X_l = X_k = 1 \end{cases}, \tag{3.150}$$

$$P_n(X_1, X_2|n_p) = \begin{cases} \frac{2n_p}{n} \times \frac{2n_p-1}{n-1} & , X_l = X_k = 0 \\ \frac{2n_p}{n} \times \frac{n-2n_p}{n-1} & , X_l = 1, X_k = 0 \\ \frac{2n_p}{n} \times \frac{n-2n_p}{n-1} & , X_l = 0, X_k = 1 \\ \frac{n-2n_p}{n} \times \frac{n-2n_p-1}{n-1} & , X_l = X_k = 1 \end{cases}, \tag{3.151}$$

$$P_n(X_1, X_2) = \begin{cases} \langle\frac{2n_p(2n_p-1)}{n(n-1)}\rangle & , X_l = X_k = 0 \\ \langle\frac{2n_p(n-2n_p)}{n(n-1)}\rangle & , X_l = 1, X_k = 0 \\ \langle\frac{2n_p(n-2n_p)}{n(n-1)}\rangle & , X_l = 0, X_k = 1 \\ \langle\frac{(n-2n_p)(n-2n_p-1)}{n(n-1)}\rangle & , X_l = X_k = 1 \end{cases}. \tag{3.152}$$

Observe that the difference between $P_n(X_1, n_p)$ and $P_n(X_1, X_2, n_p)$ for different combinations of X_1 and X_2 is in a cardinality factor, which is found by an combinatorial argument in appendix (B.12.2). Besides, the conditional distributions, namely $P_n(X_1|n_p)$ and $P_n(X_1, X_2|n_p)$, find as the ratio of the cardinality factor divided by $\binom{n}{2n_p}$. And finally, the marginal distribution is the average of ratios with respect to $P_n(n_p)$.

In general, to write the joint distribution for k elements, denoted by $P_n(X_1, \ldots, X_k, n_p)$, let us first define $l = k - \sum_{i=1}^n X_i$ as the number of zeros in a given k-tuple $(X_1, \ldots, X_k)$. Note that given l, it is possible to have $\binom{k}{l}$ distinct k-tuples. To elaborate, $\sum_{i=1}^n X_i$ is the number of ones in the k-tuple, or the number of stand-alone elements, therefore, $k - \sum_{i=1}^n X_i$ is the number of pairs. So, $\binom{k}{l}$ is the number of distinct arrangement of l zeros among k places.

As it is explained in appendix (B.12.3), for $l = k - \sum_{i=1}^n X_i$, the joint distribution derives as

$$P_n(X_1,\ldots,X_k,n_p;l) = \frac{r^{\lfloor n/2\rfloor - n_p}}{c_n(r)}\binom{n-k}{2n_p-l}(2n_p-1)!!. \tag{3.153}$$

Similarly, the conditional distribution is

$$P_n(X_1,\ldots,X_k|n_p) = \frac{(2n_p)^{(l)}(n-2n_p)^{(k-l)}}{n^{(k)}}, \tag{3.154}$$

where $x^{(l)} = x(x-1)\ldots(x-k+1)$ is a falling factorial, and the marginal is

$$P_n(X_1,\ldots,X_k) = \langle\frac{(2n_p)^{(l)}(n-2n_p)^{(k-l)}}{n^{(k)}}\rangle, \tag{3.155}$$

where the expectation is taken with respect to $P_n(n_p)$.

It is worth representing in more detail the case of $\binom{k}{l}$ distinct arrangements of k-tuples, given l. In fact, the degeneracy for a given l is represented as $\binom{k}{l}$ similar entries in the definition of $P_n(X_1,\ldots,X_k,n_p)$ in the following

$$P_n(X_1,\ldots,X_k,n_p) =$$

$$\frac{r^{\lfloor n/2\rfloor-n_p}(2n_p-1)!!}{c_n(r)}\times\begin{cases}\binom{n-k}{2n_p} & ,X_1=0,X_2=0,\ldots,X_k=0\\ \left.\begin{array}{ll}\binom{n-k}{2n_p-1} & ,X_1=1,X_2=0,\ldots,X_k=0\\ \binom{n-k}{2n_p-1} & ,X_1=0,X_2=1,\ldots,X_k=0\\ \vdots & \\ \binom{n-k}{2n_p-1} & ,X_1=0,X_2=0,\ldots,X_k=1\end{array}\right\} k \text{ times}\\ \vdots\\ \left.\begin{array}{ll}\binom{n-k}{2n_p-l} & ,X_1,X_2,\ldots,X_k\\ \binom{n-k}{2n_p-l} & ,X_1,X_2,\ldots,X_k\\ \vdots & \\ \binom{n-k}{2n_p-l} & ,X_1,X_2,\ldots,X_k\end{array}\right\} \binom{k}{l} \text{ times for} & l=k-\sum_{i=1}^n X_i\\ \vdots\\ \binom{n-k}{2n_p-k} & ,X_1=1,X_2=1,\ldots,X_k=1\end{cases}. \tag{3.156}$$

3.6 Parameters' Maximum Likelihood Estimation

So far, all the probability distributions that we have derived depends on one or two parameters, *i.e.* r for $P_n(n_p)$, r and ρ for $P_n(n_p, n_h)$, and ϵ for $P_\epsilon(m_n)$. In this section, we will derive the maximum likelihood estimation (MLE) of these parameters whenever one or more empirical values of random variables are observed.

3.6.1 MLE for the B-model

We start by finding the maximum likelihood estimation of r, when K independent $n_p^{(i)}$s for a pairing ball system are observed such that their joint probability distribution is $P_n(n_p^{(1)}, \ldots, n_p^{(K)})$. Meanwhile, we assume the consecutive observed random variables are identical and statistically independent, therefore

$$P_n(n_p^{(1)}, \ldots, n_p^{(K)}) = \prod_{i=1}^{K} P_n(n_p^{(i)}). \tag{3.157}$$

To maximise the logarithm of the joint probability for the K observed random variables, we first take the derivative of $\ln P_n(n_p^{(1)}, \ldots, n_p^{(K)})$ with respect to the parameter r

$$\frac{d}{dr} \ln P_n(n_p^{(1)}, \ldots, n_p^{(K)}) = \sum_{i=1}^{K} \frac{d}{dr} \ln P_n(n_p^{(i)}). \tag{3.158}$$

Equation (B.18) finds $\ln P_n(n_p^{(i)})$ as

$$\frac{d}{dr} \ln P_n(n_p^{(i)}) = -\frac{\frac{d}{dr} c_n(r)}{c_n(r)} + \frac{\lfloor n/2 \rfloor - n_p^{(i)}}{r}. \tag{3.159}$$

Also, equation (B.74) in appendix (B.7) finds

$$\frac{d}{dr} c_n(r) = \frac{\lfloor n/2 \rfloor}{r} c_n(r) - \frac{n(n-1)}{2r} c_{n-2}(r), \tag{3.160}$$

and therefore, equation (3.159) becomes

$$\frac{d}{dr} \ln P_n(n_p^{(i)}) = \frac{n(n-1)}{2r} \frac{c_{n-2}(r)}{c_n(r)} - \frac{n_p^{(i)}}{r}. \tag{3.161}$$

To find r_{MLE}, we set the derivative of $\ln P_n(n_p^{(1)}, \dots, n_p^{(K)})$ equal to zero to find its solution with respect to r. To do that, using the last result, equation (3.158) writes as

$$\frac{d}{dr} \ln P_n(n_p^{(1)}, \dots, n_p^{(K)}) = \sum_{i=1}^{K} (\frac{n(n-1)}{2r} \frac{c_{n-2}(r)}{c_n(r)} - \frac{n_p^{(i)}}{r})$$

$$= K \frac{n(n-1)}{2r} \frac{c_{n-2}(r)}{c_n(r)} - \frac{1}{r} \sum_{i=1}^{K} n_p^{(i)} = 0. \tag{3.162}$$

Denoting the empirical mean by $\hat{n}_p$ that is defined as

$$\hat{n}_p \equiv \frac{1}{K} \sum_{i=1}^{K} n_p^{(i)}, \tag{3.163}$$

and equation (3.162) simplifies to

$$2\hat{n}_p c_n(r) - n(n-1) c_{n-2}(r) = 0. \tag{3.164}$$

The above equation is a polynomial degree $\lfloor n/2 \rfloor$, and r_{MLE} is its root.

3.6.2 MLE for the C-model

For the B-model, the observed tuple $(n_p^{(i)}, n_h^{(i)})$ for a single observation provides

$$\frac{\partial}{\partial r} \ln P_n(n_p^{(i)}, n_h^{(i)}) = -\frac{\frac{d}{dr} c_n(r)}{c_n(r)} + \frac{n - 2n_p^{(i)}}{2r}, \tag{3.165}$$

and

$$\frac{\partial}{\partial \rho} \ln P_n(n_p^{(i)}, n_h^{(i)}) = \frac{n_h^{(i)}}{\rho} - \frac{n - 2n_p^{(i)} - n_h^{(i)}}{1-\rho}. \tag{3.166}$$

Simultaneously setting both equations equal to zero finds parameters that maximise the distribution. Each equation has only one parameter, and consequently, the first one results in a similar polynomial, by which r_{MLE} can be calculated

$$n(n-1) c_{n-2}(r) - 2\hat{n}_p c_n(r) = 0. \tag{3.167}$$

Moreover, the second equation finds

$$\frac{\sum_{i=1}^{K} n_h^{(i)}}{\rho} - \frac{\sum_{i=1}^{K} (n - 2n_p^{(i)} - n_h^{(i)})}{1-\rho} = 0 \implies$$

$$\rho_{MLE} = \frac{\hat{n}_h}{n - 2\hat{n}_p}, \tag{3.168}$$

for

$$\hat{n}_p \equiv \frac{1}{K}\sum_{i=1}^{K} n_p^{(i)}, \qquad \hat{n}_h \equiv \frac{1}{K}\sum_{i=1}^{K} n_h^{(i)}. \tag{3.169}$$

Therefore, both the average number of pairs and heads, namely $\hat{n}_p$ and $\hat{n}_h$, are sufficient statistics.

3.6.3 The Asymptotics of r_{MLE} and ρ_{MLE}

For $n \gg 1$ and constant r, equation (3.90) derived the ratio of normalisation constant as

$$\frac{c_{n-2}(r)}{c_n(r)} \sim \frac{e^{-\sqrt{\frac{r}{n}}}}{n}. \tag{3.170}$$

Note that the estimate is accurate when r is not in order of $O(n)$. So for $1 \ll n$, we have

$$\begin{aligned}\hat{n}_p &= \frac{n(n-1)}{2}\frac{c_{n-2}(r)}{c_n(r)} = \frac{(n-1)}{2}e^{-\sqrt{\frac{r}{n}}} \\ &\sim \frac{n}{2}e^{-\sqrt{\frac{r}{n}}} + O(\frac{1}{n}).\end{aligned} \tag{3.171}$$

Rearranging the terms to find r_{MLE} for a system size n obtains

$$r_{MLE} = n\ln^2\left(\frac{2\hat{n}_p}{n}\right) \implies r_{MLE} = n\ln^2\hat{m}_n. \tag{3.172}$$

And similarly

$$\rho_{MLE} = \frac{\frac{\hat{n}_h}{n}}{1 - \frac{2\hat{n}_p}{n}} \implies \rho_{MLE} = \frac{\hat{s}_n}{1 - \hat{m}_n}, \tag{3.173}$$

are maximum likelihood estimates for given empirical means as

$$\hat{m}_n = \frac{2\hat{n}_p}{n}, \qquad \hat{s}_n = \frac{\hat{n}_h}{n - 2\hat{n}_p}. \tag{3.174}$$

3.6.4 MLE for $P_\epsilon(m_n)$

In the previous part, we found the maximum likelihood estimation of r for the asymptotic case $1 \ll n$ and constant r. Recall that when r is system size dependent such that $\lim_{n\to\infty} r/n = \epsilon$ exists, the probability distribution of the random variable

m_n is $P_\epsilon(m_n)$. In this part, we will derive the maximum likelihood estimation of ϵ.

Let us start by assuming a single observed value, say $m_n^{(i)}$, is available for estimating the parameter ϵ. The normalised probability distribution , $P_\epsilon(m_n^{(i)})$, writes as

$$P_\epsilon(m_n^{(i)}) = \frac{\mathrm{e}^{-nI_1(m_n^{(i)};\epsilon)}}{\int_0^1 \mathrm{e}^{-nI_1(m_n;\epsilon)}dm_n}, \tag{3.175}$$

and therefore, taking the derivative of $\ln P_\epsilon(m_n^{(1)})$ with respect to ϵ finds

$$\frac{\partial}{\partial\epsilon}\ln P_\epsilon(m_n^{(i)}) = -n\frac{\partial I_1(m_n^{(i)};\epsilon)}{\partial\epsilon} + n\frac{\int_0^1 \frac{\partial I_1(m_n;\epsilon)}{\partial\epsilon}\mathrm{e}^{-nI_1(m_n;\epsilon)}dm_n}{\int_0^1 \mathrm{e}^{-nI_1(m_n;\epsilon)}dm_n} = 0 \implies$$

$$\frac{\partial I_1(m_n^{(i)};\epsilon)}{\partial\epsilon} = \langle\frac{\partial I_1(m_n;\epsilon)}{\partial\epsilon}\rangle_\epsilon, \tag{3.176}$$

where $\langle.\rangle_\epsilon$ is the expectation with respect to $P_n(m_n;\epsilon)$.

Similarly, for K statistically independent observed values, the joint probability distribution writes as

$$P_\epsilon(m_n^{(1)},\ldots,m_n^{(K)}) = \prod_{i=1}^{K} P_\epsilon(m_n^{(i)}), \tag{3.177}$$

and consequently,

$$\frac{\partial}{\partial\epsilon}\ln P_\epsilon(m_n^{(1)},\ldots,m_n^{(K)}) = -n\sum_{i=1}^{K}\frac{\partial I_1(m_n^{(i)};\epsilon)}{\partial\epsilon} + Kn\frac{\int_0^1 \frac{\partial I_1(m_n;\epsilon)}{\partial\epsilon}e^{-nI_1(m_n;\epsilon)}dm_n}{\int_0^1 e^{-nI_1(m_n;\epsilon)}dm_n} = 0 \implies$$

$$\sum_{i=1}^{K}\frac{\partial I_1(m_n^{(i)};\epsilon)}{\partial\epsilon} = K\langle\frac{\partial I_1(m_n;\epsilon)}{\partial\epsilon}\rangle_\epsilon. \tag{3.178}$$

Using equation (3.82), we find

$$\frac{\partial I_1(m_n;\epsilon)}{\partial\epsilon} = \frac{1}{2}\begin{cases} \frac{1}{\sqrt{\mathrm{e}\epsilon}} + \frac{m_n-1}{\epsilon} - g'(\epsilon)(1+\ln g(\epsilon)) & 0<\epsilon\le\mathrm{e} \\ \frac{m_n}{\epsilon} - g'(\epsilon)(1+\ln g(\epsilon)) & \epsilon>\mathrm{e} \end{cases}, \tag{3.179}$$

and the sum on the left-hand side of equation (3.178) written as

$$\sum_{i=1}^{K}\frac{\partial I_1(m_n^{(i)};\epsilon)}{\partial\epsilon} = \frac{K}{2}\begin{cases} \frac{\hat{m}_n-1}{\epsilon} + \frac{1}{\sqrt{\mathrm{e}\epsilon}} - g'(\epsilon)(1+\ln g(\epsilon)) & 0<\epsilon\le\mathrm{e} \\ \frac{\hat{m}_n}{\epsilon} - g'(\epsilon)(1+\ln g(\epsilon)) & \epsilon>\mathrm{e} \end{cases}, \tag{3.180}$$

where $\hat{m}_n$ is the empirical mean of the observed values

$$\hat{m}_n \equiv \frac{1}{K}\sum_{i=1}^{K} m_n^{(i)}. \tag{3.181}$$

Next, the expectation of equation (3.179) gets

$$\langle\frac{\partial I_1(m_n;\epsilon)}{\partial\epsilon}\rangle_\epsilon = \frac{1}{2}\begin{cases} \frac{\langle m_n\rangle_\epsilon - 1}{\epsilon} + \frac{1}{\sqrt{\mathrm{e}\epsilon}} - g'(\epsilon)(1+\ln g(\epsilon)) & 0<\epsilon\leq \mathrm{e} \\ \frac{\langle m_n\rangle_\epsilon}{\epsilon} - g'(\epsilon)(1+\ln g(\epsilon)) & \epsilon > \mathrm{e} \end{cases}. \tag{3.182}$$

Using the last result and equation (3.180) allows one to equate both sides of equation (3.178) and derives

$$\langle m_n\rangle_\epsilon = \hat{m}_n, \tag{3.183}$$

for both branches of $\epsilon \leq \mathrm{e}$ and $\epsilon > \mathrm{e}$. In other words, this equation implies the probability is maximum whenever the parameter ϵ makes the empirical mean equal to the expectation with respect to $P_\epsilon(m_n)$. Rewriting the last result as an integral

$$\int_0^1 (m_n - \hat{m}_n)\mathrm{e}^{-nI_1(m_n;\epsilon)} dm_n = 0, \tag{3.184}$$

we see that estimating ϵ_{MLE} for the observed empirical $\hat{m}_n$ is equivalent to solving the integral equation to find ϵ_{MLE} that makes the above integral equal to zero. However, since $1 \ll n$, the term $e^{-nI_1(m_n;\epsilon)}$ at the minimum of $I_1(m_n^*;\epsilon)$ is exponentially larger than any other $m_n \in [0,1]\setminus\{m_n^*\}$. If the minimum of $I_1(m_n;\epsilon)$ happens to be at $\hat{m}_n$, then the term $(m_n - \hat{m}_n)$ is zero and the contribution of exponentially large $e^{-nI_1(m_n;\epsilon)}$ to the integral cancels. As a result, to find ϵ_{MLE} we need to find the value of ϵ that sets the minimum of $I_1(m_n;\epsilon)$ at $\hat{m}_n$ or

$$\left.\frac{\partial I_1(m_n;\epsilon)}{\partial m_n}\right|_{m_n=\hat{m}_n} = 0. \tag{3.185}$$

For both branches of $\epsilon \leq \mathrm{e}$ and $\epsilon > \mathrm{e}$, equation (3.82) obtains

$$\frac{\partial I_1(m_n;\epsilon)}{\partial m_n} = -\frac{\partial \tilde{H}_\epsilon(m_n)}{\partial m_n} = \ln m_n - 2\ln(1-m_n) + \ln\epsilon. \tag{3.186}$$

Thus, equation (3.185) simplifies as

$$\left.\frac{\partial I_1(m_n;\epsilon)}{\partial m_n}\right|_{m_n=\hat{m}_n} = \ln \hat{m}_n - 2\ln(1-\hat{m}_n) + \ln\epsilon = 0 \implies$$

$$\epsilon_{MLE} = \frac{(1-\hat{m}_n)^2}{\hat{m}_n}. \tag{3.187}$$

This is the maximum likelihood estimate of ϵ, given empirical mean $\hat{m}_n$.

3.7 Bayesian Conjugate Prior

In a Bayesian inference setting, it is often possible to find a conjugate prior distribution that derives the posterior distribution in closed form. Consequently, the inference part of statistical modelling reduces to updating the parameter(s) of the posterior.

In this section, we show that natural conjugate priors for B and C models are well-defined distributions and are expressible in closed form. Note that for probability distributions and their parameters, we use a slightly different notation than the rest of the thesis. The new notation is more broadly accepted in Bayesian statistics related publications.

3.7.1 Conjugate Prior for B-model

Let us here reiterate what one pursues in the Bayesian inference, *e.g.*, in the B-model case. $P_n(n_p|r)$ is a one-parameter distribution, and as a statistical model, one needs to estimate r based on observed values of n_p. And in a Bayesian setting, the modeller represents its prior knowledge as a prior distribution, say $P(r)$, and next, the Bayes rule finds the posterior distribution in terms of the likelihood $P_n(n_p|r)$ and the prior distribution $P(r)$ as

$$P_n(r|n_p) = \frac{P_n(n_p|r)P(r)}{P_n(n_p)} = \frac{P_n(n_p|r)P(r)}{\int P_n(n_p|r)P(r)dr}. \tag{3.188}$$

Accordingly, the likelihood for the B-model is

$$P_n(n_p|r) = \frac{1}{c_n(r)}\binom{n}{2n_p}(2n_p-1)!!\, r^{\lfloor n/2\rfloor - n_p}, \tag{3.189}$$

hence, the posterior writes as

$$P_n(r|n_p) = \frac{\frac{1}{c_n(r)}\, r^{\lfloor n/2 \rfloor - n_p} P(r)}{\int_0^\infty \frac{1}{c_n(r)}\, r^{\lfloor n/2 \rfloor - n_p} P(r) dr}, \tag{3.190}$$

where the constant factors cancel from numerator and denominator. Since the number of stand-alone elements is equal to $n_s = n - 2n_p$, we rewrite the posterior in terms of n_s

$$P_n(r|n_s) = \frac{\frac{1}{c_n(r)}\, r^{\lfloor n_s/2 \rfloor} P(r)}{\int_0^\infty \frac{1}{c_n(r)}\, r^{\lfloor n_s/2 \rfloor} P(r) dr}, \tag{3.191}$$

to make the notation less cluttered. By inspection, we find that the conjugate prior distribution can be defined as

$$P(r) \equiv \Lambda_n(r|\alpha, \beta) = \frac{1}{\lambda_n(\alpha, \beta)} \frac{r^\alpha}{c_n^\beta(r)}, \qquad \alpha, \beta \in \mathbb{N}, \tag{3.192}$$

such that the normalisation constant, $\lambda_n(\alpha, \beta)$, is defined as

$$\lambda_n(\alpha, \beta) \equiv \int_0^\infty \frac{r^\alpha}{c_n^\beta(r)} dr, \qquad \alpha + 2 \leq \beta \lfloor n/2 \rfloor. \tag{3.193}$$

Note that the denominator is a polynomial degree $\beta\lfloor n/2 \rfloor$, and consequently, for integers α and β, the condition $\alpha + 2 \leq \beta\lfloor n/2 \rfloor$ guarantees the convergence of the integral, and consequently, a well-defined normalisation constant for $\Lambda_n(r|\alpha, \beta)$.

In general, for $\alpha, \beta \in \mathbb{R}$ given the mentioned condition, the integral is converging. However, integer α and β enable us to derive the integral in closed form. In practice, as we will show in section (3.7.3), knowing the roots of $c_n(r)$ is enough to turn $\lambda_n(\alpha, \beta)$ to partial fractions and to express it in closed form.

Since $\lambda_n(\alpha, \beta)$ is well-defined, plugging the prior $\Lambda_n(r|\alpha, \beta)$ in equation (3.191) obtains

$$P_n(r|n_s) = \frac{\frac{1}{c_n(r)}\, r^{\lfloor n_s/2 \rfloor} \times \frac{1}{\lambda_n(\alpha,\beta)} \frac{r^\alpha}{c_n^\beta(r)}}{\int_0^\infty \frac{1}{c_n(r)}\, r^{\lfloor n_s/2 \rfloor} \times \frac{1}{\lambda_n(\alpha,\beta)} \frac{r^\alpha}{c_n^\beta(r)} dr} = \frac{1}{\lambda_n(\alpha + \lfloor \frac{n_s}{2} \rfloor, \beta + 1)} \times \frac{r^{\lfloor n_s/2 \rfloor + \alpha}}{c_n^{\beta+1}(r)} \implies$$

$$P_n(r|n_s) = \Lambda_n(r|\alpha + \lfloor n_s/2 \rfloor, \beta + 1), \tag{3.194}$$

which is the same probability distribution with shifted parameters. Simply put, observing n_p, or equivalently n_s, updates the parameter of $\Lambda_n(r|\alpha, \beta)$ by changing its parameters to $\Lambda_n(r|\alpha+\lfloor n_s/2\rfloor, \beta+1)$. Also, the convergence condition for $\lambda_n(\alpha+\lfloor n_s/2\rfloor, \beta+1)$ is satisfied, since

$$\alpha + 2 \leq \beta\lfloor\frac{n}{2}\rfloor \implies \alpha + 2 + \lfloor\frac{n_s}{2}\rfloor \leq \beta\lfloor\frac{n}{2}\rfloor + \lfloor\frac{n}{2}\rfloor - n_p \leq (\beta+1)\lfloor n/2\rfloor. \quad (3.195)$$

Similar to the maximum likelihood section, we can assume there are K independent observations, and the empirical mean of stand-alone elements, namely $\hat{n}_s$, is defined as

$$\hat{n}_s = \frac{2}{K}\sum_{i=1}^{K}\lfloor\frac{n_s^{(i)}}{2}\rfloor = \frac{2}{K}\sum_{i=1}^{K}(\lfloor\frac{n}{2}\rfloor - n_p^{(i)}). \quad (3.196)$$

The likelihood function for K independent observations is

$$P_n(n_s^{(1)}, \dots, n_s^{(K)}|r) = \prod_{i=1}^{K} P_n(n_s^{(i)}|r) \propto \prod_{i=1}^{K} \frac{r^{\lfloor\frac{n_s^{(i)}}{2}\rfloor}}{c_n(r)} = \frac{r^{\sum_{i=1}^{K}\lfloor\frac{n_s^{(i)}}{2}\rfloor}}{c_n^K(r)} \implies$$

$$P_n(n_s^{(1)}, \dots, n_s^{(K)}|r) \propto \frac{r^{\frac{K\hat{n}_s}{2}}}{c_n^K(r)}. \quad (3.197)$$

So, the posterior distribution derives like

$$P_n(r|n_s^{(1)}, \dots, n_s^{(K)}) = \frac{P_n(n_s^{(1)}, \dots, n_s^{(K)}|r)\Lambda_n(r|\alpha,\beta)}{P_n(n_s^{(1)}, \dots, n_s^{(K)})}$$

$$= \frac{\frac{r^{\frac{K\hat{n}_s}{2}+\alpha}}{c_n^{K+\beta}(r)}}{\int_0^\infty \frac{r^{\frac{K\hat{n}_s}{2}+\alpha}}{c_n^{K+\beta}(r)}dr} = \frac{1}{\lambda_n(\alpha+\frac{K\hat{n}_s}{2}, \beta+K)}\frac{r^{\frac{K\hat{n}_s}{2}+\alpha}}{c_n^{K+\beta}(r)} \implies$$

$$P_n(r|n_s^{(1)}, \dots, n_s^{(K)}) = \Lambda_n(r|\alpha+\frac{K\hat{n}_s}{2}, \beta+K). \quad (3.198)$$

Figure (3.17) shows $\Lambda_n(r|\alpha,\beta)$ for $n=8$. The left panel plots it for $\beta=1$, and the right one is for $\beta=2$. We see that when α is getting close to $\beta\lfloor n/2\rfloor$ the uncertainty in r, or the desperation of the distribution, increases. Also, figure (3.18) plots $\lambda_n(\alpha,\beta)$ versus α for different βs and for $n=32$.

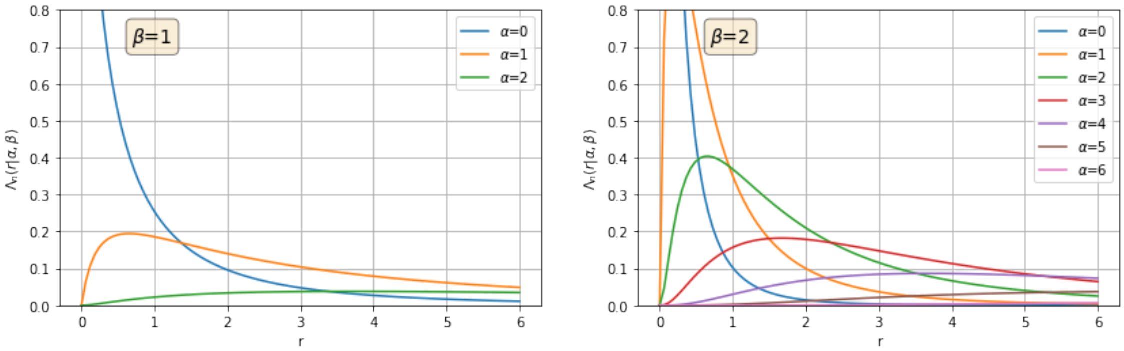


Figure 3.17: For $n = 8$, the left panel plots $\Lambda_n(r|\alpha, \beta)$ for $\beta = 1$, and similarly, the right panel shows the case for $\beta = 2$.

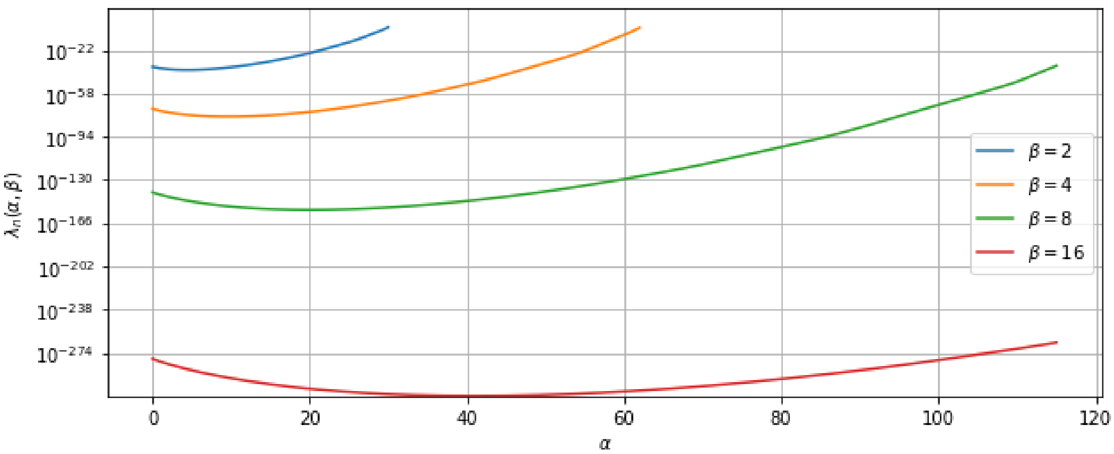


Figure 3.18: The graph of $\lambda_n(\alpha, \beta)$ for $n = 32$ for different values of βs are plotted along α. Note that $\lambda_n(\alpha, \beta)$ is continuous in $\alpha \in \mathbb{R}$, but for integer values $\alpha \in \mathbb{N}$, we can express it in closed form.

3.7.1.1 Maximum a posteriori estimation

By taking the derivative of $\Lambda_n(r|\alpha, \beta)$ with respect to r, one can derive the maximum a posteriori estimation, r_{MAP}, as follows

$$\frac{d\Lambda_n(r|\alpha, \beta)}{dr} = \frac{1}{\lambda_n(\alpha, \beta)} \times \frac{\alpha r^{\alpha-1} c_n^{\beta}(r) - \beta r^{\alpha} c_n^{\beta-1}(r) \frac{dc_n(r)}{dr}}{c_n^{2\beta}(r)} = 0 \implies$$

$$\frac{\alpha}{\beta}c_n(r) - r\frac{dc_n(r)}{dr} = 0 \implies$$

$$\left(n - \frac{2\alpha}{\beta}\right)c_n(r) - n(n-1)c_{n-2}(r) = 0, \tag{3.199}$$

where we used equation (B.74) in the last step to replace $dc_n(r)/dr$. So, the root of the last polynomial equation is r_{MAP}. Compare equation (3.199) with (3.164) to see the difference between r_{MLE} and r_{MAP}.

3.7.1.2 The asymptotic form of r_{MAP}

For $1 \ll n$ and constant n, equation (3.90) finds the ratio of $c_{n-2}(r)/c_n(r)$ as

$$\frac{c_{n-2}(r)}{c_n(r)} \sim \frac{\mathrm{e}^{\sqrt{\frac{r}{n}}}}{n}. \tag{3.200}$$

Thus, by replacing the asymptotic leading term of the ratio in equation (,3.199) one finds

$$r_{MAP} = n\ln^2\left(1 - \frac{2\alpha}{n\beta}\right). \tag{3.201}$$

Compare this result with the asymptotic estimate of r_{MLE} in equation (3.172).

3.7.1.3 Posterior Predictive Distribution

The posterior predictive distribution simply derives by integrating the likelihood and the posterior with respect to r

$$P_n(n_p|D) = \int_0^\infty P_n(n_p|r)\Lambda_n(r|D)dr = \binom{n}{2n_p}(2n_p-1)!!\int_0^\infty \frac{r^{\lfloor n/2\rfloor - n_p}}{c_n(r)} \times \frac{r^\alpha}{\lambda_n(\alpha,\beta)c_n^\beta(r)}dr$$

$$= \binom{n}{2n_p}(2n_p-1)!!\frac{\lambda_n(\alpha + \lfloor n/2\rfloor - n_p, \beta+1)}{\lambda_n(\alpha,\beta)}. \tag{3.202}$$

3.7.1.4 Moments of $\Lambda_n(r|\alpha,\beta)$

The moments of the probability distribution $\Lambda_n(r|\alpha,\beta)$ are expressible in closed form. To find the kth moment, we write

$$\langle r^k \rangle = \int_0^\infty r^k \Lambda_n(r|\alpha,\beta) dr = \frac{1}{\lambda(\alpha,\beta)} \int_0^\infty \frac{r^{\alpha+k}}{c_n^\beta(r)} dr \implies$$

$$\langle r^k \rangle = \frac{\lambda(\alpha+k,\beta)}{\lambda(\alpha,\beta)}. \tag{3.203}$$

3.7.2 Conjugate Prior for the C-model

The probability distribution of the C-model, say $P_n(n_p, n_h|r,\rho)$, can be seen as the probability distribution of the B-model multiplied by a Binomial distribution. Therefore, if we assume parameter r and ρ are independent, the prior for the C-model is simply the multiplication of $\Lambda_n(r|\alpha_1,\beta_1)$ by the Beta distribution for the Binomial part, $Beta(\rho|\alpha_2,\beta_2)$ as

$$\Lambda_n^{(2)}(r,\rho|\alpha_1,\beta_1,\alpha_2,\beta_2) = \Lambda_n(r|\alpha_1,\beta_1) Beta(\rho|\alpha_2,\beta_2). \tag{3.204}$$

So, the posterior writes as

$$\Lambda_n^{(2)}(r,\rho|\alpha_1,\beta_1,\alpha_2,\beta_2) = \frac{P_n(n_p,n_h|r,\rho)\Lambda_n(r|\alpha_1,\beta_1)Beta(\rho|\alpha_2,\beta_2)}{\int_0^\infty \int_0^1 P_n(n_p,n_h|r,\rho)\Lambda_n(r|\alpha_1,\beta_1)Beta(\rho|\alpha_2,\beta_2) d\rho dr}$$

$$= \Lambda_n(r|\alpha_1 + \lfloor n/2 \rfloor - n_p, \beta_1 + 1) Beta(\rho|\alpha_2 + n_h, \beta_2 + n - 2n_p - n_h)$$

$$= \Lambda_n^{(2)}(r,\rho|\alpha_1 + \lfloor n/2 \rfloor - n_p, \beta_1 + 1, \alpha_2 + n_h, \beta_2 + n - 2n_p - n_h), \tag{3.205}$$

which is again expressed in the same mathematical form as the prior.

3.7.3 Finding $\lambda_n(\alpha,\beta)$ in closed form

The function $\lambda_n(\alpha,\beta)$ is defined as

$$\lambda_n(\alpha,\beta) = \int_0^\infty \frac{r^\alpha}{c_n^\beta(r)} dr, \qquad \alpha,\beta \in \mathbb{N}, \tag{3.206}$$

in which the normalisation constant of the pairing models, namely $c_n(r)$, is a degree $\lfloor n/2 \rfloor$ polynomial with positive coefficients. In principle, the Fundamental Theorem of Algebra guarantees the existence of $c_n(r)$ roots in the complex plane [50], and factorises like

$$c_n(r) = f_1(r) \dots f_k(r), \tag{3.207}$$

where $f_i(r)$ is either a power of a linear factor $(r + a_i)^{l_i}$ or a quadratic factor $(r^2 + b_i r + c_i)^{l_i}$ (The quadratic factors are resulted in by complex conjugate solutions, and therefore, we must have $b_i^2 < 4c_i$). Note that $l_i \in 1, 2, \dots, \lfloor n/2 \rfloor$ is the degeneracy of the roots. However, we will show that for $n \geq 2$, $c_n(r)$ has $\lfloor n/2 \rfloor$ distinct roots, and therefore, $l_i = 1$ for all roots and $f_i(r)$ is only a linear factor such that

$$c_n(r) = (r + a_1)(r + a_2) \dots (r + a_{\lfloor n/2 \rfloor}) = \prod_{i=1}^{\lfloor n/2 \rfloor} (r + a_i). \tag{3.208}$$

Also for $0 \leq r$, we have $c_n(r) > 0$. Hence, all the roots must be negative, which assert $a_i > 0$. Considering these observations and for $\beta \in \mathbb{N}$, the ratio $r^\alpha / c_n^\beta(r)$ can be rewritten as partial fractions like

$$\frac{r^\alpha}{c_n^\beta(r)} = \frac{r^\alpha}{\prod_{i=1}^{\lfloor n/2 \rfloor} (r + a_i)^\beta} = \sum_{i=1}^{\lfloor n/2 \rfloor} \sum_{j=1}^{\beta} \frac{A_{ij}}{(r + a_i)^j}, \tag{3.209}$$

for real numbers A_{ij}. Later, we shall show that for $\beta \in \mathbb{N}$ and all such that $\alpha \leq \beta \lfloor n/2 \rfloor - 2$, we have

$$\sum_{i=1}^{\lfloor n/2 \rfloor} A_{i1} = 0. \tag{3.210}$$

At this stage, we assume it is true. Next, plugging equation (3.209) in equation (3.206) yields

$$\begin{aligned}
\lambda_n(\alpha, \beta) = \int_0^\infty \sum_{i=1}^{\lfloor n/2 \rfloor} \sum_{j=1}^{\beta} \frac{A_{ij}}{(r + a_i)^j} dr &= \sum_{i=1}^{\lfloor n/2 \rfloor} A_{i1} \int_0^\infty \frac{dr}{(r + a_i)} + \sum_{i=1}^{\lfloor n/2 \rfloor} \sum_{j=2}^{\beta} A_{ij} \int_0^\infty \frac{dr}{(r + a_i)^j} \\
&= \lim_{b \to \infty} \sum_{i=1}^{\lfloor n/2 \rfloor} A_{i1} \ln(r + a_i) \Bigg|_0^b - \sum_{i=1}^{\lfloor n/2 \rfloor} \sum_{j=2}^{\beta} \frac{A_{ij}}{(j-1)(r + a_i)^{j-1}} \Bigg|_0^\infty \\
&= \lim_{b \to \infty} \sum_{i=1}^{\lfloor n/2 \rfloor} A_{i1} \ln(b + a_i) - \sum_{i=1}^{\lfloor n/2 \rfloor} A_{i1} \ln(a_i) + \sum_{i=1}^{\lfloor n/2 \rfloor} \sum_{j=2}^{\beta} \frac{A_{ij}}{(j-1) a_i^{j-1}}.
\end{aligned} \tag{3.211}$$

However,

$$\lim_{b\to\infty}\sum_{i=1}^{\lfloor n/2\rfloor} A_{i1}\ln(b+a_i) = \lim_{b\to\infty}\ln\left(\prod_{i=1}^{\lfloor n/2\rfloor}(b+a_i)^{A_{i1}}\right)$$
$$= \lim_{b\to\infty}\ln b^{\sum_{i=1}^{\lfloor n/2\rfloor} A_{i1}} = \lim_{b\to\infty}\ln b^0 = 0, \tag{3.212}$$

where we used equation (3.210) in the last line. Finally, $\lambda_n(\alpha,\beta)$ can be expressed as

$$\lambda_n(\alpha,\beta) = -\sum_{i=1}^{\lfloor n/2\rfloor} A_{i1}\ln(a_i) + \sum_{i=1}^{\lfloor n/2\rfloor}\sum_{j=2}^{\beta}\frac{A_{ij}}{(j-1)a_i^{j-1}}. \tag{3.213}$$

To reiterate the steps to write $\lambda_n(\alpha,\beta)$ in closed form, remember that for $c_n(r)$ we can numerically find $\lfloor n/2\rfloor$ distinct roots, namely a_i, up to the required precision once, and use them to find $\lambda_n(\alpha,\beta)$. Next, for given α and β, one constructs a linear system for A_{ij} to use the last result to find $\lambda_n(\alpha,\beta)$.

In addition, there exist some iterative algorithms that find the partial decomposition coefficients directly with efficiency and speed [12, 41, 44]. However, writing $\lambda_n(\alpha,\beta)$ in closed form is independent of the estimated values of A_{ij}, by which, one can investigate the problem of interest further analytically.

Returning back to the sum of A_{i1}, let us rewrite the partial fraction of the ratio

$$\frac{r^\alpha}{c_n^\beta(r)} = \frac{r^\alpha}{\prod_{i=1}^{\lfloor n/2\rfloor}(r+a_i)^\beta} = \sum_{i=1}^{\lfloor n/2\rfloor}\sum_{j=1}^{\beta}\frac{A_{ij}}{(r+a_i)^j}, \tag{3.214}$$

where we will show that

$$\sum_{i=1}^{\lfloor n/2\rfloor} A_{i1} = 0, \tag{3.215}$$

for $\beta\in\mathbb{N}$ and all $\alpha\in\{0,1,\ldots,\beta\lfloor n/2\rfloor-2\}$. Observe that for $r\to\infty$, the A_{i1} are the dominant partial decomposition coefficients. Therefore, we multiply both side of equation (3.214) by r

$$\frac{r^{\alpha+1}}{\prod_{i=1}^{\lfloor n/2\rfloor}(r+a_i)^\beta} = \sum_{i=1}^{\lfloor n/2\rfloor}\sum_{j=1}^{\beta}\frac{A_{ij}r}{(r+a_i)^j}, \tag{3.216}$$

and take the limit $r\to\infty$. Since $\alpha+1<\beta\lfloor n/2\rfloor$, the left hand side approaches zero

$$\lim_{r\to\infty}\frac{r^{\alpha+1}}{\prod_{i=1}^{\lfloor n/2\rfloor}(r+a_i)^\beta}\to 0. \tag{3.217}$$

At the same time,

$$\lim_{r\to\infty}\sum_{i=1}^{\lfloor n/2\rfloor}\sum_{j=1}^{\beta}\frac{A_{ij}r}{(r+a_i)^j}=\lim_{r\to\infty}\sum_{i=1}^{\lfloor n/2\rfloor}\sum_{j=1}^{\beta}A_{ij}r^{1-j}=\sum_{i=1}^{\lfloor n/2\rfloor}A_{i1}. \tag{3.218}$$

Combining both results finds

$$\sum_{i=1}^{\lfloor n/2\rfloor}A_{i1}=0, \tag{3.219}$$

which complete the proof of the claim.

3.7.4 Finding the number of distinct roots of $c_n(r)$

In this section, we will show that the polynomial $c_n(r)$ has $\lfloor n/2\rfloor$ distinct roots. To start, we need to introduce the Sturm sequence [4]. For Polynomial P the Sturm sequence, denotes by $SS(P, P') = P_0, P_1, \ldots, P_i$, is defined as

$$P_0 = P,$$

$$P_1 = P',$$

$$P_i = -rem(P_{i-1}, P_{i-2}) \tag{3.220}$$

where P' is the first derivative of the polynomial P, $rem(P_i, P_{i-1})$ is the remainder of the Euclidean division of P_{i-1} by P_{i-2}, and the sequence terminates when $rem(P_i, P_{i-1})$ is a constant. In addition, the number of sign variations [4], denoted by $V(\mathcal{A})$, for a sequence $\mathcal{A} = a_0, a_1, \ldots, a_i$ of elements in $\mathbb{R}\setminus\{0\}$ is defined as

$$V(a_0) = 0,$$

$$V(a_0, a_1, \ldots, a_i) = \begin{cases} V(a_1,\ldots,a_i)+1 & \text{if } a_0a_1<0 \\ V(a_1,\ldots,a_i) & \text{if } a_0a_1>0 \end{cases}. \tag{3.221}$$

In other words, $V(\mathcal{A})$ gives the number of sign changes in the sequence $\mathcal{A}$. And finally, if $\mathcal{P} = P_0, P_1, \ldots P_i$ is a sequence of polynomials, the number of sign variations of $\mathcal{P}$ at a, denotes by $V(\mathcal{P}, a)$, is defined as

$$V(\mathcal{P}, a) = V(P_0(a), P_1(a), \ldots, P_i(a)). \tag{3.222}$$

Theorem 3.7.1. ***[Sturm' Theorem]*** *[4] Given a and b in $\mathbb{R} \cup \{-\infty, \infty\}$,*

$$V(SS(P, P'), a) - V(SS(P, P'), b)$$

is the number of roots of P in the interval (a, b).

To see the detail of proving that the polynomial $c_n(r)$ has $\lfloor n/2 \rfloor$ distinct roots, check appendix (B.13). In there, we use Sturm's Theorem and prove the claim for even and odd values separately.

3.8 Immigration Models

Using pairs or aggregates as building blocks of statistical modelling is not unique to this study and has already been investigated in statistical and probability modelling literatures. For instance, two relevant models, namely death-multiple immigration and birth-death-multiple immigration [30, 32, 47], model a stochastic process by which the population size increases by the arrival of singles, pairs, ..., m-tuples of migrants.

Let us discuss the case of death-multiple immigration in more detail here in order to explain the differences between the assumptions and approaches of this study and what has already been discussed in the publication. According to this model, population size is depleted at a constant rate μ by death that occurs in proportion to the instantaneous size of the population, and the population size increases through immigration of m-aggregates that arrive at the rate $\alpha_m \geq 0$, for $m = 1, 2, \ldots$ and the momentary population size N. The Kolmogorov forward rate equation for the process is written as [30]

$$\frac{dP_N(t)}{dt} = \mu(N+1)P_{N+1}(t) - \mu N P_N(t) - P_N(t) \sum_{m=1}^{\infty} \alpha_m + \sum_{m=1}^{N} \alpha_m P_{N-m}(t), \quad (3.223)$$

in which $P_N(t)$ is the probability that the population comprises N members at time t. In the case of birth-death-multiple immigration models, the birth rate of the elements is included.

Interestingly, these stochastic models have analytic solutions. For example, only including pairs of immigration, or for

$$\alpha_m = \begin{cases} \nu, & m = 2 \\ 0, & \text{otherwise} \end{cases}, \tag{3.224}$$

the equilibrium probabilities asymptotically are derived as [47]

$$P_{2n}(\infty) = \bar{N}^{n+1} \exp(-3\bar{N}/4) L_n^{(-1/2)}(-\bar{N}/4) n!/(2n)! \tag{3.225}$$

$$P_{2n+1}(\infty) = \bar{N}^{n+1} \exp(-3\bar{N}/4) L_n^{(1/2)}(-\bar{N}/4) n!/2(2n+1)!, \tag{3.226}$$

where $\bar{N} = 2\nu/\mu$ and $L_n^{(b)}(x)$ are Laguerre polynomials.

The similarity between these models and the pairing models in this thesis is striking, e.g., when we compare the equilibrium probabilities with phase space volumes in equations (2.16) and (2.17). However, we must emphasise what makes our results different and how these two models are related.

First, in deriving pairing models, we start from combinatorial arguments and the focus is on the states the aggregates can occupy in an arbitrary configuration. Second, there is no assumption about the existence of a dynamic or a stochastic process, and the resulting probability distribution is for the number of pairs (or, in the case of the C-model, the number of pairs and heads). And finally, when in chapter (5), by using statistical mechanics, we use the pairing model to study systems in equilibrium, the dynamics are governed by the Hamiltonian that we construct for some prototype models.

In comparison, the immigration processes are by design out of an equilibrium although they have an equilibrium solution, and the random variable in these equilibrium probability distributions is the population numbers instead of the number of pairs in our case. Simultaneously, the dynamics specify by death and migration rates, namely μ and α_m, unlike our models that govern by the Hamiltonian structure that we impose.

We must emphasise that the paring models that we have introduced, or any other aggregate models, are prototypes that can be used to investigate the notion of emergence in the state space. So, it is an interesting question to find how the new definition of emergence can be revisited in immigration processes, however, it is not the intention and goal of our study.

3.9 Conclusion

This chapter derived two probability distributions, namely $P_n(n_p)$ and $P_n(n_p, n_h)$, and their equivalent large deviation probabilities, namely $P_n(m_n)$ and $P_n(m_n, s_n)$, for B and C models, respectively. Due to emerging states and faster than the exponential growth rate, we observed that the large deviation speeds of $P_n(m_n)$ and $P_n(m_n, s_n)$ are in order $O(n \ln n)$. In contrast, for the limiting case $\lim_{n\to\infty} r/n = \epsilon$, the probability distributions $P_\epsilon(m_n)$ and $P_\epsilon(m_n, s_n)$ have linear large deviation speed, or $O(n)$.

One significant result of this chapter is the normalisation constant recursive relation in equation (3.46), which resembles the pairing model recursive relation in equation (2.9). We argue that this property must generally be valid for every n-tet compounding mechanism. Recall that the state space recursive relation in equation (2.9) was based on a combinatorial argument, and the degeneracies in the probability distributions in equation (2.15) are also based on the same argument, which resulted in equation (3.46).

Therefore, the state space geometry of compounding mechanisms encodes as a recursive relation like equation (2.37). Moreover, a probability distribution with its normalisation constant must exist that governs by a recursive equation. Apart from that, the normalisation constant in equation (3.38) is a polynomial with degeneracies as its coefficients. So do the normalisation constants of other compounding mechanisms.

In addition, the normalisation constant's polynomial imposed the form of the Bayesian conjugate prior. Thus, replacing it with other compounding mechanisms introduces a class of well-defined conjugate priors. Hence, again, we see that starting from the emerging states based on the compounding mechanism has far-reaching consequences in the mathematical forms down the line.

As we mentioned at the start of this chapter, the B and C models' averages exhibit second-order phase transition[8]. To be precise, defining the ratio of the number of stand-alone elements to the total as

$$m_n^{'} \equiv 1 - m_n = \frac{n_s}{n}, \tag{3.227}$$

in thermodynamic limit, the averages with respect to $P_n(m_n)$ or $P_n(m_n, s_n)$ finds

$$\lim_{n\to\infty} \langle m_n^{'} \rangle = 0. \tag{3.228}$$

Simultaneously, the same averages for the corresponding distributions $P_\epsilon(m_n)$ and $P_\epsilon(m_n, s_n)$ are non-zero

$$\lim_{n\to\infty} \langle m_n^{'} \rangle = 1 - \mathrm{e}^{-\sqrt{\epsilon}}. \tag{3.229}$$

Likewise, the order parameters of the second-order phase transition have the same property. Due to this observation, assume a secondary model that utilises the pairing mechanism as its building blocks such that some other internal mechanism controls the parameter r. If the mechanism changes the scaling property of r, the phase transition must occur with the number of stand-alone elements as its order parameter.

Lastly, in light of the results in the large deviation section, we review the additivity of the pairing models further. Recall that the large deviation probability of $P_n(m_n)$ in equation (3.63) writes

$$\lim_{n\to\infty} -\frac{1}{n \ln n} \ln P_n(m_n) = \frac{1-m_n}{2}. \tag{3.230}$$

As a first step and based on the asymptotic extensivity definition in section (1.3) we ask what other continuous, strictly increasing function $f(.)$ can be replaced by the logarithm of $P_n(m_n)$ such that the speed of LDP – $n \ln n$ – is replaced by a linear speed n?

$$\lim_{n\to\infty} -\frac{1}{n} f(P_n(m_n)) < \infty. \tag{3.231}$$

One might propose to replace the natural logarithm with the logarithm base n. We see that since logarithms satisfy the identity

$$\ln_n(x) = \frac{\ln(x)}{\ln(n)}, \tag{3.232}$$

the LDP limit writes as

$$\lim_{n\to\infty} -\frac{1}{n} \ln_n P_n(m_n) = \frac{1-m_n}{2}, \tag{3.233}$$

and the function $f(x) \equiv \ln_n(x)$. However, $f(x)$ is a class of functions that parametrises by the system size, n, and therefore, different than a single size-independent function. Furthermore, we will see there is a second possibility. Using equation (3.59), we write

$$\ln P_n(m_n) = -(\frac{1-m_n}{2})n\ln n - \frac{n}{2}\left[m_n \ln m_n + (1-m_n)\ln\frac{(1-m_n)^2}{er}\right] + O(\sqrt{n})$$

$$= -\frac{n(1-m_n)}{2}\ln\frac{n(1-m_n)}{2} + \frac{n}{2}H(m_n) - \frac{n(1-m_n)}{2}\ln\frac{2}{er} + O(\sqrt{n}), \quad (3.234)$$

where $H(m_n)$ is Shannon entropy of m_n, explained in equation (3.60). Consequently, if we define the transformation function as

$$\phi(x) = \begin{cases} \mathrm{e}^{L(-x)} & x \leq 0 \\ \mathrm{e}^{L(x)} & x > 0 \end{cases}, \quad (3.235)$$

and its inverse as

$$\phi^{-1}(x) = \begin{cases} -L^{-1}(\ln x) & x \leq 0 \\ L^{-1}(\ln x) & x > 0 \end{cases}, \quad (3.236)$$

where $L(x)$ is the Lambert function[2], we find

$$\lim_{n\to\infty}\frac{1}{n}\phi(-\ln P_n(m_n)) = \lim_{n\to\infty}\frac{1}{n}\mathrm{e}^{L(-\ln P_n(m_n))} = \frac{(1-m_n)}{2}. \quad (3.238)$$

The last result is the limit of the transformed large deviation with speed n. So, one can say this transformation makes the logarithm of the probability distribution $P_n(m_n)$ asymptotically extensive. It is interesting to mention that, the same conjugate functions transform $\ln P_n(m_n, s_n)$ to an asymptotically extensive quantity.

Mathematically, using a different one-to-one function other than logarithm to write the large deviation limit might seem as not more than a triviality. However, in chapter (5), we will see its applicability for a system with dynamics, specifically, statistical mechanics systems. Despite the seeming triviality, the large deviation estimate and its rate function are unaffected by the $\phi(x)$ transformation. *e.g.*, the position of the rate function maximum, or the distribution expectation, is the same in the transformed case and written as

$$P_n(m_n) \asymp \mathrm{e}^{-L^{-1}(\ln nI(m_n))}. \quad (3.239)$$

[2]The Lambert L function is defined as

$$L(x\ln x) = \ln x. \quad (3.237)$$

To finish this chapter, we show that the logarithm of the normalisation constant $\ln c_n(r)$ is asymptotically extensive when it is transformed by $\phi(x)$

$$\lim_{n\to\infty}\frac{1}{n}\phi(\ln c_n(r)) = \lim_{n\to\infty}\frac{\mathrm{e}^{L(\ln c_n(r))}}{n} = \frac{1}{2}. \tag{3.240}$$

Also, we find the limit of the transformed degeneracy in equation (2.15), $W_n(n_p)$. First, observe that

$$\ln(W_n(m_n)) = \frac{m_n}{2} n \ln n - \frac{n}{2}\left[m_n \ln\frac{m_n}{p} + 2(1-m_n)\ln\frac{1-m_n}{s} + m_n\right]$$

$$= \frac{nm_n}{2}\ln\frac{nm_n}{2} - \frac{n}{2}\left[m_n \ln\frac{m_n^2}{p} + (1-m_n)\ln\left(\frac{1-m_n}{s}\right)^2 + m_n\right]. \tag{3.241}$$

Then, the limit writes as

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\phi(\ln W_n(m_n)) &= \lim_{n\to\infty}\frac{\mathrm{e}^{L(\ln W_n(m_n))}}{n}\\ &= \lim_{n\to\infty}\frac{\mathrm{e}^{L(\frac{nm_n}{2}\ln\frac{nm_n}{2})}}{n}\\ &= \lim_{n\to\infty}\frac{\mathrm{e}^{\ln\frac{nm_n}{2}}}{n} = \frac{m_n}{2}.\end{aligned} \tag{3.242}$$

We will use the last result in chapter (5).

Chapter

FOUR

Pairing Models in Information Theory

The probability distributions that we studied in the previous chapter provide an opportunity to explore the state space emergence through information-theoretic measures, such as Shannon entropy, mutual information, and joint entropy [15, 45].

Before reporting the results, let us elaborate on the goals that we will follow in this chapter. Ideally, suppose in comparison to typical systems, complex systems have inherently different emergent properties. In that case, measures for a complex system deviate from their usual counterparts, and we assert that the deviation is due to emergent properties. We must emphasise that the deviation we are looking for is in the mathematical form of the measures, which might be some additional terms in the formula. Consequently, by studying quantities of information for emergent systems, one can look for features specific to complex systems as a new or extra term compared to well-known results.

Apparently, this goal is achievable since we have pairing probability distributions, and more importantly, most of their relevant statistics are expressible in closed form, and as a result, it is possible to study the pertinent quantities analytically. At the same time, we need to select a so-called typical system such that its inherent properties be relevant to the pairing models and their state spaces.

At this stage, it is evident that the Binomial distribution has a Cartesian state space. Consequently, the states of its constituent elements are statistically independent, and its statistical and information-theoretic quantities are expressible in closed form. In addition, numerous mathematical models constitute a binary random variable and are modelled based on the Binomial distribution. Due to its simplicity and ubiquity, we use the Binomial distribution as a benchmark to show the deviation in mathematical form.

Meanwhile, emergent properties of the pairing models can inspire us to look at some theorems and applications in information theory from a new angle. To do so, we shall construct a transmission model with a pairing time series and show for this model that the rate entropy [15, 42, 45] is a well-defined quantity under certain assumptions.

In this brand-new transmission model, the pairing time series distinguishes from the usual one so that the momentary state of the received signal can carry with certainty the value of the signal in the future. Furthermore, we shall see the Shannon entropy results in converging rate entropy depending on the distance distribution between paired elements.

4.1 Information-theoretic Measures

The Shannon entropy [15, 45] measures the associated uncertainty for a random variable, given its probability distribution function. So, for a set of probability distributions of system size n, denoted by $\mathcal{P}_n$, the Shannon entropy is a functional such that

$$H[P_n] : \mathcal{P}_n \to \mathbb{R}^+ \cup \{0\}, \tag{4.1}$$

and is defined as

$$H[P_n] = -\sum_i p_i \ln p_i, \qquad p_i \in P_n. \tag{4.2}$$

Besides, for a system with n statistically independent elements, each with two states – say head and tail – Bernoulli random variables describe the randomness in observations – we call them Bernoulli coins. In contrast, if they are pairing balls or coins, one can use statistical random variables introduced in the previous chapter. At the same time, the states of Bernoulli coins are independent of the rest of the system, and if the head and tail probabilities are equal to $(\rho, 1-\rho)$ respectively, the entropy finds as

$$H[(\rho, 1-\rho)] = -\rho \ln \rho - (1-\rho) \ln(1-\rho). \tag{4.3}$$

On the contrary, in pairing models, the state of a single element is system-dependent. Thus, using marginals in equation (3.129) from chapter (3), one can write the entropy of a single pairing ball as

$$H_B[P_n(X_l)] = -\langle \frac{2n_p}{n} \rangle \ln \langle \frac{2n_p}{n} \rangle - (1 - \langle \frac{2n_p}{n} \rangle) \ln(1 - \langle \frac{2n_p}{n} \rangle), \tag{4.4}$$

and similarly, by using equation (3.145) for a single pairing coin the entropy is equal to

$$\begin{aligned} H_C[P_n(X_l)] = & -\langle \frac{2n_p}{n} \rangle \ln \langle \frac{2n_p}{n} \rangle - (1 - \langle \frac{2n_p}{n} \rangle) \ln(1 - \langle \frac{2n_p}{n} \rangle) \\ & - (1 - \langle \frac{2n_p}{n} \rangle)(\rho \ln \rho + (1-\rho) \ln(1-\rho)). \end{aligned} \tag{4.5}$$

Interestingly, the pairing mechanism changes the entropy of the Bernoulli coin from a local quantity to a system-dependent one in the pairing ball case since the mean values in the derived equations above are system-dependent. Meanwhile, later we will derive that entropy is decomposable to a pairing effect and a Bernoulli effect for a pairing coin

$$H_C[P_n(X_l)] = H_B[P_n(X_l)] + \langle \frac{n_s}{n} \rangle H[(\rho, 1-\rho)], \tag{4.6}$$

where $\langle \frac{n_s}{n} \rangle = 1 - \langle \frac{2n_p}{n} \rangle$ is the average of the ratio of elements in a stand-alone state.

Furthermore, unlike Bernoulli's entropy, the system-dependency of pairing entropies has an asymptotic leading term when $1 \ll n$. For example, for $H_B[P_n(X_l)]$, we will show

$$H_B[P_n(X_l)] = \sqrt{\frac{r}{n}} \left[1 - \ln \sqrt{\frac{r}{n}} \right] + O(\frac{1}{n}), \tag{4.7}$$

and for $H_C[P_n(X_l)]$

$$H_C[P_n(X_l)] = \sqrt{\frac{r}{n}} \left[1 - \ln \sqrt{\frac{r}{n}} + H[(\rho, 1-\rho)] \right] + O(\frac{1}{n}). \tag{4.8}$$

This introductory explanation can give us the general theme of what we should expect in the following parts: We first study information-theoretic quantities of paring random variables and then compare them to an ordinary one. Next, we derive the asymptotic leading term and discuss its consequence and interpretation provided system size effects.

The asymptotic entropies of a handful of pairing random variables are microscopic quantities since they belong to a minuscule subsystem inside a macroscopic system (assuming $1 \ll n$). Thus, for $1 \ll n$, based on the number of random variables of a quantity and their relative size to the system, we can classify the derived measure as microscopic, mesoscopic or macroscopic. *e.g.* if k represents the number of random variables in, say, joint entropy, $k = O(1)$ is microscopic, $k = O(\sqrt{n})$ is mesoscopic, and $k = O(n)$ is macroscopic. Therefore, we shall use the joint entropy for two or more elements to study the spectrum of subsystem sizes using the joint distribution we derived in section (3.5).

Nonetheless, for macroscopic entities, we study the ensemble entropy over the probabilities of all configurations. Let us elaborate on the difference between the ensemble entropy and the entropy of a system-wide random quantity like the number of heads by giving an example for Bernoulli coins: the probability of observing a given configuration as a sequence of head and tail states is

$$P_i = \rho^{n_h}(1-\rho)^{n-n_h}, \tag{4.9}$$

for 2^n distinct configurations. So, the ensemble entropy finds as

$$\begin{aligned} H[Ber_n] &= -\sum_{i=1}^{2^n} \rho^{n_h}(1-\rho)^{n-n_h}(n_h \ln\rho + (n-n_h)\ln(1-\rho)) \\ &= -\langle n_h\rangle \ln\rho - (n - \langle n_h\rangle)\ln(1-\rho), \end{aligned} \tag{4.10}$$

and since $\langle n_h\rangle = n\rho$, it simplifies to

$$H[Ber_n] = nH[(\rho, 1-\rho)]. \tag{4.11}$$

At the same time, the Binomial distribution for the number of heads, denoted by n_h, is defined as

$$Bin_n(n_h) = \binom{n}{n_h}\rho^{n_h}(1-\rho)^{n-n_h}. \tag{4.12}$$

However, n_h is a system-wide quantity, and its entropy derives as

$$H[Bin_n(n_h)] = -\sum_{n_h=0}^{n}\binom{n}{n_h}\rho^{n_h}(1-\rho)^{n-n_h}(n_h\ln\rho + (n-n_h)\ln(1-\rho) + \ln\binom{n}{n_h})$$

$$= H[Ber_n] + \langle \ln n_h! \rangle + \langle \ln(n - n_h)! \rangle - \ln n!, \tag{4.13}$$

which is different from the ensemble entropy. In essence, the point we try to make here is that the ensemble entropy has a subtle difference from its corresponding random variable's entropy, $H[Bin_n(n_h)]$. This situation is more pronounced for the B-model when the probability of configurations writes as

$$P_n(c_i) = \frac{r^{\lfloor n/2 \rfloor - i}}{c_n(r)}, \tag{4.14}$$

whereas the probability of the number of pairs is

$$P_n(n_p) = \frac{\binom{n}{2n_p}(2n_p - 1)!!r^{\lfloor n/2 \rfloor - i}}{c_n(r)}. \tag{4.15}$$

This distinction is the case for the C-model too, and we will be precise to express which entropy we refer to in the following sections.

Before starting, we define some notations that simplify the results further in the following parts. Recall that the Shannon entropy is a functional over the set of probability distributions of system size n

$$H[P_n] : \mathcal{P}_n \to \mathbb{R}^+ \cup \{0\}. \tag{4.16}$$

At the same time, for $k \in \mathbb{N}$, we define a Shannon *function* $H_k(x_1, x_2, \ldots, x_k)$: $[0,1]^k \to \mathbb{R}^+ \cup \{0\}$ as

$$H_k(x_1, x_2, \ldots, x_k) = -\sum_{i=1}^{k} x_i \ln x_i - (1 - \sum_{i=1}^{k} x_i) \ln(1 - \sum_{i=1}^{k} x_i). \tag{4.17}$$

For instance, $H_2(x_1) : [0,1] \to \mathbb{R}^+ \cup \{0\}$ is

$$H_2(x_1) = -x_1 \ln x_1 - (1 - x_1) \ln(1 - x_1), \tag{4.18}$$

and $H_3(x_1, x_2) : [0,1]^2 \to \mathbb{R}^+ \cup \{0\}$ is

$$H_3(x_1, x_2) = -x_1 \ln x_1 - x_2 \ln x_2 - (1 - x_1 - x_2) \ln(1 - x_2 - x_2). \tag{4.19}$$

Using this notation, the following entropies rewrite as

$$H[(\rho, 1 - \rho)] = H_2(\rho), \tag{4.20}$$

$$H_B[(P_n(X_l))] = H_2(\langle \frac{n_p}{n} \rangle), \tag{4.21}$$

$$H_C[(P_n(X_l))] = H_2(\langle \frac{n_p}{n} \rangle) + \langle \frac{n_2}{n} \rangle H_2(\rho), \tag{4.22}$$

and

$$H[Ber_n] = nH_2(\rho). \tag{4.23}$$

4.1.1 Ensemble Entropy

This section derives the ensemble entropy of pairing models. In appendix (C.1.2), in equation (C.9), we see that the ensemble entropy for the B-model is

$$H_B[P_n] = \langle n_p \rangle_n \ln r + \sum_{i=1}^{n-1} \ln(1 + \frac{\langle n_s \rangle_i}{r}), \tag{4.24}$$

in which, $\langle n_s \rangle_x$ denotes the average number of stand-alone elements for a system size x. Hence, it must be recognisable in much the same way the kth moment was derived in equation (3.111), here, the entropy is expressed in terms of the first moment of smaller system sizes.

Likewise, for the C-model, equation (C.16) finds the ensemble entropy as

$$H_C[P_n] = H_B[P_n] + \langle n_s \rangle H_2(\rho). \tag{4.25}$$

The term $H_2(\rho)$ on the right-hand side is the entropy of a Bernoulli sequence in equation (4.23). Therefore, the C-model's entropy decomposes to the sum of the B-model and the Bernoulli's sequence entropies

$$H_C[P_n] = H_B[P_n] + \langle \frac{n_s}{n} \rangle H[Ber_n]. \tag{4.26}$$

This result proves that the C-model's entropy is an additive quantity of its constructive mechanisms, namely pairing and Bernoulli, even if the $H_B[P_n]$ part might not be additive itself. Besides, the effect of the pairing mechanism is decomposable into a purely pairs entropy ($H_B[P_n]$) and a normal binary states entropy ($H[Ber_n]$) whenever the Bernoulli part is scaled by the average ratio of elements in a stand-alone state. This result is understandable since only these elements are in a head or tail state.

Recall that r is a free parameter that defines the ratio of abundance of stand-alone elements to pair ones. Accordingly, the limit $r \to \infty$ is equivalent to removing the pairing mechanism from the model, whereas $r \to 0$ forces all elements into a pair state. In the following parts, we first look at both cases and study their effect on entropies, and further, we will explain a qualitative picture of the obtained results.

4.1.1.1 The limiting case $r \to \infty$

Without the pairing mechanism, all the B-model elements are in a stand-alone state, so only one configuration must be allowed, and consequently, the entropy must be zero. To show that, let us start from the probability distribution of this case. Since the probability of a configuration with n_p pairs is

$$P_n(c_{n_p}) = \frac{r^{\lfloor \frac{n}{2} \rfloor - n_p}}{\sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{2i}(2i-1)!!r^{\lfloor n/2 \rfloor - i}}, \tag{4.27}$$

then, for constant n

$$\lim_{r\to\infty} P_n(c_{n_p}) = \lim_{r\to\infty} \frac{r^{\lfloor \frac{n}{2} \rfloor - n_p}}{r^{\lfloor n/2 \rfloor}} = \begin{cases} 1 & , n_p = 0 \\ 0 & , n_p > 0 \end{cases} . \tag{4.28}$$

Clearly, from the definition of entropy ($-\sum_i p_i \ln p_i$) the last result implies

$$\lim_{r\to\infty} H[P_n(c_{n_p})] = 0. \tag{4.29}$$

Also, equation (4.28) asserts that we must have

$$\langle n_s \rangle_n = n + O(\frac{1}{r}), \qquad \langle n_p \rangle_n = 0 + O(\frac{1}{r}), \tag{4.30}$$

and hence, for constant n, we get the same result from equation (4.24)

$$\lim_{r\to\infty} \left[\langle n_p \rangle_n \ln r + \sum_{i=1}^{n-1} \ln(1 + \frac{\langle n_s \rangle_i}{r}) \right] \to 0, \tag{4.31}$$

as expected. For the B-model, the probability of observing a configuration with n_p pairs and n_h heads is

$$P_n(c_{n_p,n_h}) = \frac{r^{\lfloor \frac{n}{2} \rfloor - i} \rho^j (1-\rho)^{n-2i-j}}{\sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{2i}(2i-1)!!r^{\lfloor n/2 \rfloor - i}}, \tag{4.32}$$

and therefore,

$$\lim_{r\to\infty} P_n(c_{n_p,n_h}) = \lim_{r\to\infty} \frac{r^{\lfloor \frac{n}{2} \rfloor - n_p}}{r^{\lfloor n/2 \rfloor}} \rho^{n_h}(1-\rho)^{n-2n_p-n_h}$$

$$= \begin{cases} \rho^{n_h}(1-\rho)^{n-n_h} & , n_p = 0 \\ 0 & , n_p > 0 \end{cases}, \tag{4.33}$$

which is the probability of observing a sequence of Bernoulli random variables with n_h heads. To check that in the limit $H_C[P_n]$ is equal to the Bernoulli entropy, recall the average number of heads in the Binomial distribution is $\langle n_h \rangle_n = n\rho$. Meanwhile, for this limiting case, we showed $H_B[P_n] = 0$ and $\langle n_s \rangle_n = n$ in equations (4.30) and (4.29), respectively. Hence, equation (4.26) writes

$$\lim_{r\to\infty} H_C[P_n] = H[Ber_n]. \tag{4.34}$$

4.1.1.2 The limiting case $r \to 0$

Similar to the previous section, for $r \to 0$ and constant n, we have

$$\lim_{r\to 0} P_n(c_{n_p}) = \lim_{r\to 0} \frac{r^{\lfloor \frac{n}{2} \rfloor - n_p}}{\sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{2i}(2i-1)!!r^{\lfloor n/2 \rfloor - i}} = \begin{cases} \frac{1}{(n-1)!!} & , n_p = \lfloor \frac{n}{2} \rfloor \\ 0 & , \text{ otherwise} \end{cases}, \tag{4.35}$$

[1] and

$$\langle n_s \rangle_n = 0 + O(r), \qquad \langle n_p \rangle_n = \lfloor \frac{n}{2} \rfloor + O(r), \tag{4.36}$$

which for $(n-1)!!$ distinct configurations with equal probabilities we get

$$\lim_{r\to 0} H_B[P_n] = \ln(n-1)!!. \tag{4.37}$$

As a side note, it is not clear at first sight how in the limit $r \to 0$, equation (4.24) becomes zero. Therefore, we have to return to equation (C.7), from which we derived

[1] In fact, for even $2n$s it finds as

$$\binom{2n}{2n}(2n-1)!! = (2n-1)!!,$$

and for odd $2n-1$, it is

$$\binom{2n-1}{2\lfloor \frac{2n-1}{2} \rfloor}(2\lfloor \frac{2n-1}{2} \rfloor - 1)!! = (2n-1)!!.$$

So, for every two consecutive odd and even numbers, we get the same limit.

equation (4.24)

$$H_B[P_n] = \Big[\langle n_p\rangle_n - \lfloor\frac{n}{2}\rfloor\Big]\ln r + \ln c_n(r). \tag{4.38}$$

We see that in the limit $r \to 0$, the term $\big[\langle n_p\rangle_n - \lfloor\frac{n}{2}\rfloor\big]\ln r$ approaches zero, since the $\langle n_p\rangle_n - \lfloor\frac{n}{2}\rfloor$ has the order $O(r)$. Meanwhile,

$$\lim_{r\to 0} c_n(r) \to (n-1)!!, \tag{4.39}$$

which implies

$$\lim_{r\to 0} H_B[P_n] = \lim_{r\to 0}\Big[\langle n_p\rangle_n - \lfloor\frac{n}{2}\rfloor\Big]\ln r + \ln c_n(r) \to \ln(n-1)!!, \tag{4.40}$$

as expected.

For the C-model, when n is even[2], we get

$$\begin{aligned}\lim_{r\to 0} P_n(c_{n_p,n_h}) &= \lim_{r\to 0}\frac{r^{\frac{n}{2}-n_p}\rho^{n_h}(1-\rho)^{n-2n_p-n_h}}{\sum_{i=0}^{n/2}\binom{n}{2i}(2i-1)!!r^{n/2-i}} \\ &= \begin{cases}\frac{1}{(n-1)!!} & , n_p = \frac{n}{2}, n_h = 0 \\ 0 & , \text{ otherwise}\end{cases}.\end{aligned} \tag{4.43}$$

Consequently,

$$\langle n_s\rangle_n = 0 + O(r), \qquad \langle n_p\rangle_n = \frac{n}{2} + O(r), \qquad \langle n_h\rangle_n = 0 + O(r) \tag{4.44}$$

and

$$\lim_{r\to 0} H_C[P_n] = \ln(n-1)!!. \tag{4.45}$$

[2]For odd n we have

$$\lim_{r\to 0} P_n(c_{n_p,n_h}) = \begin{cases}\frac{\rho}{(n-1)!!} & , n_p = \lfloor\frac{n}{2}\rfloor, n_h = 1 \\ \frac{(1-\rho)}{(n-1)!!} & , n_p = \lfloor\frac{n}{2}\rfloor, n_h = 0 \\ 0 & , \text{ otherwise}\end{cases}, \tag{4.41}$$

and

$$\lim_{r\to 0} H_C[P_n] = \ln(n-1)!! + H_2(\rho). \tag{4.42}$$

4.1.1.3 Maximum of $H_B[P_n]$

$H_B[P_n]$ is an r-dependent, non-negative, continuous function, and let us say its maximum value happens in its domain at $r^* \in [0, \infty)$. In appendix (C.1.4), equation (C.19) finds the derivative of $H_B[P_n]$, and by setting it equal to zero, we find its maximum location for constant n as

$$\frac{dH_B[P_n]}{dr} = (\langle n_p \rangle_n^2 - \langle n_p^2 \rangle_n) \ln r = 0 \implies r^* = 1. \tag{4.46}$$

Recall that r equal to one corresponds to a uniform distribution over all the configurations.

4.1.1.4 Ensemble entropy: Conclusion

Previous results show that the ensemble entropies for both models are well defined. Moreover, the choice of the parameter r controls the strength of the pairing mechanism on the entropy, *e.g.*, in the limit $r \to \infty$, the pairing mechanism is removed.

To have a qualitative understanding of the calculated quantities, figure (4.1) shows the entropy of the B-model for different system sizes. We see the entropy is approaching zero for $r \to \infty$, and is maximum at $r^* = 1$, whilst it is equal to $\ln(2n-1)!!$ for both sizes $2n - 1$ and $2n$.

Also, we can see in figure (**??**) entropy increases when the system size increases. However, unlike entropies in Cartesian spaces, it does not grow linearly with the system size. To elaborate, recall the entropy of a Bernoulli sequence is

$$H[Ber_n] = nH_2(\rho), \tag{4.47}$$

and it has a Cartesian state space. Consequently, since the term $H_2(\rho)$ is the entropy of a single element and is constant, Bernoulli's entropy is of order $H[Ber_n] = O(n)$. On the other hand, in appendix (C.2), equation (C.22) obtains the asymptotic leading terms of $H_B[P_n]$ as

$$H_B[P_n] \sim \frac{n}{2} \ln \frac{n}{r} + O(n), \tag{4.48}$$

and we see its order is $O(n \ln n)$.

Figure (4.2) plots the ensemble entropy of the C-model with respect to ρ. For $\rho = 0$

or $\rho = 1$, we observe $H_C[P_n]$ is the same as $H_B[P_n]$ in figure (4.1). Meanwhile, changing ρ from these two extremes moves the maximum from $r^* = 1$ to higher values. Recall that $r^* = 1$ is where $H_B[P_n]$ is maximum. And finally, the overall entropy is the highest for $\rho = 1/2$ in comparison to other values.

As discussed for the C-model, even when the pairing mechanism is present, the entropy can be separated into the contribution of pairing random variables and Bernoulli random variables, so as it manifests itself as head or tail states for stand-alone elements. Defining the difference between $H_C[P_n]$ and Bernoulli part as

$$G(r, \rho) \equiv H_C[P_n] - \langle \frac{n_s}{n} \rangle_n H[Ber_n] = H_B[P_n], \tag{4.49}$$

we obtain a ρ-independent, continuous function. This fact is depicted in figure (4.3). Note that since $G(r, \rho)$ is independent of ρ, one might say that all curves overlap for different values of ρ, which represents a universality class among all C-models.

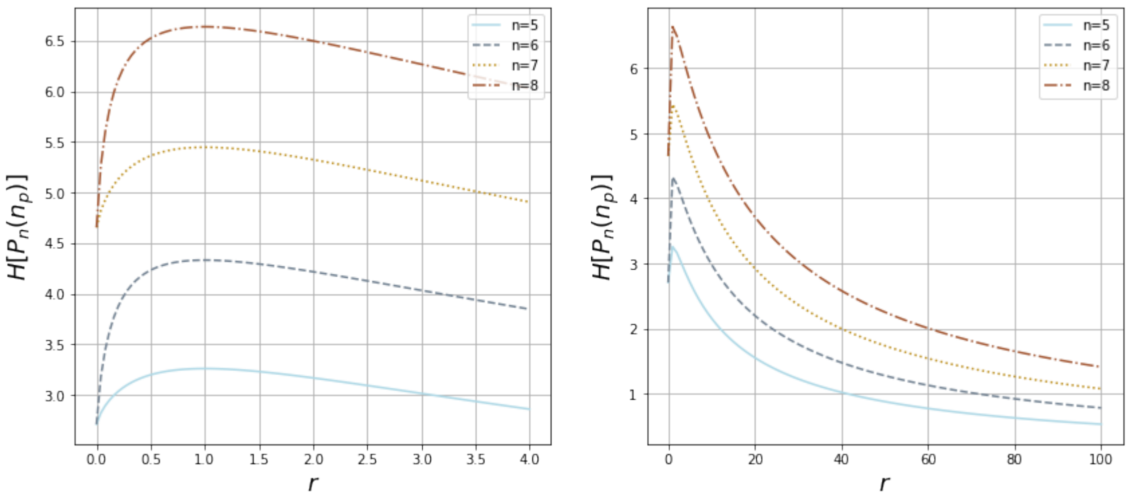


Figure 4.1: The B-model's Shannon entropy, $H[P_n(n_p)]$, for different system sizes. Notice that the maximum is at $r = 1$, and for $r \to \infty$, it approaches zero.

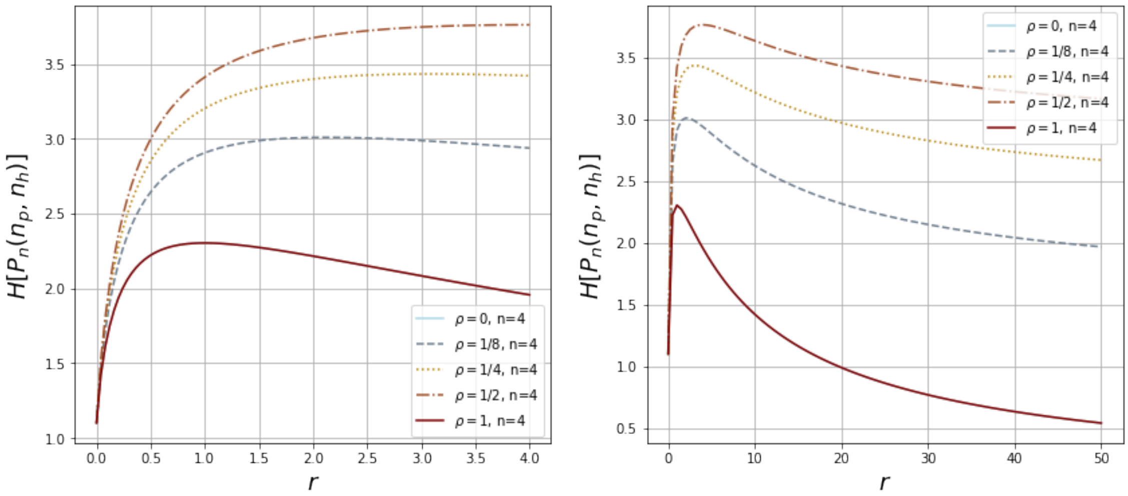


Figure 4.2: The C-model's Shannon entropy, $H[P_n(n_p, n_h)]$ for different ρs. Notice that the curve is the same for $\rho = 0$ and $\rho = 1$. Also, ρ shifts the position of the maximum entropy to $r > 1$.

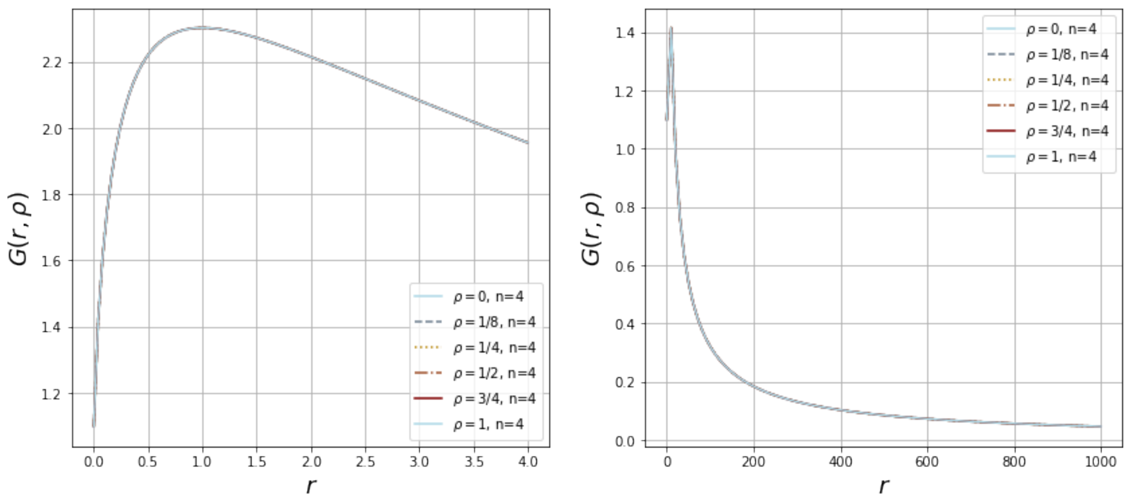


Figure 4.3: The plot of $G(r, \rho)$ for different values of ρ. As it is depicted, $G(r, \rho)$ represents a universality property and is independent of ρ.

4.1.2 Joint Entropy

Recall that in equation (3.155), we derived the joint probability distribution for k arbitrary elements as

$$P_n(X_1, \dots, X_k; l) = \langle \frac{2n_p^{(l)} n_s^{(k-l)}}{n^{(k)}} \rangle, \tag{4.50}$$

when $n^{(k)}$ is a falling factorial and l out of k elements are in a pair state, *i.e.*, for a given k-tuple $(X_1, \dots, X_k)$, we have $l = k - \sum_{i=1}^{k} X_i$. And also, given l, there are $\binom{k}{l}$ possible combinations among the arguments of $P_n(X_1, \dots, X_k; l)$. Hence, the joint entropy derives as

$$\begin{aligned} H[P_n(X_1, \dots, X_k)] &= -\sum_{l=0}^{k} \binom{k}{l} P_n(X_1, \dots, X_k; l) \ln P_n(X_1, \dots, X_k; l) \\ &= -\sum_{l=0}^{k} \binom{k}{l} \langle \frac{2n_p^{(l)} n_s^{(k-l)}}{n^{(k)}} \rangle \ln \langle \frac{2n_p^{(l)} n_s^{(k-l)}}{n^{(k)}} \rangle. \end{aligned} \tag{4.51}$$

Although, in theory, this result is the joint entropy in closed form, no clear insight can be gained about its property. At the same time, as discussed in the introduction, for $k = O(1)$, the joint entropy of such a small subsystem is a microscopic quantity compared to the ensemble entropy. It is needless to say that the ensemble entropy is equal to the joint probability for $k = n$. Based on this fact, it is interesting to find the asymptotic behaviour of the joint entropy, especially for the case $k \sim O(\sqrt{n})$, in which we categorise a *mesoscopic* quantity.

Accordingly, for $1 \ll n$, and by using equations (B.82) and (B.86), the asymptotic leading terms of the ratio of pairs and stand-alone elements are

$$\langle \left(\frac{2n_p}{n}\right)^k \rangle \sim e^{-k\sqrt{r/n}} + O(\frac{1}{n}), \qquad \langle \left(\frac{n_s}{n}\right)^k \rangle \sim \left(1 - e^{-\sqrt{r/n}}\right)^k + O(\frac{1}{n}). \tag{4.52}$$

However, to find the asymptotic leading term of the joint entropy, one needs to evaluate the asymptotic of the falling factorial. So, we have

$$\frac{2np^{(l)} n_s^{(k-l)}}{n^{(k)}} = \frac{2n_p(2n_p - 1) \dots (2np - l + 1) \times n_s(n_s - 1) \dots (n_s - k + l + 1)}{n(n-1) \dots (n - k + 1)}$$

$$= \frac{(2n_p)^l n_s^{k-l} + O(n_p^{l-1} n_s^{k-l-1})}{n^k + O(n^{k-1})} \sim \left(\frac{2n_p}{n}\right)^l \left(\frac{n_s}{n}\right)^{k-l} + O(\frac{1}{n}) \implies$$

$$\langle \frac{2np^{(l)} n_s^{(k-l)}}{n^{(k)}} \rangle \sim \langle \left(\frac{2n_p}{n}\right)^l \rangle \langle \left(\frac{n_s}{n}\right)^{k-l} \rangle + O(\frac{1}{n})$$

$$\sim e^{-l\sqrt{r/n}} \left(1 - e^{-\sqrt{r/n}}\right)^{k-l} + O(\frac{1}{n}). \tag{4.53}$$

Using the last result and equation (4.51), we get

$$H[P_n(X_1, \ldots, X_k)] \sim \sum_{l=0}^{k} \binom{k}{l} l \sqrt{\frac{r}{n}} e^{-l\sqrt{\frac{r}{n}}} \left(1 - e^{-\sqrt{r/n}}\right)^{k-l}$$

$$- \sum_{l=0}^{k} \binom{k}{l} (k-l) e^{-l\sqrt{r/n}} \left(1 - e^{-\sqrt{r/n}}\right)^{k-l} \ln\left(1 - e^{-\sqrt{\frac{r}{n}}}\right), \tag{4.54}$$

which in appendix (C.3), equation (C.28) finds it as

$$H[P_n(X_1, \ldots, X_k)] = k\sqrt{\frac{r}{n}} \left[1 - \ln\sqrt{\frac{r}{n}}\right] + O(\frac{1}{n}). \tag{4.55}$$

Before discussing the consequences of the last result, let us re-examine the interpretation of the joint entropy. Suppose we are studying a subsystem composed of k elements in a system with n constituent elements. For a system with a Cartesian state space and statistically independent elements, the subsystems' state is unaffected by the rest of the $n-k$ other elements. Moreover, technically, the probability distribution of the whole system is the multiplication of the distribution of both k and $n-k$ subsystems.

However, for non-Cartesian state spaces and statistically dependent subsystems, one must first derive the subsystem's joint distribution by marginalising the rest of the system to obtain the subsystem's entropy. In other words, the subsystem's entropy is identical to the joint entropy of k elements. For example, this is the case for pairing systems.

Considering this preliminary explanation, equation (4.55) finds the entropy of a subsystem whenever the whole system is macroscopic ($1 \ll n$). For $k = 1$, we get the entropy of a single element, which is exactly the same as equation (4.7). And as it is shown in appendix (C.3), equation (C.26) derives the joint entropy in terms of the entropy of a single element as

$$H[P_n(X_1,\dots,X_k)] = kH_B[P_n(X_1)] + O(\frac{1}{n}). \tag{4.56}$$

Therefore, for macroscopic systems, the entropy of subsystems is the additive quantity of its elements' entropy

$$H[P_n(X_1,\dots,X_k)] = \sum_{i=1}^{k} H_B[P_n(X_i)] + O(\frac{1}{n}). \tag{4.57}$$

We must stress this result is valid up to mesoscopic subsystem sizes, or $k = O(\sqrt{n})$. It is explained in appendix (B.9), the asymptotic leading term of $\langle n_p^k \rangle$ and $\langle n_s^k \rangle$ are valid for $k = O(\sqrt{n})$. Consequently, the additive result that we get here is valid up to mesoscopic subsystem sizes.

We must remind ourselves, unlike statistically independent elements, that the entropy of microscopic subsystems is system-size-dependent. In addition, when we repeat equation (4.7) here

$$H_B[P_n(X_l)] = \sqrt{\frac{r}{n}}\left[1 - \ln\sqrt{\frac{r}{n}}\right] + O(\frac{1}{n}), \tag{4.58}$$

we see that in thermodynamic limit $n \to \infty$, the entropy of a single element approaches zero. Consequently, for subsystems with size $k = o(\sqrt{n})$, the entropy approaches zero too. But, for subsystems size $k = O(\sqrt{n})$ the entropy is finite

$$H[P_n(X_1,\dots,X_k)] = \sqrt{r}, \qquad k = O(\sqrt{n}). \tag{4.59}$$

Nonetheless, the macroscopic entropy has an order $O(n \ln n)$. So, the additivity property cannot stay intact from mesoscopic to macroscopic sizes; what happens in between is an open question that needs careful analysis of the asymptotic terms for order other than $O(\sqrt{n})$.

4.1.3 Mutual Information

In much of the same way as the previous section, suppose a system composed of pairing balls or coins is divided into two subsystems. When a pair happens between elements in different subsystems, the pairing link can be considered an interaction between subsystems. So it is illuminating to study the mutual information between two subsystems and find its dependence in inter-system pairs. We found the system's ensemble entropy and its subsystems in the previous section as a measure of uncertainty/information about the system and its sub-components. Nevertheless, the mutual information is equivalent to the information content of one subsystem about the other one [61].

This part will derive the mutual information of two kinds of system decomposition for a system of pairing balls. In the first case, a system with n elements, namely S_n, is divided into S_{n-1} and S_1 subsystems, containing $n-1$ and a single element, respectively. We denote the mutual information by $I_n(S_{n-1}, S_1)$, and eventually, since all elements are identical, the mutual information of each is the same with respect to every other one as a whole. Also, for $1 \ll n$, $I_n(S_{n-1}, S_1)$ is the information contribution of a microscopic element to a macroscopic one.

In the second case, we will derive the mutual information between two arbitrary balls, or $I_n(S_1, S_1)$. One can do further and derive higher-order interactions by including three or more elements. In general, interaction information [72] is defined as

$$I_n(X_1, X_2, \ldots, X_k) = \sum_{T \subseteq \{X_1, X_2, \ldots, X_k\}} (-1)^{|T|-1} H_B[T], \tag{4.60}$$

where T runs over all the subset of $\{X_1, X_2, \ldots, X_k\}$, and $|T|$ is its cardinality. Note that, for $n = 2$, the interaction information is identical to the mutual information.

4.1.3.1 Finding $I_n(S_{n-1}, S_1)$

The mutual information of the subsystems S_{n-1} and S_1 is defined as

$$I_n(S_{n-1}, S_1) = H_B[P_n(X_1)] - H_B[P_n(X_1|n_p)], \tag{4.61}$$

where $H[P_n(X_1)]$ is the entropy of a single element and $H[P_n(X_1|n_p)]$ is the entropy of a single element conditioned on the remaining part of the system. It is necessary to write down the marginal distribution of a single element to find the first term and the conditional distribution of a single element given the number of pairs for the second term. We have already obtained the marginal entropy in equation (4.21) as

$$H_B[P_n(X_1)] = H_2(\langle \frac{n_p}{n} \rangle). \tag{4.62}$$

Note that $H_2(\langle\frac{2n_p}{n}\rangle)$ is the Shannon function evaluated at $\langle\frac{2n_p}{n}\rangle \in [0,1]$ – equation (4.18). In other words, if we interpret $\langle\frac{2n_p}{n}\rangle$ like a probability, $H_2(\langle\frac{2n_p}{n}\rangle)$ is its Shannon entropy.

However, the quantity $\langle\frac{2n_p}{n}\rangle$ is a system-wide average and a global quantity for the system, and in the case $1 \ll n$, this is a macroscopic quantity. So we can say that a single element's entropy is equal to the Shannon entropy evaluated at the value of the macroscopic quantity $\langle\frac{2n_p}{n}\rangle$ for the whole system. To emphasis the size-dependency, we use the subscript on $\langle\frac{2n_p}{n}\rangle_n$ and write

$$H_B[P_n(X_1)] = H_2\left(\langle\frac{2n_p}{n}\rangle_n\right). \tag{4.63}$$

Next, equation (3.148) derives the conditional distribution as

$$P_n(X_1|n_p) = \begin{cases} \frac{2n_p}{n} & ,X_1 = 0 \\ \frac{n-2n_p}{n} & ,X_1 = 1 \end{cases}, \tag{4.64}$$

and the conditional entropy, $H[P_n(X_1|n_p)]$, must be equal to

$$H[P_n(X_1|n_p)] = -\sum_{n_p=0}^{\lfloor n/2\rfloor} P_n(n_p) \sum_{X_l=0}^{1} P_n(X_1|n_p)\ln P_n(X_1|n_p)$$

$$= -\sum_{n_p=0}^{\lfloor n/2\rfloor} \binom{n}{2n_p}(2n_p-1)!!\frac{r^{\lfloor n/2\rfloor-n_p}}{c_n(r)}\left[\frac{2n_p}{n}\ln\frac{2n_p}{n} + (1-\frac{2n_p}{n})\ln(1-\frac{2n_p}{n})\right]$$

$$= \langle -\frac{2n_p}{n}\ln\frac{2n_p}{n} - (1-\frac{2n_p}{n})\ln(1-\frac{2n_p}{n})\rangle_n = \langle H_2(\frac{2n_p}{n})\rangle_n. \tag{4.65}$$

Finally, the mutual information writes as

$$I_n(S_{n-1}, S_1) = H_2\left(\langle\frac{2n_p}{n}\rangle_n\right) - \left\langle H_2(\frac{2n_p}{n})\right\rangle_n. \tag{4.66}$$

To put it as a sentence, the mutual information between each element and the rest of the system is equal to the Shannon function of the macroscopic quantity $\langle\frac{2n_p}{n}\rangle_n$ minus the ensemble average of the Shannon function at $\frac{2n_p}{n}$ of the microstates.

In Appendix (C.5), equation (C.37) expands $I_n(S_{n-1}, S_1)$ as a power series in terms of $\langle 2n_p/n\rangle$ and $\langle n_s/n\rangle$ like

$$I_n(S_{n-1}, S_1) = \sum_{k\geq 2} \frac{\langle (\frac{2n_p}{n})^k\rangle - \langle \frac{2n_p}{n}\rangle^k + \langle (\frac{n_s}{n})^k\rangle - \langle \frac{n_s}{n}\rangle^k}{k(k-1)}. \tag{4.67}$$

This power series is converging fast, by which one can find a numerical estimation for both $I_n(S_{n-1}, S_1)$ and $\langle H_2(\frac{2n_p}{n})\rangle_n$. Check appendix (C.5) for details. Also, equation (C.40) obtains the asymptotic leading term as

$$I_n(S_{n-1}, S_1) \sim \frac{1}{n} + O(\frac{1}{n^2}). \tag{4.68}$$

Subsequently, this result implies that increasing the system size reduces the mutual information, and in the limit $n \to \infty$, it approaches zero. In other words, the information about S_1 reduces the uncertainty about the whole system. Nevertheless, when the system size increases, the information about the whole system from a single element is negligible, as expected.

4.1.3.2 Finding Interaction Information

This part first finds the mutual information, $I_n(S_1, S_1)$, and then moves to derive the interaction information. For the B-model, the mutual information between two elements equals

$$\begin{aligned} I_n(S_1, S_1) &= H_B[P_n(X_1)] + H_B[P_n(X_2)] - H_B[P_n(X_1, X_2)], \\ &= 2H_B[P_n(X_1)] - H_B[P_n(X_1, X_2)], \end{aligned} \tag{4.69}$$

where X_1 and X_2 are the random variables representing the states of two arbitrary indices. Note that, since the marginal of a single element is the same for all indices, we used $H_B[P_n(X_1)] + H_B[P_n(X_2)] = 2H_B[P_n(X_1)]$ in the last step.

Equation (3.152) derives the marginal for two elements as

$$P_n(X_1, X_2) = \begin{cases} \langle \frac{2n_p(2n_p-1)}{n(n-1)}\rangle & , X_l = X_k = 0 \\ \langle \frac{2n_p n_s}{n(n-1)}\rangle & , X_l = 1, X_k = 0 \\ \langle \frac{2n_p n_s}{n(n-1)}\rangle & , X_l = 0, X_k = 1 \\ \langle \frac{n_s(n_s-1)}{n(n-1)}\rangle & , X_l = X_k = 1 \end{cases}. \tag{4.70}$$

Observe that the terms in $P_n(X_1, X_2)$ entries show a pattern. For example, the term

$\langle 2n_p(2n_p-1)\rangle_n$ is quadratic in $2n_p$, but it is neither a second moment nor a cumulant. Instead, the falling factorial $\langle 2n_p(2n_p-1)\rangle_n$ equals the correlation between $2n_p$, the number of pairs in the system, and $2n_p - 1$, the number of pairs after removing the first one. Moreover, the identity

$$\langle 2n_p(2n_p-1)\rangle_n + 2\langle 2n_p n_s\rangle_n + \langle n_s(n_s-1)\rangle_n = n(n-1), \tag{4.71}$$

implies for a system size n, knowing two of these quantities are enough to find the third one. So that, using the definition of $H_3(x, y)$, the joint entropy writes as

$$H[P_n(X_1, X_2)] = H_3\left(\langle \frac{2n_p}{n}\frac{(2n_p-1)}{(n-1)}\rangle_n, \langle \frac{n_s}{n}\frac{(n_s-1)}{(n-1)}\rangle_n\right). \tag{4.72}$$

Finally, the mutual entropy is

$$I_n(S_1, S_1) = 2H_2\left(\langle \frac{2n_p}{n}\rangle_n\right) - H_3\left(\langle \frac{2n_p}{n}\frac{(2n_p-1)}{(n-1)}\rangle_n, \langle \frac{n_s}{n}\frac{(n_s-1)}{(n-1)}\rangle_n\right). \tag{4.73}$$

Asymptotically, for $1 \ll n$, equation (4.56) writes the asymptotic expansion of $H[P_n(X_1, X_2)]$ as

$$H[P_n(X_1, X_2)] = 2H_B[P_n(X_1)] + O(\frac{1}{n}), \tag{4.74}$$

and therefore,

$$I_n(S_1, S_1) = 0 + O(\frac{1}{n}). \tag{4.75}$$

In short, the mutual information is negligible for large systems, and the knowledge about an element does not obtain any information about another randomly selected one.

Next, to derive higher-order interactions by including three or more elements, we derive the interaction information or information correlation [72], which is defined as

$$I_n(X_1, X_2, \ldots, X_k) = \sum_{T\subseteq\{X_1, X_2, \ldots, X_k\}} (-1)^{|T|-1} H_B[T], \tag{4.76}$$

where T runs over all the subset of $\{X_1, X_2, \ldots, X_k\}$, and $|T|$ is its cardinality. Observe that the marginal in equation (3.155) is the same for all elements in different indices

$$P_n(X_1, \ldots, X_k;\, l) = \langle \frac{2np^{(l)}n_s^{(k-l)}}{n^{(k)}}\rangle. \tag{4.77}$$

Using this fact and considering the cardinality of the subsets, say $\binom{k}{i}$, the interaction information equals to

$$I(X_1, X_2, \ldots, X_k) = \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} H_B[P_n(X_1, \ldots, X_i)]. \tag{4.78}$$

But, the joint entropy $H_B[P_n(X_1, \ldots, X_i)]$ writes as

$$H[P_n(X_1, \ldots, X_i)] = -\sum_{l=0}^{i} \binom{i}{l} P_n(X_1, \ldots, X_i; l) \ln P_n(X_1, \ldots, X_i; l)$$

$$= -\sum_{l=0}^{i} \binom{i}{l} \langle \frac{2np^{(l)} n_s^{(i-l)}}{n^{(i)}} \rangle \ln \langle \frac{2np^{(l)} n_s^{(i-l)}}{n^{(i)}} \rangle, \tag{4.79}$$

where $\binom{i}{l}$ is the cardinality of the i-tuple when $l = i - \sum_j X_j$. Therefore, The interaction information derives as

$$I_n(X_1, X_2, \ldots, X_k) = -\sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \sum_{l=0}^{i} \binom{i}{l} \langle \frac{2np^{(l)} n_s^{(i-l)}}{n^{(i)}} \rangle \ln \langle \frac{2np^{(l)} n_s^{(i-l)}}{n^{(i)}} \rangle. \tag{4.80}$$

Again, no insight is directly gained from this result. However, we can study the asymptotic form of the interaction information, and by using equation (4.56), the asymptotic expansion of $H[P_n(X_1, \ldots, X_k)]$ is equal to

$$H[P_n(X_1, \ldots, X_k)] = kH_B[P_n(X_1)] + O(\frac{1}{n}). \tag{4.81}$$

Therefore, the asymptotic expansion of the interaction information obtains as

$$I_n(X_1, X_2, \ldots, X_k) = \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} iH_B[P_n(X_1)] + O(\frac{1}{n})$$

$$= kH_B[P_n(X_1)] \sum_{i=1}^{k} \binom{k-1}{i-1} (-1)^{i-1} + O(\frac{1}{n})$$

$$kH_B[P_n(X_1)](1-1)^{k-1} + O(\frac{1}{n}) = 0 + O(\frac{1}{n}). \tag{4.82}$$

So, for a large system size, the interaction information is negligible for any number of elements up to the mesoscopic scale, $k \sim O(\sqrt{n})$. One can say pairing models do not exhibit higher-order information correlation. It seems reasonable since the pairs are the most complex entity constructed in the pairing mechanism. So, for large system sizes, pairs and stand-alone elements are the only atomic structures that exist. It is an interesting open question to study the interaction information when we include a variety of compounding mechanisms.

4.1.4 Non-extensive Entropies

Both B and C probability distributions are polynomials of a single variable r. For entropies written in terms of probabilities' power, such as Rényi's and Tsallis entropies, this form is conveniently written in closed form. Also, later in the coming section for pairing time series, we will discuss the non-extensivity property of Rényi's and Tsallis entropies, and therefore, it is helpful to write them in terms of pairing models probability distribution.

4.1.4.1 Tsallis Entropy

Tsallis's entropy [67] is defined as

$$H_q[P_n] = \frac{1}{q-1}\left(1 - \sum_{i=1}^{W(n)} p_i^q\right), \tag{4.83}$$

for positive $q \neq 1$, where $W(n)$ is the number of elements in the state space. In appendix (C.7), equation (C.42), for B-model we find

$$H_q[P_n] = \frac{1}{q-1}\left(1 - \frac{c_n(r^q)}{c_n^q(r)}\right). \tag{4.84}$$

For the C-Model, equation (C.45) derives it as

$$H_\alpha[P_n] = \frac{1}{q-1}\left(1 - \frac{c_n(r^q\left[\rho^q + (1-\rho)^q\right]^2)}{c_n^q(r)}\right). \tag{4.85}$$

4.1.4.2 Rényi's Entropy

Rényi's entropy [56] is defined as

$$H_\alpha[P_n] = -\frac{1}{1-\alpha}\ln\left(\sum_{i=1}^{W(n)} p_i^\alpha\right), \tag{4.86}$$

for positive $\alpha \neq 1$, where $W(n)$ is the number of elements in the state space. In appendix (C.6), equation (C.41), for B-model we find

$$H_\alpha[P_n] = -\frac{\ln c_n(r^\alpha) - \alpha \ln c_n(r)}{1-\alpha}. \tag{4.87}$$

Recall that $c_n(r^\alpha)$ is a polynomial degree n, evaluates at r^α. And for the C-Model, it derives as

$$H_\alpha[P_n] = -\frac{\ln c_n(r^\alpha \left[\rho^\alpha + (1-\rho)^\alpha\right]^2) - \alpha c_n(r)}{1-\alpha}. \tag{4.88}$$

Again, $c_n(r^\alpha \left[\rho^\alpha + (1-\rho)^\alpha\right]^2)$ is a polynomial degree n, evaluates at $r^\alpha \left[\rho^\alpha + (1-\rho)^\alpha\right]^2$.

4.2 Pairing Time Series

4.2.1 Introduction

Pairing models are intuitively simple and fortunately have quantities that can be expressed in closed form. This combination opens up opportunities to use them as the building block of other mathematical models. To show that, we will propose a transmission model with a pairing time series.

The pairing time series distinguishes from the ordinary time series so that the current state of the received signal can carry a definite value of the state of the signal in the future. To a certain extent, one can find time series with similar properties in the real world.

For example, imagine a faulty mechanical clock such that now and then its hand ticks backwards. So, the clock's time series can be modelled as a Bernoulli sequence of successful or failed ticks – see figure (4.4). Next, suppose a mechanical constraint like a cogwheel forces the clock to fail once again exactly after one rotation of the wheel, or equivalently, exactly after some pre-determined number of ticks. Therefore, regarding the clock behaviour, the first failed tick indicates the subsequent failure in the future, although other backward ticks can happen by chance between these two moments.

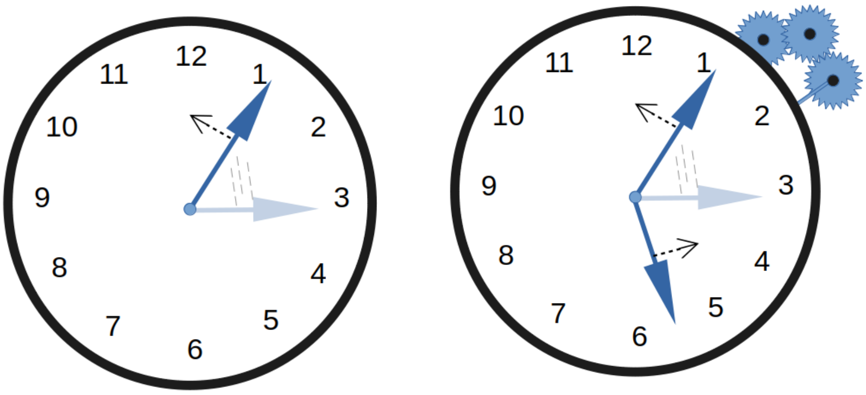


Figure 4.4: Left panel: A faulty clock that randomly fails. **Right panel:** A faulty clock repeats itself. After the first backward tick, the next one happens at a pre-determined number of ticks since the cogwheel mechanism forces it to repeat itself.

We can even think of more than one cogwheel and more than one type of failed state. Thus, different types of backward ticks repeat once more with different duration or, say, make a pair with a different moment in future. Nevertheless, the pairing time series works the same.

While such a paring time series can have applications in the real world, at the same time, we shall see under some assumptions for infinite time series that the Shannon entropy of a single moment becomes undefined. Also, without providing proof, we will propose a scheme that might resolve this problem with non-extensive entropies. Despite this failure, we have to remark that the Shannon entropy is well behaved for a large class of conditions, and a pairing time series can be utilised in modelling.

Non-extensive entropies have been introduced in publications for many years [20], and the criteria to choose them are not clear cut. However, at least for this model, there is no doubt that the Shannon entropy is problematic, and non-extensive ones might have their merits.

All in all, the pairing time series is an example that paring models, in general, have the potential not just to impose interesting mathematical questions but also to be the building block of practically valuable models. Nevertheless, they can show the limit of the applicability of information-theoretic quantities like Shannon entropy, which is usually accepted as universally valid.

After this preliminary introduction, let us make our definition mathematically precise. Intuitively, a configuration that is composed of pairing coins can be arranged in space. Moreover, when elements are ordered along one dimension, by definition, pairing between any two elements is possible, irrespective of the distance between them.

For example, in figure (4.5), the head and tail states are represented by up and down arrows, respectively, and the pairs that are denoted by P are linked along a one-dimensional string of pairing coins.

Let us replace the tail and head states with 0 and 1. Moreover, pair coins can be replaced by a number that represents the distance between them (figure 4.5, 4.6 and 4.7). So, the set $\mathcal{A}_L = \{0, 1, 2, \ldots, L\}$ is the alphabet that constructs all the admissible strings composed of pairing coins for any configuration with length L.

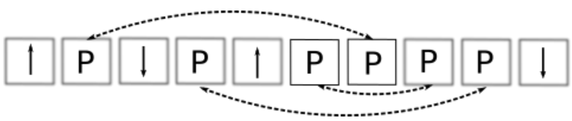


Figure 4.5: An example of paring coins configuration.

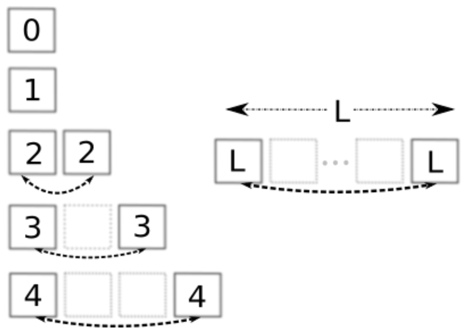


Figure 4.6: Labelling pair coins by the distance between them.

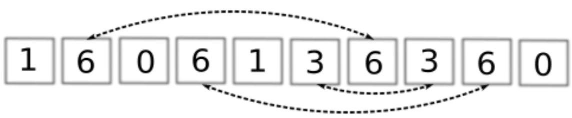


Figure 4.7: Replacing the configuration in figure (4.5) with its elements corresponding alphabets.

Note that the set of all possible strings that one can construct by $\mathcal{A}_L$ is larger than the admissible pairing coins strings. To be precise, the volume of the former is L^{L+1}, and the latter one is $(L/\mathrm{e})^{L/2}$.

For instance, although "222" is a string constructed by elements of $\mathcal{A}_3$, this is not an admissible pairing coin string. The reason for rejecting the admissibility of this configuration is understood if one considers two coins at each end of the configuration link to the middle one. The C-model requires at most two coins in a pair state with each other.

Generally, ordering binary random variables in time constructs a stochastic time series. Let's say there is a receiver that can register the arrival of these random variables in discrete time steps. So, each recorded string is a realisation of the stochastic process. Then, we may ask, is it possible to order a pairing coin string in time? What does that mean to say two random variables make a pair by each other along the arrow of time? To elaborate on the new model, let us assume that $X_t \in \mathcal{A}_L$ is a random variable that the subscript t denotes its time-step index. In other words, it is ordered in time, and the receiver gets a new input at each step t. See figure (4.8).

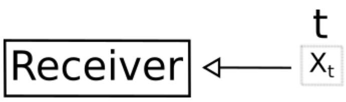


Figure 4.8: At the time t, the random viable X_t arrives at the receiver.

When $X_t \in \{0, 1\}$, we have a usual binary random variable. However, when $X_t > 1$, the future value at index $t + X_t - 1$ will be the same as the current one (figure 4.9). In other words, there is no uncertainty at time step $t + X_t - 1$ whenever the receiver registers a random variable $X_t > 1$ at time t. This assumption is equivalent to having the two states as a pair.

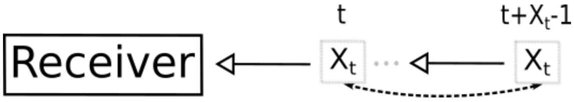


Figure 4.9: A random variable X_t can make a pair with the one that is $X_t - 1$ steps away in the future. In other words, after receiving X_t at t, the receiver certainly gets X_t once more at time step $t + X_t - 1$.

Therefore, the time series is a mixture of randomness and deterministic inputs. For example, in figure (4.10) at time t (between two dashed lines), we see that future states are partially known based on what has been already seen in the past, while other future states are uncertain.

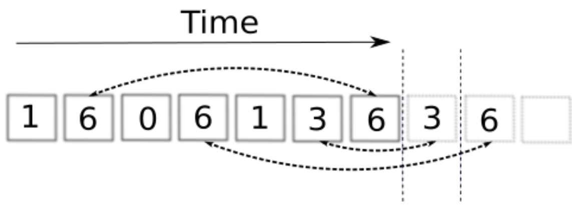


Figure 4.10: The receiver has observed values in the past up to the time-step t (between dashed lines). The state at t and $t+1$ is definite, although farther than that is uncertain.

After this introduction, in what follows, we shall look at a model of pairing time series to show the possibility of mixing certainty and uncertainty in the explained fashion. First, we shall see, using the Shannon entropy, one can define the rate entropy when the length of the string approaches infinity. Later, we show that depending on the probability distribution one imposes on the length of the pairs, the Shannon entropy of each time-step is a well-defined quantity.

4.2.2 Enumerating admissible configurations

To enumerate the set of admissible configurations, we suppose the time series size is finite. So, let us denote by L the maximum length of a time series. *e.g.*, X_L is the last random variable that arrives at the receiver in figure (4.8). This assumption permits us to take a finite alphabet, $\mathcal{A}_L$, to enumerate the state space.

Simultaneously, assuming that $n \leq L$ random variables have arrived at the receiver, Λ_n denotes the state space of all *admissible* configurations with length n. Furthermore, when the joint probability $P(X_1, \ldots, X_n)$ is defined on Λ_n, the Shannon entropy must be equal to

$$H_L(X_1, \ldots, X_n) = - \sum_{(X_1,\ldots,X_n)\in\Lambda_n} P(X_1, \ldots, X_n) \log P(X_1, \ldots, X_n). \qquad (4.89)$$

Note that the state space of admissible configurations is not the Cartesian product of the alphabet set

$$\Lambda_n \neq \underbrace{\mathcal{A} \times \mathcal{A} \times \cdots \times \mathcal{A}}_{n} = \mathcal{A}_L^n. \tag{4.90}$$

Therefore, we have to devise a method to enumerate the admissible configurations properly. To make the formulas compact, we represent a sequence of n random variables as a vector

$$\boldsymbol{X}_n \equiv (X_1, \ldots, X_n). \tag{4.91}$$

For example, the Shannon entropy writes as

$$H_L(\boldsymbol{X}_n) = -\sum_{\boldsymbol{X}_n \in \Lambda_n} P(\boldsymbol{X}_n) \log P(\boldsymbol{X}_n). \tag{4.92}$$

Also, the chain rule for the joint distribution of n random variables writes the joint distribution as the past conditioned on the present

$$P(X_1, X_2, \ldots, X_n) = P(X_n) P(X_1, \ldots, X_{n-1} | X_n), \tag{4.93}$$

or in vector notation

$$P(\boldsymbol{X}_n) = P(X_n) P(\boldsymbol{X}_{n-1} | X_n). \tag{4.94}$$

4.2.2.1 Decomposing the state space

The state-space structure, which is imposed by pairing coins' emergent properties, is not multiplicative. And therefore, it does not easily enumerate two independent sums over states in the present and the past. We shall elaborate on this point later, but first, we decompose the state space to disjoint subsets. To enumerate the admissible configurations, we divide Λ_n into two disjoint subsets

$$\Lambda_n = \Lambda_n^1 \cup \Lambda_n^2, \qquad \Lambda_n^1 \cap \Lambda_n^2 = \emptyset, \tag{4.95}$$

such that:

- Λ_n^1 is a subset of configurations in which a pair exists between X_n and one and only one time-step in the past – see figure (4.11). To emphasise that the pairs are part of this subset, the state of the present random variable is indicated by a left arrow over their values

$$X_n \in \overleftarrow{X_n} = \{\overleftarrow{2}, \ldots, \overleftarrow{n}\}. \tag{4.96}$$

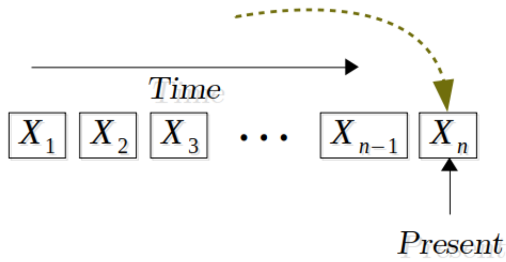


Figure 4.11: X_n is making a pair with one and only one time-step in the past. In other words, the state of X_n has been known in the past.

- Λ_n^2 is a subset of configurations with no pair state between X_n and previous ones – see figure (4.12). Hence, it can be in a head or tail state or make a pair to a single step in the future. Similar to the first case, X_n states are indicated by a right arrow over their values

$$X_n \in \overrightarrow{X_n} = \{0, 1, \overrightarrow{2}, \ldots, \overrightarrow{L+1-n}\}. \tag{4.97}$$

Note that the most distant step from $t = n$ is $t = L + 1 - n$.

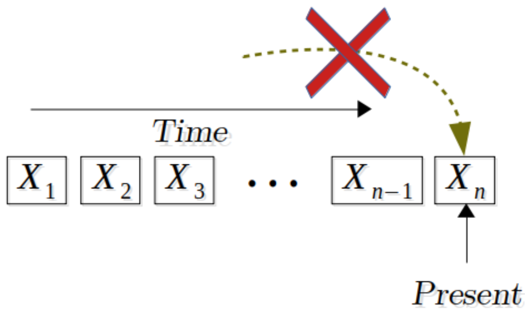


Figure 4.12: X_n can be in a head or tail state or make a pair with one and only one time-step in the future. In other words, the state of X_n contains the information about now or carries the information about the future.

4.2.2.2 Decomposing Λ_n^1

To begin the enumeration, we define a new set, namely Γ_{n-1}^1, that contains all the configurations in Λ_n^1 such that their last random variables at n are excluded – see figure (4.13). In other words, it says elements of Γ_{n-1}^1 are the first $n-1$ states of configurations in Λ_n^1.

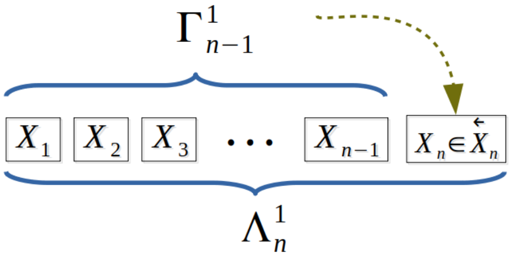


Figure 4.13: Constructing Γ_{n-1}^1 from Λ_n^1.

We understand that the Cartesian product of Γ_{n-1}^1 and $\overleftarrow{X_n}$ introduces non-admissible configurations,

$$\Lambda_n^1 \neq \Gamma_{n-1}^1 \times \overleftarrow{X_n}, \tag{4.98}$$

since every element in Γ_{n-1}^1 makes an admissible configuration with one and only one element in the set $\overleftarrow{X_n}$. The consequence of this observation implies

$$\sum_{\boldsymbol{X}_n \in \Lambda_n} (.) \neq \sum_{\boldsymbol{X}_{n-1} \in \Gamma_{n-1}^1} \sum_{X_n \in \overleftarrow{X_n}} (.)\,, \tag{4.99}$$

and makes the calculation of the Shannon entropy problematic.

To resolve this problem, observe that from the definition of Λ_n^1 every configuration in Γ_{n-1}^1 has a time-step in the past, say i, which makes a pair with X_n, and we can partition Γ_{n-1}^1 to disjoint subsets, denotes by $\Gamma_{n-1}^{1,i}$, such that

$$\Gamma_{n-1}^1 = \bigcup_{i=1}^{n-1} \Gamma_{n-1}^{1,i}, \qquad \Gamma_{n-1}^{1,i} \bigcap \Gamma_{n-1}^{1,j} = \emptyset, \quad i \neq j. \tag{4.100}$$

Note that, when $\boldsymbol{X}_n \in \Lambda_n^1$, the domain of conditional probability $P(\boldsymbol{X}_{n-1}|X_n = i)$ depends on X_n. In short, for $X_n = i$

$$P(\boldsymbol{X}_{n-1}|X_n = i) : \Gamma_{n-1}^{1,i} \to [0, 1], \qquad P(\boldsymbol{X}_n) : \Lambda_n^1 \to [0, 1]. \tag{4.101}$$

So, extending the domain of the conditional probability $P(\boldsymbol{X}_{n-1}|X_n = i)$ from $\Gamma_{n-1}^{1,i}$

to Γ^1_{n-1} is simply equivalent to

$$P_E(\boldsymbol{X}_{n-1}|X_n = i) \equiv \begin{cases} P(\boldsymbol{X}_{n-1}|X_n = i) & , \boldsymbol{X}_{n-1} \in \Gamma^{1,i}_{n-1} \\ 0 & , \text{ otherwise} \end{cases}, \tag{4.102}$$

where

$$P_E(\boldsymbol{X}_{n-1}|X_n = i) : \Gamma^1_{n-1} \to [0,1], \tag{4.103}$$

and it introduces an extended joint probability such that

$$P_E(\boldsymbol{X}_n) = P(X_n) P_E(\boldsymbol{X}_{n-1}|X_n), \qquad P_E(\boldsymbol{X}_n) : \Gamma^1_{n-1} \times \overleftarrow{X_n} \to [0,1]. \tag{4.104}$$

Thus, using the convention $0 \log 0 = 0$, the extension results in

$$\sum_{\boldsymbol{X}_n \in \Lambda^1_n} P(\boldsymbol{X}_n) \log P(\boldsymbol{X}_n)$$
$$= \sum_{X_n \in \overleftarrow{X_n}} P(X_n) \sum_{\boldsymbol{X}_{n-1} \in \Gamma^1_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n) \log \left[P(X_n) P_E(\boldsymbol{X}_{n-1}|X_n)\right]. \tag{4.105}$$

Likewise, including $X_n \in \overrightarrow{X_n}$ in the domain of the extension such that the conditional probability is zero, $P_E(\boldsymbol{X}_n)$ becomes

$$P_E(\boldsymbol{X}_n) : \Gamma^1_{n-1} \times (\overleftarrow{X_n} \cup \overrightarrow{X_n}) \to [0,1]. \tag{4.106}$$

Then we can write

$$\sum_{\boldsymbol{X}_n \in \Lambda^1_n} P(\boldsymbol{X}_n) \log P(\boldsymbol{X}_n)$$
$$= \sum_{X_n \in \overleftarrow{X_n} \cup \overrightarrow{X_n}} P(X_n) \sum_{\boldsymbol{X}_{n-1} \in \Gamma^1_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n) \log \left[P(X_n) P_E(\boldsymbol{X}_{n-1}|X_n)\right]. \tag{4.107}$$

4.2.2.3 Decomposing Λ^2_n

From the definition of Λ^2_n, we know it contains configurations in which there is no pair between X_n and previous time-steps. Thus, similar to how Γ^1_{n-1} was constructed, Γ^2_{n-1} constructs after removing random variable at n from configurations in Λ^2_n – see figure (4.14). Surely,

$$\Lambda^2_n \neq \Gamma^2_{n-1} \times \overrightarrow{X_n}. \tag{4.108}$$

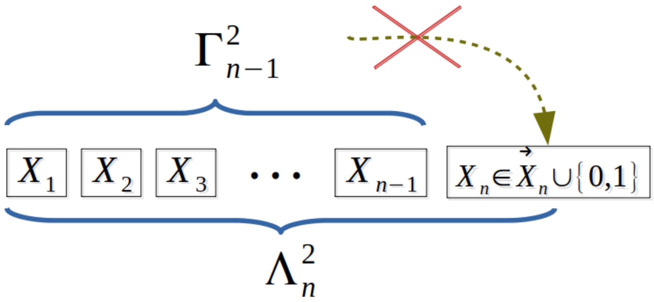


Figure 4.14: Constructing Γ^2_{n-1} from Λ^2_n.

Furthermore, we also partition Γ^2_{n-1} into disjoint subsets, say $\Gamma^{2,i}_{n-1}$. For $X_n = \vec{i}$, observe that the time-step at $n+i-1$ in the future must make a pair state with X_n. Therefore, a configuration that makes a pair from any time-step except n to $n+i-1$ is not admissible. So, Γ^2_{n-1} partitions into $\Gamma^{2,i}_{n-1}$ for all $i \in \{\vec{2}, \ldots, \overrightarrow{L+1-n}\}$.

Afterward, for $\boldsymbol{X}_n \in \Lambda^2_n$, the domain of the conditional probability $P(\boldsymbol{X}_{n-1}|X_n)$ depends on X_n. In other words, for $X_n = i$ and $i \in \{\vec{2}, \ldots, \overrightarrow{L+1-n}\}$

$$P(\boldsymbol{X}_{n-1}|X_n = i) : \Gamma^{2,i}_{n-1} \to [0,1], \tag{4.109}$$

while for $i \in \{0,1\}$

$$P(\boldsymbol{X}_{n-1}|X_n = i) : \{0,1\} \to [0,1]. \tag{4.110}$$

Notice that in the last relation, the domain of the conditional probability is Γ^2_{n-1} since for $X_n = 0$ or 1, there is no pair between the present and the future time steps.

Similar to the case $\boldsymbol{X}_n \in \Lambda^1_n$, we extend the conditional probability $P(\boldsymbol{X}_{n-1}|X_n = i)$ from $\Gamma^{2,i}_{n-1}$ to Γ^2_{n-1} like

$$P_E(\boldsymbol{X}_{n-1}|X_n = i) \equiv \begin{cases} P(\boldsymbol{X}_{n-1}|X_n = i) & , \boldsymbol{X}_{n-1} \in \Gamma^{2,i}_{n-1} \cup \{0,1\} \\ 0 & , \text{ otherwise} \end{cases}, \tag{4.111}$$

and

$$P_E(\boldsymbol{X}_{n-1}|X_n = i) : \Gamma^2_{n-1} \to [0,1]. \tag{4.112}$$

Consequently, the extended joint probability distribution defines as

$$P_E(\boldsymbol{X}_n) \equiv P(X_n)P_E(\boldsymbol{X}_{n-1}|X_n), \qquad P_E(\boldsymbol{X}_n) : \Gamma^2_{n-1} \times \vec{X}_n \to [0,1]. \tag{4.113}$$

Thus

$$\sum_{\boldsymbol{X}_n \in \Lambda_n^2} P(\boldsymbol{X}_n) \log P(\boldsymbol{X}_n)$$

$$= \sum_{X_n \in \overrightarrow{X_n}} P(X_n) \sum_{\boldsymbol{X}_{n-1} \in \Gamma_{n-1}^2} P_E(\boldsymbol{X}_{n-1}|X_n) \log \left[P(X_n) P_E(\boldsymbol{X}_{n-1}|X_n) \right]. \tag{4.114}$$

Furthermore, by defining the joint probability for $X_n \in \overleftarrow{X_n}$ equal to zero, we can write

$$P_E(\boldsymbol{X}_n) : \Gamma_{n-1}^2 \times (\overleftarrow{X_n} \cup \overrightarrow{X_n}) \to [0,1], \tag{4.115}$$

and hence, using the convention $0 \log 0 = 0$, the extension results in

$$\sum_{\boldsymbol{X}_n \in \Lambda_n^2} P(\boldsymbol{X}_n) \log P(\boldsymbol{X}_n)$$

$$= \sum_{X_n \in \overleftarrow{X_n} \cup \overrightarrow{X_n}} P(X_n) \sum_{\boldsymbol{X}_{n-1} \in \Gamma_{n-1}^1} P_E(\boldsymbol{X}_{n-1}|X_n) \log \left[P(X_n) P_E(\boldsymbol{X}_{n-1}|X_n) \right]. \tag{4.116}$$

4.2.2.4 Decomposition theorems

Before proceeding further, we prove the following theorems, and we shall use them to enumerate the state space for calculating the entropy.

Theorem 4.2.1. $\Gamma_{n-1}^1 \cap \Gamma_{n-1}^2 = \emptyset$.

Proof. Let us say $c_{n-1} \in \Gamma_{n-1}^1$. From the definition of Γ_{n-1}^1, the configuration c_{n-1} is constructed from c_n such that $c_n \in \Lambda_n^1$. Since Λ_n^1 and Λ_n^2 are disjoint sets, then $c_n \notin \Lambda_n^2$. Consequently, $c_{n-1} \notin \Gamma_{n-1}^2$. Similarly, $c_{n-1} \in \Gamma_{n-1}^2$ results in $c_{n-1} \notin \Gamma_{n-1}^1$, which implies the claim of the theorem. □

Lemma 4.2.1. *All admissible configurations, say c_{n-1}, is in Λ_{n-1} if and only if there exists at least one admissible configuration $c_n \in \Lambda_n$.*

Proof. We star from the forward case, namely $c_{n-1} \in \Lambda_{n-1}$. The configuration c_{n-1} may or may not have a pair that links to the site n. Let assume it does not have one. So, for example $X_n = 0$ constructs an admissible configuration, and consequently $c_n \in \Lambda_n$. Otherwise, there is a time step in c_{n-1} which makes a pair to X_n, and from the definition of admissible configurations it implies $c_n \in \Lambda_n$.

The converse case is straightforward. For any $c_n \in \Lambda_n$, constructing c_{n-1} by dropping the element's states at n must create an admissible configuration, otherwise c_n cannot be an admissible one. Hence, $c_{n-1} \in \Lambda_{n-1}$. □

Theorem 4.2.2. $\Lambda_{n-1} = \Gamma^1_{n-1} \cup \Gamma^2_{n-1}$.

Proof. Let us assume $c_{n-1} \in \Lambda_{n-1}$. From lemma 4.2.1, there exists one configuration $c_n \in \Lambda_n$. And since c_n is admissible, constructing a configuration c'_{n-1} by dropping its last state at n implies that c'_{n-1} is in Γ^1_{n-1} or Γ^2_{n-1}. This conclusion is justified based on the definition of Γ^1_{n-1} and Γ^2_{n-1}.

Consequently, $c'_{n-1} \in \Gamma^1_{n-1} \cup \Gamma^2_{n-1}$. It only remains to show that $c_{n-1} = c'_{n-1}$. This must be trivially true, since constructing c_n from c_{n-1} was equivalent to add one element at n in the start of the argument, which we dropped it later to make c'_{n-1}. So, all the states in the previous $n-1$ positions are intact and $c_{n-1} = c'_{n-1}$. Therefore, $c_{n-1} \in \Gamma^1_{n-1} \cup \Gamma^2_{n-1}$, which implies $\Lambda_{n-1} \subset \Gamma^1_{n-1} \cup \Gamma^2_{n-1}$.

For the converse case, Let us assume $c_{n-1} \in \Gamma^1_{n-1} \cup \Gamma^2_{n-1}$. From definitions of Γ^1_{n-1} and Γ^2_{n-1}, we can always construct c_n such that it is in Λ_n, therefore, lemma 4.2.1 requires $c_{n-1} \in \Lambda_{n-1}$, which implies $\Gamma^1_{n-1} \cup \Gamma^2_{n-1} \subset \Lambda_{n-1}$. This complete the proof of the claimed statement.

□

4.2.2.5 Shannon entropy for pairing time series

In this part for a finite pairing time series, we show that the Shannon entropy is an additive, recursive relation over an ensemble of strings with length n and writes as

$$H_L(\boldsymbol{X}_n) = H_L(X_n) + H_L(\boldsymbol{X}_{n-1}|X_n), \tag{4.117}$$

such that *the present entropy* is defined as

$$H_L(X_n) \equiv -\sum_{X_n \in \overleftarrow{X_n} \cup \overrightarrow{X_n}} P(X_n) \log P(X_n), \tag{4.118}$$

and *the past conditional entropy*, given present, is

$$H_L(\boldsymbol{X}_{n-1}|X_n) \equiv -\sum_{X_n \in \overleftarrow{X_n} \cup \overrightarrow{X_n}} P(X_n) \sum_{\boldsymbol{X}_{n-1} \in \Lambda_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n) \log P_E(\boldsymbol{X}_{n-1}|X_n). \tag{4.119}$$

Here, we show only the steps that use the results from previous sections. The details of the derivation are explained in appendix (C.8).

Proof:

We write the Shannon entropy over an ensemble of strings with length n as

$$H_L(\boldsymbol{X}_n) = -\sum_{\boldsymbol{X}_n \in \Lambda_n} P(\boldsymbol{X}_n) \log P(\boldsymbol{X}_n)$$

$$= -\sum_{\boldsymbol{X}_n \in \Lambda_n^1} P(\boldsymbol{X}_n) \log P(\boldsymbol{X}_n) - \sum_{\boldsymbol{X}_n \in \Lambda_n^2} P(\boldsymbol{X}_n) \log P(\boldsymbol{X}_n) \qquad (\Lambda_n = \Lambda_n^1 \cup \Lambda_n^2)$$

$$= -\sum_{X_n \in \overleftarrow{X_n}} \sum_{\boldsymbol{X}_{n-1} \in \Gamma_{n-1}^1} P(X_n) P_E(\boldsymbol{X}_{n-1}|X_n) \left[\log P(X_n) + \log P_E(\boldsymbol{X}_{n-1}|X_n)\right]$$

(from 4.105)

$$-\sum_{X_n \in \overrightarrow{X_n}} \sum_{\boldsymbol{X}_{n-1} \in \Gamma_{n-1}^2} P(X_n) P_E(\boldsymbol{X}_{n-1}|X_n) \left[\log P(X_n) + \log P_E(\boldsymbol{X}_{n-1}|X_n)\right]$$

(from 4.114)

$$= -\sum_{X_n \in \overleftarrow{X_n} \cup \overrightarrow{X_n}} P(X_n) \log P(X_n)$$

$$-\sum_{X_n \in \overleftarrow{X_n} \cup \overrightarrow{X_n}} P(X_n) \sum_{\boldsymbol{X}_{n-1} \in \Lambda_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n) \log P_E(\boldsymbol{X}_{n-1}|X_n).$$

$(\Gamma_{n-1}^1 \cup \Gamma_{n-1}^2 = \Lambda_{n-1})$

4.2.3 Entropy for Infinite Alphabet

The entropy $H_L(\boldsymbol{X}_n)$ is defined for a finite alphabet and finite-length time series. However, for infinite time series, or the limit $L \to \infty$, one must prove the existence of entropy and examine that it is well-define. Put it differently, the Shannon entropy of infinite time series at n is defined as

$$H(\boldsymbol{X}_n) \equiv \lim_{L \to \infty} H_L(\boldsymbol{X}_n). \tag{4.120}$$

And therefore, $H_L(\boldsymbol{X}_n)$ must be bounded and converging to have a non-absurd Shannon entropy $H(\boldsymbol{X}_n)$ for infinite time series. To find an upper bound for the limit, we use the inequality that governs the conditional entropy [15]

$$H_L(\boldsymbol{X}_{n-1}|X_n) \leq H_L(\boldsymbol{X}_{n-1}), \tag{4.121}$$

and write

$$\begin{aligned}
H_L(\boldsymbol{X}_n) &= H_L(X_n) + H_L(\boldsymbol{X}_{n-1}|X_n) \\
&\leq H_L(X_n) + H_L(\boldsymbol{X}_{n-1}) && \text{(from (4.121))} \\
&= H_L(X_n) + H_L(X_{n-1}) + H_L(\boldsymbol{X}_{n-2}|X_{n-1}) && \text{(from (4.117))} \\
&= \dots \\
&\leq \sum_{i=1}^{n} H_L(X_i),
\end{aligned} \tag{4.122}$$

where

$$X_i \in \overleftarrow{X_i} \cup \overrightarrow{X_i}. \tag{4.123}$$

Let us assume the marginal entropy $H_L(X_i)$ has an upper bound such that

$$\forall i \leq L: \qquad H_L(X_i) \leq h_c. \tag{4.124}$$

Then

$$H_L(\boldsymbol{X}_n) \leq \sum_{i=1}^{n} H_L(X_i) \leq n h_c. \tag{4.125}$$

The upper bound of $H_L(\boldsymbol{X}_n)$ is n-dependent, and in the limit $L \to \infty$, we have

$$H(\boldsymbol{X}_n) \equiv \lim_{L\to\infty} H_L(\boldsymbol{X}_n) \leq n h_c. \tag{4.126}$$

Consequently, $H(\boldsymbol{X}_n)$ does not diverge and is well-defined. Conversely, the non-negativity of conditional entropy $H_L(\boldsymbol{X}_{n-1}|X_n)$ in equation (4.117) implies

$$H_L(X_n) \leq H_L(\boldsymbol{X}_n), \tag{4.127}$$

and therefore, in the limit $L \to \infty$, if $H_L(X_n)$ is unbounded, we must have

$$\lim_{L\to\infty} H_L(\boldsymbol{X}_n) \to \infty. \tag{4.128}$$

So, for $\exists i \in \mathbb{N}$ and a diverging entropy $H_L(X_i)$, equation (4.127) implies for every $i \leq n$ the entropy of the sequence, namely $H_L(\boldsymbol{X}_n)$, is diverging too. Consequently, the joint entropy $H_L(\boldsymbol{X}_n)$ is not defined.

Finally, these two results enable us to investigate the existence of Shannon entropy by studying the divergence of the entropy at an arbitrary time-step, say $H_L(X_i)$, for the limit $L \to \infty$ as follows:

1. The diverging $H_L(X_i)$ for $\exists i \in \mathbb{N}$ implies the divergence of $\forall i \leq n : H_L(\boldsymbol{X}_n)$. Thus, $H(\boldsymbol{X}_n)$ is undefined.
2. The boundedness of $H_L(X_i)$ for $\forall i \in \mathbb{N}$ implies the convergence of $H_L(\boldsymbol{X}_n)$ for all n. Hence, $H(\boldsymbol{X}_n)$ exists and is well-defined.

In addition, to study the entropy of a single moment, we must find the marginal probability and recall that the marginal probability is defined as

$$P(X_n = x) = \sum_{\boldsymbol{X}_n \in \Lambda_n : X_n = x} P(\boldsymbol{X}_n). \tag{4.129}$$

4.2.3.1 The Rate Entropy

The rate entropy [64] is defined as

$$H(\mathcal{X}) \equiv \lim_{n\to\infty} \frac{H(X_1, \ldots, X_n)}{n} = \lim_{n\to\infty} \frac{H(\boldsymbol{X}_n)}{n}. \tag{4.130}$$

Whenever the Shannon entropy $H(\boldsymbol{X}_n)$ is defined, using equation (4.126), we find

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{H(\boldsymbol{X}_n)}{n} = \lim_{n\to\infty} \lim_{L\to\infty} \frac{H_L(\boldsymbol{X}_n)}{n} \leq h_c. \tag{4.131}$$

And since $H(\mathcal{X})$ is a non-negative function and has an upper bound, it must have a limit. In other words, the rate entropy exists and is well-defined.

For ergodic, stationary stochastic processes, the Shannon-McMillan-Breiman theorem [15] finds that

$$-\frac{1}{n} \ln P(X_0, \ldots, X_{n-1}) \to H(\mathcal{X}), \tag{4.132}$$

with probability one. This theorem is used to prove the Asymptotic Equipartition property (AEP) [15] and is an important result in Information Theory and its application. However, a pairing time series is not stationary, and at the moment, we do not know the validity of AEP for them. Therefore, we leave it as an open question that can be pursued as a future research problem.

4.2.4 Uniform Distribution

Similar to the first case, we start with the uniform distribution for configurations with length L. Knowing the state space volume, say $\Omega_2(L)$, the uniform distribution is equal to

$$P(\boldsymbol{X}_L) = \frac{1}{\Omega_2(L)}. \tag{4.133}$$

From the combinatorial argument about the C-model, we remember that the $\Omega_2(L)$ is composed of $2\Omega_2(L-1)$ configurations whenever we take into account the effect of adding a new head or tail state of a single coin and $(L-1)\Omega_2(L-1)$ for making a pair that coin can make with any other ones. So, accordingly, the marginal sum in equation (4.129) has $\Omega_2(L-1)$ equal terms when $X_n = 0, 1$, and $\Omega_2(L-1)$ otherwise. As a result

$$P(X_n = x) = \begin{cases} \frac{\Omega_2(L-1)}{\Omega_2(L)} & X_n = 0, 1 \\ \frac{\Omega_2(L-2)}{\Omega_2(L)} & X_n \in \{\overleftarrow{2}, \dots, \overleftarrow{n}, \overrightarrow{2}, \dots, \overrightarrow{L+1-n}\}. \end{cases} \tag{4.134}$$

In appendix (C.9), equation (C.59) finds as

$$\frac{\Omega_2(L-1)}{\Omega_2(L)} \sim \frac{1}{\sqrt{L}}, \quad (L-1)\frac{\Omega_2(L-2)}{\Omega_2(L)} \sim 1 - \frac{2}{\sqrt{L}}, \tag{4.135}$$

and, for $1 \ll L$, the Shannon entropy for the uniform distribution in equation (4.134) writes

$$\begin{aligned} H_L(X_n) = &- \sum_{X_n \in \{\overleftarrow{2}, \dots, \overleftarrow{n}, 0, 1, \overrightarrow{2}, \dots, \overrightarrow{L+1-n}\}} P(X_n) \log P(X_n) \\ = &-2\frac{\Omega_2(L-1)}{\Omega_2(L)} \log \frac{\Omega_2(L-1)}{\Omega_2(L)} - (L-1)\frac{\Omega_2(L-2)}{\Omega_2(L)} \log \frac{\Omega_2(L-2)}{\Omega_2(L)} \\ = &-\frac{2}{\sqrt{L}} \log \frac{1}{\sqrt{L}} - (1 - \frac{2}{\sqrt{L}}) \log \frac{1 - \frac{2}{\sqrt{L}}}{L-1} \qquad \text{(from eq. (4.135))} \\ &\sim (1 - \frac{1}{\sqrt{L}}) \log L \sim \log L. \end{aligned} \tag{4.136}$$

We see that, in the limit $L \to \infty$, the above result is unbounded, and consequently, $H_L(\boldsymbol{X}_n)$. Hence, based on the explanation in the previous section, the unboundedness of $H_L(X_n)$ implies $H_L(\boldsymbol{X}_n)$ and the rate entropy is undefined for uniform distributions.

4.2.5 Exponential distribution

Observe that the random variable X_n in equation (4.129) can be partitioned into the following subsets

$$\{\overleftarrow{2}, \ldots, \overleftarrow{n}\} \cup \{0, 1\} \cup \{\overrightarrow{2}, \ldots, \overrightarrow{L+1-n}\}, \tag{4.137}$$

and note that the size of the last set is L-dependent. In this section, we assume the probability of observing $X_n \in \{\overrightarrow{2}, \ldots, \overrightarrow{L+1-n}\}$ is exponentially decreasing with respect to n

$$P(X_n = l) \propto \mathrm{e}^{-\lambda l}, \qquad \lambda > 0. \tag{4.138}$$

To keep the assumption as general as possible, no other condition is assumed for the probability $P(X_n)$ for $X_n \in \{\overleftarrow{2}, \ldots, \overleftarrow{n}, 0, 1\}$. So, we define

$$P(X_n) = \begin{cases} A\mathrm{e}^{-\lambda l} & X_n = l \in \{\overrightarrow{2}, \ldots, \overrightarrow{L+1-n}\} \\ f(l) & X_n = l \in \{\overleftarrow{2}, \ldots, \overleftarrow{n}, 0, 1\}, \end{cases} \tag{4.139}$$

where A is the normalisation constant and $f(l)$ is an arbitrary function

$$f : \{\overleftarrow{2}, \ldots, \overleftarrow{n}, 0, 1\} \to [0, 1]. \tag{4.140}$$

Applying the normalisation condition on $P(X_n)$, we get

$$A = \frac{1 - \sum_{l \in \{\overleftarrow{2}, \ldots, \overleftarrow{n}, 0, 1\}} f(l)}{\sum_{l=2}^{L+1-n} \mathrm{e}^{-\lambda l}}. \tag{4.141}$$

In the limit

$$\lim_{L \to \infty} \sum_{l=2}^{L+1-n} e^{-\lambda l} = \frac{e^{-\lambda}}{e^{\lambda} - 1}, \tag{4.142}$$

therefore, for $L \to \infty$, the normalisation constant is well-defined. Next, we write the entropy of the moment n as

$$H_L(X_n) = -\sum_{X_n \in \{\overleftarrow{2},\ldots,\overleftarrow{n},0,1,\overrightarrow{2},\ldots,\overrightarrow{L+1-n}\}} P(X_n) \log P(X_n)$$

$$= -\sum_{X_n \in \{\overleftarrow{2},\ldots,\overleftarrow{n},0,1\}} f(X_n) \log f(X_n) - A \log A \sum_{l=2}^{L+1-n} e^{-\lambda l} + A\lambda \sum_{l=2}^{L+1-n} l e^{-\lambda l}. \quad (4.143)$$

In the limit, the third term is equal to

$$\lim_{L\to\infty} \sum_{l=2}^{L+1-n} l e^{-\lambda l} = -\frac{1}{(1-e^{-\lambda})^2} - e^{-\lambda}. \quad (4.144)$$

Hence, using equations (4.142) and (4.144)

$$H(X_n) = \lim_{L\to\infty} H_L(X_n)$$

$$= -\sum_{X_n \in \{\overleftarrow{2},\ldots,\overleftarrow{n},0,1\}} f(X_n) \log f(X_n) - \frac{A \log A e^{-\lambda}}{e^{\lambda} - 1} - \frac{A}{(1-e^{-\lambda})^2} - A e^{-\lambda}, \quad (4.145)$$

and it is bounded, which implies $H(\boldsymbol{X}_n)$ is well-defined.

4.2.6 Power law

Similar to the previous section, we study the power law probability distribution on the length of pairs into the future, or for pairing $X_n \in \{\overrightarrow{2}, \ldots, \overrightarrow{L+1-n}\}$

$$P(X_n = l) \propto l^{-\lambda}, \qquad \lambda > 1. \quad (4.146)$$

So the probability distribution is

$$P(X_n) = \begin{cases} A l^{-\lambda} & X_n = l \in \{\overrightarrow{2}, \ldots, \overrightarrow{L+1-n}\} \\ f(l) & X_n = l \in \{\overleftarrow{2}, \ldots, \overleftarrow{n}, 0, 1\}, \end{cases} \quad (4.147)$$

where A is the normalisation constant, and $f(l)$ is an arbitrary function

$$f : \{\overleftarrow{2}, \ldots, \overleftarrow{n}, 0, 1\} \to [0, 1]. \quad (4.148)$$

So, the normalisation constant is

$$A = \frac{1 - \sum_{l \in \{\overleftarrow{2}, \ldots, \overleftarrow{n}, 0, 1\}} f(l)}{\sum_{l=2}^{L+1-n} l^{-\lambda}}, \tag{4.149}$$

and, in the limit $L \to \infty$, we get

$$\lim_{L \to \infty} \sum_{l=2}^{L+1-n} l^{-\lambda} = \zeta(\lambda) - 1, \tag{4.150}$$

where $\zeta(\lambda)$ is the Riemann's zeta function. Therefore, for $L \to \infty$

$$\lim_{L \to \infty} A = \frac{1 - \sum_{l \in \{\overleftarrow{2}, \ldots, \overleftarrow{n}, 0, 1\}} f(l)}{\zeta(\lambda) - 1}, \tag{4.151}$$

is well defined. Then, using equation (4.150), the entropy obtains as

$$H_L(X_n) = - \sum_{X_n \in \{\overleftarrow{2}, \ldots, \overleftarrow{n}, 0, 1, \overrightarrow{2}, \ldots, \overrightarrow{L+1-n}\}} P(X_n) \log P(X_n)$$

$$= - \sum_{X_n \in \{\overleftarrow{2}, \ldots, \overleftarrow{n}, 0, 1\}} f(X_n) \log f(X_n) - A \log A \sum_{l=2}^{L+1-n} l^{-\lambda} + A\lambda \sum_{l=2}^{L+1-n} l^{-\lambda} \log l. \tag{4.152}$$

As we have already shown for the normalisation constant, the second sum on the right-hand side has a limit

$$\lim_{L \to \infty} \sum_{l=2}^{L+1-n} l^{-\lambda} = \zeta(\lambda) - 1. \tag{4.153}$$

It remains to check the convergence of the last term to conclude that $H_L(X_n)$ is bounded for all ns. In the limit $L \to \infty$, we have

$$\lim_{L \to \infty} \sum_{i=2}^{L+1-n} i^{-\lambda} \log i = \sum_{i=1}^{\infty} \frac{\log(i+1)}{(i+1)^{\lambda}}, \qquad \lambda > 1. \tag{4.154}$$

And to check the convergence of the above series, we use the integral test by taking the summand as a continuous function in $(1, \infty)$

$$\int_1^{\infty} \frac{\log(x+1)}{(x+1)^{\lambda}} dx, \tag{4.155}$$

and check the convergence of the integral instead. Using the integration by parts technique

$$\int_1^\infty \frac{\log(x+1)}{(x+1)^\lambda}dx = \frac{1}{(1-\lambda)}\frac{\log(x+1)}{(x+1)^{\lambda-1}}\Bigg|_1^\infty - \frac{1}{(1-\lambda)}\int_1^\infty \frac{1}{(x+1)^\lambda}dx. \tag{4.156}$$

For $\lambda > 1$, the term $\frac{\log(x+1)}{(x+1)^{\lambda-1}}$ is zero in ∞. Therefore

$$\begin{aligned}\int_1^\infty \frac{\log(x+1)}{(x+1)^\lambda}dx &= -\frac{\log 2}{2^{\lambda-1}(1-\lambda)} - \frac{1}{(1-\lambda)}\int_1^\infty \frac{1}{(x+1)^\lambda}dx \\ &= -\frac{\log 2}{2^{\lambda-1}(1-\lambda)} - \frac{1}{(1-\lambda)^2}\frac{1}{(x+1)^{\lambda-1}}\Bigg|_1^\infty \\ &= \frac{1}{2^{\lambda-1}(1-\lambda)}\left[\frac{1}{(1-\lambda)} - \log 2\right]. \end{aligned} \tag{4.157}$$

So, since the integral is bounded, the series is bounded too, and it concludes that the $H_L(X_n)$ is bounded for $\lambda > 1$.

4.2.7 Conclusion

The previous section showed that for an ensemble of finite pairing strings, the Shannon entropy is additive and is equal to the sum of the present entropy plus the past conditional entropy, given the present.

This result is not surprising considering that the Shannon entropy is the unique function that satisfies a set of axioms, including the additivity [21, 36]. In contrast, for infinite time series, we found that the Shannon entropy is well-defined as long as the distribution over the length of pairs is an exponential distribution or a power law with an exponent greater than one.

The Asymptotic Equipartition Property theorem [15] separates the ensemble of ergodic, stationary stochastic sequences into a set of typical and atypical sets, such that the probability of observing an atypical sequence is negligible. We do not know a similar characteristic is held by pairing time series, but proving the existence of a well-defined Shannon entropy is the first step in that direction that we did here.

At the same time, for uniform distribution and power law with an exponent equal to or lower than one, the Shannon entropy is not defined. Reminding that our proof was based on the additivity of Shannon entropy, we can ask ourselves, is it possible to use a non-extensive entropy and get a well-defined state entropy when the Shannon entropy fails for infinite time series?

To elaborate, let us use the Tsallis entropy [67, 69]. For a composed system, Tsallis entropy writes as

$$H_q(A \cup B) = H_q(A) + H_q(B) + (1-q)H_q(A)H_q(B), \tag{4.158}$$

for a positive, real number $q \neq 1$. We see that, for $q > 1$, the above equality finds

$$H_q(A \cup B) < H_q(A) + H_q(B), \tag{4.159}$$

and incorporating Tsallis entropy in equations (4.121) and (4.122), we expect to get a sharper upper bound depending on the choice of q. So, we conjecture that in the cases that the Shannon entropy diverges, a sharper upper bound finds a well-defined entropy for infinite time series when Tsallis entropy is used. It is particularly interesting if the choice of q can be written in terms of the parameter of the distribution, *e.g.* the exponent of the power-law distribution [53].

Similarly, Rényi's entropy [56] satisfies an inequality for two different parameters as

$$\alpha_1 \leq \alpha_2 \implies H_{\alpha_1}[P] \geq H_{\alpha_1}[P]. \tag{4.160}$$

So, the same property can be exploited to find a proper upper bound.

Chapter

FIVE

Pairing Models: Applications

This chapter uses the pairing models in two different contexts: (1) Statistical mechanics and (2) Delivery joint venture, described in the introduction. Note that, as was mentioned in chapter (3), we shall use N to denote the number of elements or system size. This notation is common in physics publications.

Statistical mechanics models are probabilistic ones that incorporate energy function, known as Hamiltonian, for different microscopic states of a system. Similarly, we will introduce energy levels for stand-alone and pair states to calculate standard macroscopic quantities. Our toy models have discrete phase space (state space), and therefore, the Hamiltonian is defined for discrete energy levels. Hence, any difference in the resulting quantities from their corresponding ones in an ordinary model must be due to the emerging states.

As it is common practice in statistical mechanics modelling, we will review the microcanonical and canonical models. For the former ensemble, the system is closed, and the total energy and number of elements are conserved, while in the latter one, the number of constituent elements is conserved. Also, the system in the canonical ensemble is in contact with a heat bath at a constant temperature, so it exchanges energy with the heat bath.

Unlike standard statistical mechanics, we shall see that the specific free energy is diverging and ill-defined in the thermodynamic limit. At this stage, before seeing the result, the reason is apparent. Free energy is the logarithm of the normalisation constant of the Boltzmann distribution, such that its degeneracy of states is equal to the pairing model. Therefore, it scales as $N \ln N$.

However, specific free energy is scaled by $1/N$, and in the limit, it diverges. Therefore, the standard statistical mechanics does not apply to the pairing model or emergence state models.

As we mentioned in the introduction, section (1.2.2), we can find the cost of the delivery joint venture analytically. We will touch on this problem briefly in the last section.

5.1 Standard Statistical Mechanics: Problem

One of the extensively studied models in statistical mechanics is the Ising model [49]. Besides its intuitive simplicity, it has an analytical solution that makes it a suitable model to inspire and outreach other branches of science.

For the B-model, stand-alone elements have a binary state, say, head or tail state. So, naturally, one can construct an Ising model with stand-alone elements while actively introducing emergent states among the pair states. Energy levels are the only missing ingredient to turning the pairing models into statistical mechanics models. In doing so, in this section, we try to construct models similar to standard statistical mechanics ones, and from the onset, we know that these models have faster than exponential growth phase spaces. Therefore, it is insightful to see how the standard statistical model can capture their properties or fails and results in unbounded quantities.

Based on this roadmap, we introduce a system with different energy levels and then use *micro-canonical* and *canonical* ensemble settings to find the subsequent thermodynamic quantities. We will see that these systems have unboundedness in free energy and entropy in the thermodynamic limit.

5.1.1 Introducing Energy Levels

Suppose a system of N pairing coins and three distinct energy levels, ϵ_1, ϵ_2 and ϵ_3, such that, n_1, n_2 and n_3 coins are in each level respectively. Obviously

$$N = n_1 + n_2 + n_3. \tag{5.1}$$

Let us say ϵ_2 is the energy level of coins in pair state. Therefore, there must be even number of coins in ϵ_2 such that, $n_2 = 2k$ for $k \in \mathbb{N} \cup \{0\}$. Then, for any given (n_1, k, n_3) the phase space volume partitions as

$$\Omega_2(n_1, k, n_3) = \frac{N!(2k-1)!!}{(2k)!n_1!n_3!} = \frac{N!}{2^k k! n_1! n_3!}. \tag{5.2}$$

Figure (5.1) shows the combinatorial argument schematically. Without loss of generality, to simplify the calculations, let us suppose

$$\epsilon_1 = -1, \quad \epsilon_2 = 0, \quad \epsilon_3 = 1, \tag{5.3}$$

while the total energy of the system, say E, is obtained as

$$E = n_1\epsilon_1 + n_2\epsilon_2 + n_3\epsilon_3 = n_3 - n_1. \tag{5.4}$$

Therefore in terms of (E, N, k), the volume of the partition of the phase space is equal to

$$\Omega_2(E, N, k) = \frac{N!}{2^k k!(\frac{N-E}{2} - k)!(\frac{N+E}{2} - k)!}. \tag{5.5}$$

We shall use these energy levels and their corresponding phase space partitions for both micro-canonical and canonical ensemble settings in the following sections.

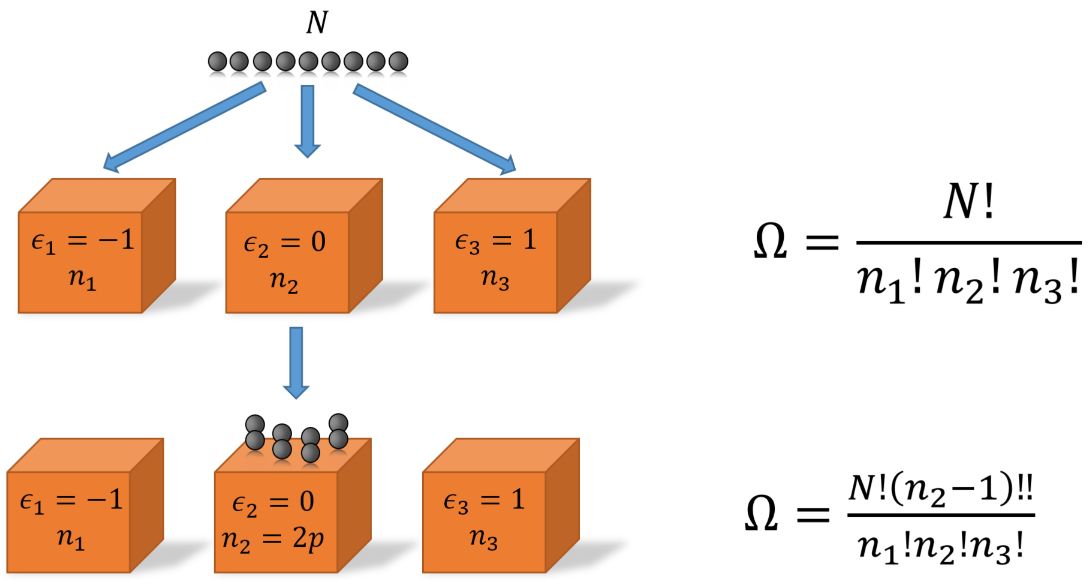


Figure 5.1: The combinatorial problem of distributing paring coins in three energy levels.

5.1.2 Micro-canonical Ensemble

For the phase space volume in equation (5.5), when the specific energy, namely $u = \frac{E}{N}$, is kept constant, we can show $\Omega_2(E, N, k)$ is asymptotically maximum at k^* for $1 \ll N, E$

$$k^* = \begin{cases} \frac{N}{2}(1-u), & u \geq 0 \\ \\ \frac{N}{2}(1+u), & u < 0 \end{cases} = \begin{cases} \frac{N-E}{2}, & E \geq 0 \\ \\ \frac{N+E}{2}, & E < 0 \end{cases}, \tag{5.6}$$

where $-N \leq E \leq N$. Check appendix (D.1) for details. Defining

$$E_* = E_+ = -\left|E_-\right|, \tag{5.7}$$

the phase space volume is maximum at $k^* = \frac{N-E_*}{2}$, and therefore,

$$\Omega_2(E_*, N) = \frac{N!}{2^{\frac{N-E_*}{2}}(\frac{N-E_*}{2})!E_*!}. \tag{5.8}$$

Meanwhile, the phase space volume for fixed (E, N) is

$$\Omega_2(E, N) = \sum_{k=0}^{\lfloor N/2 \rfloor} \Omega_2(E, N, k). \tag{5.9}$$

However, in the above sum, almost all the mass of $\Omega_2(E, N)$ concentrates around its maximum at k^*, with a deviation equal to $\sqrt{N}$. So, as is shown in appendix (D.1), we can safely write

$$\Omega_2(E, N, k^*) \approx \sum_{k=0}^{\lfloor N/2 \rfloor} \Omega_2(E, N, k) = \Omega_2(E, N). \tag{5.10}$$

For an isolated system in equilibrium, the micro-canonical entropy is defined as

$$S_B(N, E) = k_B \ln \Omega(N, E), \tag{5.11}$$

where the total energy, E, and the number of elements, N, are conserved and k_B is the Boltzmann constant [64]. Therefore in the thermodynamic limit, the system's *specific entropy* is

$$s_B \equiv \lim_{N\to\infty} \frac{S_B}{N} = k_B \lim_{N\to\infty} \frac{\log \Omega_2(E, N)}{N}. \tag{5.12}$$

Defining the *specific internal energy*

$$u \equiv \lim_{N\to\infty} \frac{E}{N}, \qquad -1 \leq u \leq 1, \tag{5.13}$$

and using the Stirling's approximation for $N!$, we get

$$\lim_{N\to\infty} \frac{\log \Omega_2(E,N)}{N} = \frac{1}{2}(1-u) \lim_{N\to\infty} (\log N) - u \log u - \frac{1}{2}(1-u)\log(1-u)$$
$$- \frac{1}{2}(1-u) \to \infty, \tag{5.14}$$

which is diverging, and therefore, s_B is unbounded. Check appendix (D.2) for details. The divergence of s_B is equivalent to say S_B is non-extensive. In other words, for the micro-canonical ensemble of paring coins, by using the standard statistical mechanics and the Boltzmann definition of entropy, we get a non-extensive entropy. This is the direct result of a faster than exponential growth phase space of the paring mechanism that manifests itself as $\lim_{N\to\infty} (\log N)$ in the Boltzmann entropy.

5.1.3 Canonical Ensemble

Suppose the previous model is not in isolation, and a heat bath at temperature T is in thermal contact with the system. Also, we will define the Hamiltonian of a one-dimensional Ising model with no neighbour interaction for a configuration $\boldsymbol{\sigma_i}$. As is depicted in figure (5.2), for a system size n, $\boldsymbol{\sigma_i}$ is a sequence of paring coins

$$\boldsymbol{\sigma_i} \equiv (\sigma_{i1}, \ldots, \sigma_{ij}, \ldots, \sigma_{in}), \tag{5.15}$$

such that σ_{ij} is the state of the element at index j in the configuration i. As usual, $\sigma_{ij} = -1$ and $\sigma_{ij} = 1$ corresponds to tail and head states respectively.

We assume the energy contribution of a single element in the head or tail state is caused by its interaction with an external magnetic field, namely B. And to make the notation less cluttered, taking $B = 1$ in an arbitrary unit system. Therefore for tail state or $\sigma_{ij} = -1$, we must have

$$\epsilon_{ij} = B\sigma_{ij} = -1, \tag{5.16}$$

and for $\sigma_{ij} = 1$ or the head state

$$\epsilon_{ij} = B\sigma_{ij} = 1. \tag{5.17}$$

In addition, similar to the micro-canonical model in the previous section, the pair coins are at a zero-energy level, which is equivalent to saying they do not interact with the magnetic field or, simply put, their energy contribution is zero. Consequently, there are three different energy levels

$$\epsilon_{ij} \in \{-1, 0, 1\}, \tag{5.18}$$

and the Hamiltonian of the configuration $\boldsymbol{\sigma_i}$ must be

$$\mathcal{H}(\boldsymbol{\sigma_i}) = -\sum_{j=1}^{N} \epsilon_{ij}. \tag{5.19}$$

Surely, the energy contribution of the pair coins is zero in the Hamiltonian. In other words, removing them from the sum does not change its value.

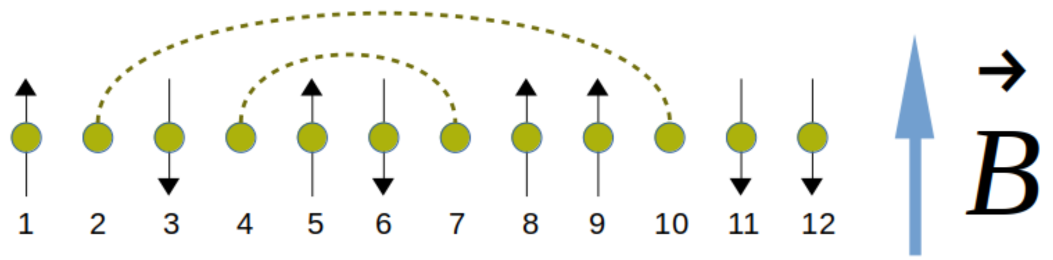


Figure 5.2: One dimensional Ising model including pair states. The up and down states interact with an external magnetic field.

In canonical ensemble, finding thermodynamic quantities corresponds to calculating the partition function. To find it for the one-dimensional paring model, suppose M coins among N are in head or tail states. Thus, the partition function over all the canonical ensemble, say $\{\boldsymbol{\sigma_i}\}$, is defined as

$$Z_M = \sum_{\{\boldsymbol{\sigma_i}\}} \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{\sigma_i})} = \sum_{\{\boldsymbol{\sigma_i}\}} \mathrm{e}^{-\beta \sum_{j=1}^{N} \epsilon_{ij}}$$

$$= \sum_{\{\boldsymbol{\sigma_i}\}} \mathrm{e}^{-\beta \sum_{j=1}^{M} \epsilon_{ij}}, \tag{5.20}$$

where $\beta = \frac{1}{k_B T}$. In appendix (D.3), equation (D.33) for $B = 1$ finds

$$Z_M = 2^M \cosh^M(\beta). \tag{5.21}$$

Notice again that the pair coins do not contribute to Hamiltonian, and we removed their energy contributions from the Hamiltonian. Therefore, Z_M is not the partition function of the paring model, say $Z(N, \beta)$. To find $Z(N, \beta)$, we need to include the effect of pairing in the phase space.

We can see, for $2k$ pair coins there are $\begin{pmatrix} N \\ 2k \end{pmatrix}$ ways to construct distinct configurations. At the same time, there are $(2k-1)!!$ *distinguishable* pairs among $2k$ pair states. By enumerating Z_M, or Z_{N-2k}, over all possible pairs, the partition function derives as

$$Z(N, \beta) = \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} (2k-1)!! \begin{pmatrix} N \\ 2k \end{pmatrix} Z_{N-2k}$$

$$= \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} 2^{N-2k} (2k-1)!! \begin{pmatrix} N \\ 2k \end{pmatrix} \cosh^{N-2k}(\beta). \tag{5.22}$$

Let us define the summand in equation (5.22) as

$$t_{N,k} = 2^{N-2k} (2k-1)!! \begin{pmatrix} N \\ 2k \end{pmatrix} \cosh^{N-2k}(\beta). \tag{5.23}$$

So, the free energy per coin (*specific free energy*) is

$$f(\beta) = \lim_{N \to \infty} \frac{1}{N} \ln \left(\sum_{k=0}^{\frac{N}{2}} t_{N,k} \right), \tag{5.24}$$

for a system in contact with heat bath with constant temperature, and after taking the thermodynamic limit. From the fact that all the terms in equation (5.22) are positive, each summand, namely $t_{N,k}$, is smaller than the sum, or

$$t_{N,k} \leq \sum_{k=0}^{\frac{N}{2}} 2^{N-2k}(2k-1)!! \begin{pmatrix} N \\ 2k \end{pmatrix} \cosh^{N-2k}(\beta B). \tag{5.25}$$

Hence, if we show that in the limit $N \to \infty$ the term $\ln t_{N,k}/N$ is unbounded for a single k, then $f(\beta)$, or specific free energy, is necessarily unbounded. Furthermore, since $0 < k < N/2$, the limit $\lim_{k,N\to\infty} k/N = \epsilon$ for $\epsilon > 0$ exists, and in appendix (D.4) we show that

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \ln t_{N,2k} &= (1-3\epsilon)\ln 2 - \epsilon - \ln(1-2\epsilon) + 2\epsilon \ln(1-2\epsilon) \\ &+ (1-2\epsilon)\ln(\cosh(\beta B)) - \epsilon \ln \epsilon + \lim_{N\to\infty} \epsilon \ln N \to \infty. \end{aligned} \tag{5.26}$$

Consequently, the free energy per element is unbounded too

$$f(\beta) = \lim_{N\to\infty} \frac{1}{N} \log \left(\sum_{k=0}^{\frac{N}{2}} t_{N,k} \right) \to \infty. \tag{5.27}$$

Unboundedness of specific free energy means free energy is non-extensive. It is worth noting that the average energy and the heat capacity are well-defined

$$\langle u \rangle = \frac{\langle U \rangle}{N} = -B \tanh(\beta B) \left(1 - \frac{\langle k \rangle}{N} \right), \tag{5.28}$$

and

$$c_B = \frac{C_B}{N} = k_B \beta^2 B^2 \left[\operatorname{sech}^2(\beta B)(1 - \frac{\langle k \rangle}{N}) + \tanh^2(\beta B) \frac{\langle k^2 \rangle - \langle k \rangle^2}{N} \right], \tag{5.29}$$

and in the thermodynamic limit we find

$$\langle u \rangle = 0, \qquad c_B = 0. \tag{5.30}$$

(the expectations are ensemble averages). We have shown this final result, both numerically and analytically. But details are not included here.

Also, we have to emphasise that the above result is not a pathological case specific to the energy levels, which we defined at the start of this section. To check this claim further, we can use another Hamiltonian with different energy levels

$$\sigma_1 = -1, \quad \sigma_2 = 2L, \quad \sigma_3 = 1, \tag{5.31}$$

where $L \geq 1$, i.e., the new energy level for paired states is higher than the other two. Interestingly, we get

$$\langle u \rangle = L. \tag{5.32}$$

Since $\langle u \rangle$ is the average energy per particle, for any pair of two elements, it becomes $2L$, as we expected, while the free energy is non-extensive.

5.2 Standard Statistical Mechanics: Remedy

In the previous section, we saw the problem of diverging specific free energy in the thermodynamic limit. This section will propose a remedy by using the conjugate functions of pairing model state space, introduced in section (3.9)

$$\phi(x) = \mathrm{e}^{L(x)}, \qquad \phi^{-1}(x) = L^{-1}(\ln x). \tag{5.33}$$

Before starting, let us review the standard statistical mechanics procedure that finds thermodynamic specific free energy from the partition function.

5.2.1 Standard Procedure

In canonical ensemble, one assumes the system is in thermal equilibrium with a heat bath in temperature T and exchanges energy with no particle exchange. In this ensemble, the Boltzmann distribution finds the probability of a microstate – configuration –, say c, with its Hamiltonian $\mathcal{H}(c)$ and inverse temperature $\beta = 1/T$ as[1]

$$P(c) = \frac{\mathrm{e}^{-\beta\mathcal{H}(c)}}{Z_N}, \tag{5.34}$$

where Z_n is the normalisation constant or the partition function, and is defined as

$$Z_N = \sum_{i=1}^{\Omega(N)} \mathrm{e}^{-\beta\mathcal{H}(c_i)}. \tag{5.35}$$

[1] We assumed the Boltzmann constant $k = 1$ in any relevant unit system.

Here again, $\Omega(N)$ denotes the volume of the state space or the number of distinct configurations. In general, one can factorise the degeneracy of configurations that have the same Hamiltonian value. For example, factorising the partition function for l as a parameter that specifies the energy levels of microstates and $W_N(l)$ as the degeneracy of the number of microstates at energy level $\mathcal{H}(l)$ writes as

$$Z_n = \sum_l W_N(l) \mathrm{e}^{-\beta \mathcal{H}(l)}. \tag{5.36}$$

For $1 \ll N$, the summand in partition function concentrates around its peak value, and the sum can be approximated with its maximum, such that for $n \to \infty$, the peak turns into Dirac delta and the dispersion around the maximum decreases as $1/\sqrt{N}$. This fact is explained in detail, *e.g.*, in [27, 54, 58, 63, 64], and more systematic techniques like *steepest decent* can be employed to estimate Z_N. However, we take this assumption for granted and for the maximum of the Boltzmann distribution at l^*, write the estimate as

$$Z_N = \sum_l W_N(l) \mathrm{e}^{-\beta \mathcal{H}(l)} \approx \mathrm{e}^{[\ln W_N(l^*) - \beta \mathcal{H}(l^*)]}. \tag{5.37}$$

At the same time, since the free energy is defined as [49]

$$F_N = -T \ln Z_N, \tag{5.38}$$

one writes the equation (5.36) as

$$F_N = -T \ln \left(\mathrm{e}^{\ln W_N(l^*) - \beta \mathcal{H}(l^*)} \right) = \mathcal{H}(l^*) - T \ln W_N(l^*) \implies$$

$$F_N = U_N - T S_N, \tag{5.39}$$

where $U_N \equiv \mathcal{H}(l^*)$ is the internal energy and $S_N \equiv \ln W_N(l^*)$ is the entropy. Notice that the internal energy equals the energy level at the maximum of the Boltzmann distribution. This assumption is reasonable since the system is in thermal equilibrium with a heat bath, and the energy fluctuation is negligible compared to its value.

It is essential to recognise that a microscopic model in statistical mechanics obtains the last result, so, to obtain the equivalent thermodynamic quantities, one needs to take the thermodynamic limit. Therefore, to derive the well-known thermodynamic identity between specific free energy, entropy, and internal energy in the thermodynamic limit [23, 27, 28], we must write

$$f \equiv \lim_{N\to\infty} \frac{1}{N} F_N = \lim_{N\to\infty} \frac{\ln \mathcal{H}(l^*)}{N} - \frac{T \ln W_N(l^*)}{N} \implies$$

$$f = u - Ts, \tag{5.40}$$

for $f \equiv F_N/N$ as specific free energy, $s \equiv S_N/N = \ln W_N(l^*)/n$ as specific thermodynamic entropy and $u \equiv U_N/N = \mathcal{H}(l^*)/N$ as specific internal energy.

Following this procedure, the program of statistical mechanics reduces to finding the partition function, and from there, one can find thermodynamics quantities. However, section one of this chapter showed that this procedure fails for simple Hamiltonian and its energy levels, and consequently, the free energy diverges. We have to remark that the results obtained for large deviation limits in chapter (3) and appendix (A.3) proved that the probability distributions in pairing state spaces concentrate around their peak. Therefore, the shape of the distribution is not the source of the problem, while the estimation by maximum is valid for the pairing space partition function as well.

5.2.2 A Solution

One solution that we propose in this thesis constitutes defining the *specific values* of quantities of pairing models with transformation function in equation (3.235) that was introduced in section (3.9). When the statistical mechanics' free energy is defined as

$$F_N = \ln(Z_N), \tag{5.41}$$

then, we define its specific value as

$$f \equiv -T \lim_{N\to\infty} \frac{1}{N} \phi(F_N) = -T \lim_{N\to\infty} \frac{1}{N} \mathrm{e}^{L(\ln(Z_N))}, \tag{5.42}$$

and for the entropy

$$S_N = \ln(W_N(m_N)), \tag{5.43}$$

the specific entropy is

$$s \equiv \lim_{N\to\infty} \frac{1}{N} \phi(S_N) = \lim_{N\to\infty} \frac{1}{N} \phi(\ln(W_N(m_N))). \tag{5.44}$$

Let us write the canonical ensemble partition function that we discussed in section (5.1.3), but for a general Hamiltonian function of n_p, and estimate its sum by its maximum value at n_p^*

$$Z_N = \sum_{n_p}^{\lfloor N/2 \rfloor} W_N(n_p) \mathrm{e}^{-\beta \mathcal{H}(n_p)} \approx \mathrm{e}^{\ln W_N(n_p^*) - \beta \mathcal{H}(n_p^*)}. \tag{5.45}$$

Hence, using equation (5.42), the conjugate function derives the specific free energy as

$$f = -T \lim_{N \to \infty} \mathrm{e}^{L(\ln W_N(n_p^*) - \beta \mathcal{H}(n_p^*))}. \tag{5.46}$$

The specific free energy limit has three different regimes:

1. For asymptotically extensive Hamiltonians, for $1 \ll N$ and a function $h(.)$ that is bounded above, we must have

$$\mathcal{H}(m_N^*) \sim h(m_N^*) N. \tag{5.47}$$

Using equation (3.241) for the degeneracy $W_N(m_N^*)$, the entropy writes as

$$S_N = \ln(W_N(m_N^*)) = \frac{N m_N^*}{2} \ln \frac{N m_N^*}{2} + O(N), \tag{5.48}$$

and the specific free energy limit obtains as

$$f = -T \lim_{N \to \infty} \frac{1}{N} \mathrm{e}^{L(\frac{N m_N^*}{2} \ln \frac{N m_N^*}{2} - \beta N h(m_N^*) + O(N))}$$

$$= -T \lim_{N \to \infty} \frac{1}{N} \mathrm{e}^{L(\frac{N m_N^*}{2} \ln \frac{N m_N^*}{2})} = -T \lim_{N \to \infty} \frac{\frac{N m_N^*}{2}}{N} \implies$$

$$f = -T \frac{m_N^*}{2}. \tag{5.49}$$

Note that the Hamiltonian is order $O(N)$, and the first term is dominant in the limit. Subsequently, the specific free energy is well defined. However, the energetic interaction of the system does not involve the system's dynamic. To be more precise, the minimum of the free energy, which is its thermodynamic equilibrium, is solely determined by the entropy term. Most importantly, the entropy term derives from the maximum of the degeneracy term, and from equation (5.48), the specific entropy finds as

$$s = \lim_{N\to\infty} \frac{1}{N}\phi(S_N) = \lim_{N\to\infty} \frac{1}{N} \mathrm{e}^{L(\ln(W_N(m_N^*)))} = \frac{m_N^*}{2}. \tag{5.50}$$

and therefore, the free energy is written as

$$f = -Ts. \tag{5.51}$$

2. The second possibility happens for the Hamiltonian that scales as $N \ln N$

$$\mathcal{H}(m_N^*) \sim h(m_N^*) N \ln N. \tag{5.52}$$

In this case, the Hamiltonian leading term's order is the same as $\ln(W_N(m_N^*))$. Therefore, we use equation (3.241), which derives $\ln(W_N(m_N^*))$ as

$$\begin{aligned}\ln(W_N(m_N^*)) &= \frac{m_N^*}{2} N \ln N - \frac{N}{2}\left[m_N^* \ln \frac{m_N^*}{p} + 2(1-m_N^*)\ln\frac{1-m_N^*}{s} + m_N^*\right] \\ &= \frac{m_N^*}{2} N \ln N + O(N),\end{aligned} \tag{5.53}$$

and then,

$$\begin{aligned}\ln W_N(m_N^*) - \beta\mathcal{H}(m_N^*) &= \frac{m_N^*}{2} N \ln N - \beta h(m_N^*) N \ln N + O(N) \\ = \frac{N}{2}\left[m_N^* - 2\beta h(m_N^*)\right]\ln N &+ \frac{N}{2}\left[m_N^* - 2\beta h(m_N^*)\right]\ln\left[\frac{m_N^* - 2\beta h(m_N^*)}{2}\right] \\ -\frac{N}{2}\left[m_N^* - 2\beta h(m_N^*)\right]&\ln\left[\frac{m_N^* - 2\beta h(m_N^*)}{2}\right] + O(N) \\ = \left[\frac{N}{2}(m_N^* - 2\beta h(m_N^*))\right]&\ln\left[\frac{N}{2}(m_N^* - 2\beta h(m_N^*))\right] + O(N).\end{aligned} \tag{5.54}$$

Finally, the specific free energy limit yields

$$\begin{aligned} f &= -T \lim_{N\to\infty} \frac{1}{N} \mathrm{e}^{L\left(\left[\frac{N}{2}(m_N^* - 2\beta h(m_N^*))\right]\ln\left[\frac{N}{2}(m_N^* - 2\beta h(m_N^*))\right] + O(N)\right)} \\ &= -T \lim_{N\to\infty} \frac{1}{N} \mathrm{e}^{\ln\left[\frac{N}{2}(m_N^* - 2\beta h(m_N^*))\right]} \implies \\ f &= h(m_N^*) - T\frac{m_N^*}{2}.\end{aligned} \tag{5.55}$$

The second term in the last equation is the specific entropy. The specific internal energy defines in a similar way

$$u \equiv \lim_{N\to\infty} \frac{1}{N}\phi(U_N) = \lim_{N\to\infty} \frac{1}{N}\phi(\mathcal{H}(m_N^*)) = \lim_{N\to\infty} \frac{1}{N}\mathrm{e}^{L(h(m_N^*)N\ln N + O(N))}$$

$$= \lim_{N\to\infty} \frac{1}{N}\mathrm{e}^{L(h(m_N^*)N\ln N + h(m_N^*)N\ln h(m_N^*) - h(m_N^*)N\ln h(m_N^*) + O(N))}$$

$$= \lim_{N\to\infty} \frac{1}{N}\mathrm{e}^{L(Nh(m_N^*)\ln Nh(m_N^*) + O(N))} = h(m_N^*), \tag{5.56}$$

and therefore, the free energy writes as

$$f = u - Ts. \tag{5.57}$$

Subsequently, energetic and entropic effects simultaneously evolve the system's dynamics, or say, both determine the minimum free energy.

3. Naturally, the third case happens when the Hamiltonian is the dominant one. Then, depending on the order of Hamiltonian, one has to use a different function to get a converging free and internal energy. So, the minimum of the Hamiltonian always determines the minimum free energy. Accordingly, the system freezes in the corresponding configurations, and thermodynamical evolution does not happen.

To interpret these results, we recall the definition of free energy as the maximum amount of work that a thermodynamic system can extract from the input heat at a constant temperature. So, the quantity Ts acts as an entropic sink while u is the internal energy storage. In a thermodynamic process, part of the input energy – heat – increases the internal disorder and entropy increases after absorbing the external heat. At the same time, the other part increases the internal energy, which acts as heat storage. Thus, the difference between internal energy – storage – and entropy – sink – is available as free energy that can be turned into work.

In exponential state spaces, the storage and the sink terms have the same order. However, for the pairing space in the first regime, the entropy's sink absorbs and turns all the heat into internal disorder since it has more room than exponential spaces. Surely, the emerging states cause the system to turn into a sink of heat such that the system is always at minimum free energy.

At the same time, in the second regime, the Hamiltonian has the same order as the entropy, and therefore, the usual thermodynamic processes can occur.

At first sight, it might seem unrealistic to assume Hamiltonians that are not asymptotically extensive. On the contrary, such Hamiltonians are not unknown in physics. For example, as we showed in section (1.3), the Curie–Weiss model's Hamiltonian [8, 14, 65] is order $O(N^2)$. Classical statistical mechanics books suggest scaling this Hamiltonian by $1/N$ to prevent the specific free energy from diverging without providing any sound reason – this is known as Kac prescription [8, 63]. However, we understand that this scaling converts the Hamiltonian order to $O(N)$ and makes the procedure tractable.

Although we do not report the details, it is worth mentioning that we have crafted a Hamiltonian that is of order $O(N \ln N)$ and is in the pairing state space. Interestingly, this model has well-defined free energy and a phase transition that numerical simulation confirms its prediction.

5.3 Pairing Models Applications: Delivery Joint Venture

As mentioned in section (1.2.2), we can generalise the delivery joint venture model for N different actors. At the same time, by assuming $C/2$ is the cost for driving a distance L, and $0 < \alpha$ is the free parameter of the model, the average cost of the aggregate, denoted by C_N, found as

$$C_N = C\left[N + (\alpha - 1)\langle 2n_p\rangle\right], \tag{5.58}$$

or specific cost as

$$\mathcal{C}_N \equiv \frac{C_N}{N} = C\left[1 + (\alpha - 1)\langle m_N\rangle\right]. \tag{5.59}$$

Recall that for $\alpha = 1$, all the actors are in a stand-alone state, and α less than one corresponds to cooperation, especially $\alpha = 1/2$. Similarly, for α greater than one, the system is in a competitive state, or say, the cost of delivery increases when another delivery company is in the neighbourhood.

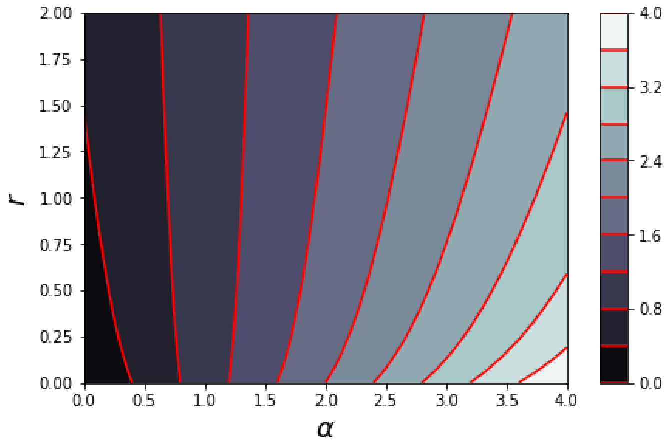


Figure 5.3: The specific cost, say $\mathcal{C}_N$, for a system of six actors.

Using equation (3.89), which finds the average of the B-model in terms of $c_N(r)$ and $c_{N-2}(r)$, the specific cost obtains as

$$\mathcal{C}_N = C\left[1 + (\alpha - 1)(N - 1)\frac{c_{N-2}(r)}{c_N(r)}\right], \tag{5.60}$$

or for $1 \ll N$, by using equation (3.91), the last results writes as

$$\mathcal{C}_N = C\left[1 + (\alpha - 1)\mathrm{e}^{-\sqrt{\frac{r}{N}}}\right]. \tag{5.61}$$

We see that $\mathcal{C}_N$ depends on two free parameters, r and α, as plotted in figure (5.3) for six actors. Here, we assumed $C = 1$. Note that α controls the cost of cooperation/competition, whereas r controls the probability of making a joint venture. In figure (5.3), the band $\mathcal{C}_N = 1$ shows there is no incentive to work together or compete. Outside this band, depending on the parameters, one regime dominates and influences the cost.

Chapter

SIX

Summary and Future Work

As we have seen, the central idea of this research program was to restrict the definition of emergence to states' emergence. However, new emerging states are not the property of stand-alone elements and cannot be deduced from their states. Therefore, the emerging states must be observed directly to be included in modelling the phenomena of interest. Despite that, the emerging states have a mathematically tractable structure suitable for analytical and numerical modelling.

We saw that the state spaces with the emerging states are growing faster than exponentially for the number of elements in the system. Faster than exponential growth breaks the additivity property of quantities usually used in information theory and statistical mechanics.

At the same time, to study the state emergence more rigorously, we further restricted the mechanism of emergence to pairwise compounding and introduced pairing models: Paring Balls and Coins models. Although we postponed studying them for future works, the two pairing models are generalisable to more complex mechanisms. In other words, the pairing model machinery by which the results were derived applies to more complex mechanisms that include more elements in a compound and allow variates of combinations. Even more, compounds can make compounds, and a hierarchy of structures emerges.

In pairing models, the faster than exponential growth of the state space volumes and their asymptotic leading terms were derived and quantitatively showed that the volume is in order $O(N^N)$. Furthermore, the derivation was based on a combinatorial argument that resulted in a recursive relation among volumes for different sizes. Thus, one can say, all the relevant information of the emerging states is encoded in the corresponding recursive relation.

The pairing model is both insightful and instrumental as a combinatorial object, and it enabled the modelling of the randomness and uncertainty with the same emerging properties. Consequently, we proposed two probability distributions with large deviation properties. Interestingly, the statistics of the distributions are expressible in closed form. We also derived the joint and marginal distributions to include them in the catalogue of analytical results of statistical pairing models.

We obtained limiting distributions that are again defined in exponential state spaces by properly scaling parameters for the statistical pairing model. Notably, the relevance of the limiting distributions to the main ones is similar to the relation between the Poisson distribution and the Binomial distribution [22]. Alongside these results, we also touched on statistical inference by finding maximum likelihood estimates of parameters and obtained a conjugate prior for the Bayesian inference.

One of the mentionable features of the limiting distribution is the similarity of its expectation to the order parameter of a second-order phase transition. Pairing model distributions have a zero average for stand-alone elements in the limit $N \to \infty$. However, for the limiting distribution, the mean is non-zero. For a continuous function, the zero to non-zero switching is a discontinuity in the first derivative, which indicates second-order phase transition occurs for the order parameter. Although the pairing models do not have dynamics and consequently do not have a phase transition per se, the change in the mean happens in the limit $N \to \infty$. Therefore, models that incorporate the pairing models as their building blocks and have an internal mechanism through which their scaling law changes from one distribution to the other will manifest the second-order phase transition.

In the meantime, studying the speed of large deviation property and the state space volume directed us to propose a mapping function to control the diverging specific values. Using the standard statistical mechanics, we observed that the free specific energy diverges. So, to overcome this problem in pairing state space, we proposed a one-to-one increasing function that maps the statistical mechanic's quantity to their thermodynamic ones such that in the limit $N \to \infty$, they are converging.

Studying the Shannon entropy of the pairing models' ensembles revealed that the entropy resulting from the uncertainty in pairing is separable from the uncertainty from stand-alone elements, say, the entropy of the C-model decomposes to the entropy of the B-model and the Binomial distribution.

The entropy of the subsystems, up to mesoscopic sizes, is additive. The mesoscopic scale is defined as order $O(\sqrt{N})$, and the entropy of subsystems in this order is the sum of all its parts. The decomposition of larger subsystems' entropy and its relation to the system entropy is an open question, and we postpone it to future works.

Unlike systems with independent elements, the mutual information between one element and the rest is non-zero. However, for large system sizes, mutual information approaches zero. One interpretation of this observation is as follows: in a macroscopic system, similar to stand-alone elements, pairs do not carry information about the system, or say, the pairs look like independent elements.

The mutual information between two elements approaches zero in the limit $N \to \infty$. It is the same for all interaction information as a higher-order form of mutual information. To have non-zero interaction information, we conjecture that one needs to include more complex emerging states by including more elements in a compound.

Inspired by paring models, we proposed a paring time series that mixes certainty and uncertainty in time. We showed that the Shannon entropy is well-defined for infinite time series when the probability distribution on the length of the pairs follows an exponential or power-law distribution. Having a well-defined Shannon entropy provides the foundation for further study. For instance, one can study similar stochastic quantities that are defined for a Markov process. Similarly, finding information-theoretic quantities like active information storage [42] or predictive information [45] for pairing time series is the next step to investigate further.

The proof of the existence of Shannon entropy for infinite pairing time series relied on an inequality that finds an upper bound in the limit. We propose that replacing the Shannon entropy with non-extensive entropies has the potential to derive the upper bound that fails for the uniform distribution or power-law distributions with an exponent smaller than one.

In addition, the mapping that we found for the pairing space divergence provides a solution to control the divergence of specific free energy, entropy and internal energy. Moreover, this program can be repeated for other growth rates when one finds the corresponding mapping. Thus, contrary to Kac's prescription, our proposal is not merely a mathematical convenience and has a more detailed explanation for the reason for applying it.

Finally, we constructed an elementary model with some dynamics and emerging states – delivery joint venture – to show that the pairing models can be utilised in more elaborated modelling.

Here, we suggest a list of future works and research aligned with the thesis:

- Generalising the compounding mechanism and studying their state-space geometry – A brief sketch of the details is included in appendix (E.1).
- Finding the corresponding probability distributions for new compounding mechanisms – A brief sketch of the details is included in appendix (E.1).
- The normalisation constants, namely $c_n(r)$, are closed under the derivative/anti-derivative operations. Therefore, they might possess an algebraic structure. Studying this abstract structure for $c_n(r)$ and normalisation constants for new compounding mechanisms is a possible open line of research.
- Revisiting the entropy, mutual entropy and interaction entropy for new compounding mechanisms.
- We showed that in pairing models, the entropy of subsystems is additive up to mesoscopic scales. Increasing system sizes from order $O(\sqrt{N})$ is an open question.
- For pairing time series, since the rate entropy is well-defined, we guess the Asymptotic Equipartition Property theorem must also be valid.
- Similar to Markov processes, the pairing time series can be represented by a finite number of states for uncertain parts of the dynamics. Then, certain evolutions can be modelled as a definite recurrence. So, exploring techniques similar to the first passage is one possibility to calculate different infinite time limits.
- As briefly mentioned in chapter (5), at the moment of writing this thesis, we are working on a Hamiltonian in order $O(N \ln N)$ for a system in the pairing state

space. Different Hamiltonians and state space growth rates are the natural next step for future research.

Bibliography

[1] *The Routledge handbook of emergence*. Routledge handbooks in philosophy. Routledge, Abingdon, Oxon ;, 2019 - 2019.

[2] P. W. Anderson. More is different. *Science*, 177(4047):393–396, 1972.

[3] Sámuel G Balogh, Gergely Palla, Péter Pollner, and Dániel Czégel. Generalized entropies, density of states, and non-extensivity. *Scientific reports*, 10(1):15516, 2020.

[4] Saugata. Basu. *Algorithms in Real Algebraic Geometry*. Algorithms and Computation in Mathematics, 10. Springer Berlin Heidelberg, Berlin, Heidelberg, 2nd ed. 2006. edition, 2006.

[5] Robert W. Batterman. Why equilibrium statistical mechanics works: Universality and the renormalization group. 65(2):183–208, June 1998.

[6] Mark A. Bedau. Weak emergence. *Philosophical Perspectives*, 11:375–399, 1997.

[7] C.M. Bender, S. Orszag, and S.A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*. Advanced Mathematical Methods for Scientists and Engineers. Springer, 1999.

[8] J.J. Binney, N.J. Dowrick, A.J. Fisher, and M.E.J. Newman. *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*. Clarendon Press, 1992.

[9] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, 2006.

[10] Herbert B. Callen. *Thermodynamics and an Introduction to Thermostatistics.* Second edition, 1985.

[11] David J. Chalmers. Strong and weak emergence. In *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion.*, Oxford, 2008. Oxford University Press.

[12] Feng-Cheng Chang. Recursive formulas for the partial fraction expansion of a rational function with multiple poles. *Proceedings of the IEEE*, 61(8):1139–1140, 1973.

[13] Irun R. Cohen and David Harel. Explaining a complex living system: Dynamics, multi-scaling and emergence, 4 2007.

[14] Marius Costeniuc, Richard S. Ellis, and Hugo Touchette. Complete analysis of phase transitions and ensemble equivalence for the Curie-Weiss-Potts model. *Journal of Mathematical Physics*, 46(6), 2005.

[15] Thomas M Cover and Joy A Thomas. *Elements of Information Theory.* John Wiley & Sons, Incorporated, Hoboken, 2006.

[16] P.C.P. Davies, P. Clayton, P. Davies, International Society for Science, Religion, and Oxford University Press. *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion.* ISSR Library. OUP Oxford, 2006.

[17] N.G. de Bruijn. *Asymptotic Methods in Analysis.* Bibliotheca mathematica. Dover Publications, 1981.

[18] George F. R. Ellis. Emergence in solid state physics and biology. 50(10):1098–1139, August 2020.

[19] Richard S Ellis. *Entropy, large deviations, and statistical mechanics.* Grundlehren der mathematischen Wissenschaften ; 271. Springer-Verlag, New York, 1985.

[20] Maria Dolores Esteban and Domingo Moralse. A summary on entropy statistics. *Kybernetika*, 31(4):337–346, 1995.

[21] Amiel Feinstein. *Foundations of Information Theory.* McGraw-Hill, 1958.

[22] W. Feller. *An Introduction to Probability Theory and Its Applications: Volume I*. Number v. 1 in Wiley series in probability and mathematical statistics. John Wiley & sons., third edition edition, 1968.

[23] E. Fermi. *Thermodynamics*. Dover books in physics and mathematical physics. Dover Publications, 1956.

[24] Jeffrey Goldstein. Emergence as a construct: History and issues. 1(1):49–72, March 1999.

[25] Jeffrey Goldstein. Emergence then and now: Concepts, criticisms, and rejoinders: Introduction to pepper's 'emergence,'. *Emergence: Complexity and Organization*, 6(4):66–71, 2004.

[26] Rudolf Hanel and Stefan Thurner. When do generalized entropies apply? how phase space volume determines entropy. *EPL (Europhysics Letters)*, 96(5):50003, 2011.

[27] T.L. Hill. *Statistical Mechanics: Principles and Selected Applications*. Dover Books on Physics. McGraw-Hill, 1956.

[28] T.L. Hill. *An Introduction to Statistical Thermodynamics*. Addison-Wesley series in chemistry. Dover Publications, 1986.

[29] J.H. Holland. *Emergence: From Chaos to Order*. Popular science / Oxford University Press. Oxford University Press, 2000.

[30] KI Hopcraft, E Jakeman, and JO Matthews. Generation and monitoring of a discrete stable random process. *Journal of Physics A: Mathematical and General*, 35(49):L745, 2002.

[31] Velimir M Ilić, Antonio Maria Scarfone, and Tatsuaki Wada. Equivalence between four versions of thermostatistics based on strongly pseudoadditive entropies. *Physical Review E*, 100(6):062135, 2019.

[32] Eric Jakeman, Sean Phayre, and Eric Renshaw. The evolution and measurement of a population of pairs. *Journal of applied probability*, 32(4):1048–1062, 1995.

[33] Henrik Jeldtoft Jensen, Roozbeh H Pazuki, Gunnar Pruessner, and Piergiulio Tempesta. Statistical mechanics of exploding phase spaces: ontic open systems. *Journal of Physics A: Mathematical and Theoretical*, 51(37):375002, aug 2018.

[34] Henrik Jeldtoft Jensen and Piergiulio Tempesta. Group entropies: From phase space geometry to entropy functionals via group theory. *Entropy*, 20(10):804, 2018.

[35] Walter T. Grandy Jr. *Entropy and the Time Evolution of Macroscopic Systems*. International Series of Monographs on Physics. 2008.

[36] A. I. Khinchin. *Mathematical Foundations of Information Theory*. Dover Books on Mathematics. 1957.

[37] Charles Kittel. *Elementary statistical physics*. Wiley, New York, 1958.

[38] Jan Korbel, Rudolf Hanel, and Stefan Thurner. Classification of complex systems by their sample-space scaling exponents. *New Journal of Physics*, 20(9):093007, 2018.

[39] Jan Korbel, Rudolf Hanel, and Stefan Thurner. Information geometric duality of ϕ-deformed exponential families. *Entropy*, 21(2):112, 2019.

[40] Jan Korbel, Rudolf Hanel, and Stefan Thurner. Information geometry of scaling expansions of non-exponentially growing configuration spaces. *The European Physical Journal Special Topics*, 229(5):787–807, 2020.

[41] H. T. Kung and D. M. Tong. Fast algorithms for partial fraction decomposition. 6(3):582–593, September 1977.

[42] Joseph T. Lizier. Jidt: An information-theoretic toolkit for studying the dynamics of complex systems. *Frontiers in Robotics and AI*, 1:11, 2014.

[43] AO Lovejoy, A Juarrero, and CA Rubino. The meanings of 'emergence' and its modes. *Emergence: Complexity and Organization*, 10(1):62, 2008.

[44] Youneng Ma, Jinhua Yu, and Yuanyuan Wang. An easy pure algebraic method for partial fraction expansion of rational functions with multiple high-order poles. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 61(3):803–810, 2014.

[45] D.J.C. MacKay, J.C. MacKay, D.J.C.M. Kay, and Cambridge University Press. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[46] Jon Mathews and R. L Walker. *Mathematical Methods Of Physics*. Addison-Wesley, second edition edition, 1970.

[47] JO Matthews, KI Hopcraft, and E Jakeman. Generation and monitoring of discrete stable random processes using multiple immigration population models. *Journal of Physics A: Mathematical and General*, 36(46):11585, 2003.

[48] K.P. Murphy. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning series. MIT Press, 2012.

[49] Giuseppe Mussardo. *Statistical Field Theory*. Oxford Graduate Texts. 2010.

[50] T. Needham. *Visual Complex Analysis*. Comparative Pathobiology - Studies in the Postmodern Theory of Education. Clarendon Press, 1998.

[51] James R. Norris. *Markov Chains*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, 15th edition, 2009.

[52] Timothy O'Connor. Emergent properties. *American Philosophical Quarterly*, 31(2):91–104, 1994.

[53] Roozbeh H Pazuki and Henrik Jeldtoft Jensen. New probability distribution describing emergence in state space. *Journal of Physics Communications*, 5(9):095002, September 2021.

[54] Oliver Penrose. *Foundations of Statistical Mechanics*. 1970.

[55] Bateman Manuscript Project, H. Bateman, A. Erdélyi, and United States. Office of Naval Research. *Higher Transcendental Functions*. Number v. 2 in Bateman manuscript project. McGraw-Hill, 1953.

[56] Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 547–561. University of California Press, 1961.

[57] Rene L. Schilling. *Measures, integrals and martingales*. Cambridge University Press, 2011.

[58] Erwin Schrodinger. *Statistical Thermodynamics*. 1989.

[59] Robert Sedgewick and Philip Flajolet. *An Introduction to The Analysis Of Algorithms*. Addison-Wesley, second edition edition, 2013.

[60] M.P Soares, F Riet-Correa, D.R Smith, M Pereira Soares, M.C Mendez, and A.L Brandolt. Experimental intoxication by larvae of perreyia flavipes konow, 1899 (hymenoptera: Pergidae) in pigs and some aspects on its biology. *Toxicon*, 39(5):669–678, 2001.

[61] JV Stone. *Information Theory: A Tutorial Introduction*. Tutorial Introduction Book. Sebtel Press, 2015.

[62] Piergiulio Tempesta and Henrik Jeldtoft Jensen. Universality classes and information-theoretic measures of complexity via group entropies. *Scientific reports*, 10(1):5952, 2020.

[63] Colin J. Thompson. *Mathematical Statistical Mechanics*. Books In Applied Mathematics. 1972.

[64] Colin J. Thompson. *Classical Equilibrium Statistical Mechanics*. 1988.

[65] Hugo Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1-3):1–69, 2009.

[66] Hugo Touchette. A basic introduction to large deviations: Theory, applications, simulations. pages 1–58, 2011.

[67] Constantino Tsallis. Possible Generalization of Boltzmann-Gibbs Statistics. 52:479–487, 1988.

[68] Constantino Tsallis. Beyond boltzmann–gibbs–shannon in physics and elsewhere. *Entropy*, 21(7):696, 2019.

[69] Constantino Tsallis et al. *Introduction to nonextensive statistical mechanics*. 2009.

[70] Juan Pablo Vigneaux. Information theory with finite vector spaces. *IEEE Transactions on Information Theory*, 65(9):5674–5687, 2019.

[71] H. S. Wilf. *Generatingfunctionology*. A K Peters/CRC Press, third edition (3rd ed.) edition, 2006.

[72] R.W. Yeung. A new outlook on shannon's information measures. *IEEE Transactions on Information Theory*, 37(3):466–474, 1991.

Appendix

A

Pairing Models

A.1 Recursive Relation: Solution

Denoting $G(z)$ as a converging power series like

$$G(z) = \sum_{n\geq 0} a_n \frac{z^n}{n!}, \tag{A.1}$$

for $z \in A \subseteq \mathbb{R}$. Next, we assume $\Omega_s(n)$ is the coefficient of term $\frac{z^n}{n!}$ in power series expansion of $G(z)$ or

$$a_n \equiv \Omega_s(n). \tag{A.2}$$

Thus, the recursive relation in equation (2.9) is written as

$$a_{n+1} = sa_n + pna_{n-1}. \tag{A.3}$$

After that, by multiplying each side of the last equation by $\frac{z^n}{n!}$ and sum for $n \geq 1$ we derive to the following equation,

$$\sum_{n\geq 1} a_{n+1} \frac{z^n}{n!} = s \sum_{n\geq 1} a_n \frac{z^n}{n!} + p \sum_{n\geq 1} na_{n-1} \frac{z^n}{n!}. \tag{A.4}$$

The left-hand side term is

$$G'(z) = \sum_{n\geq 1} a_n \frac{z^{n-1}}{n-1!} = \sum_{n\geq 0} a_{n+1} \frac{z^n}{n!} = a_1 + \sum_{n\geq 1} a_{n+1} \frac{z^n}{n!}. \tag{A.5}$$

The first term on the right-hand side is

$$G(z) = \sum_{n\geq 0} a_n \frac{z^n}{n!} = a_0 + \sum_{n\geq 1} a_n \frac{z^n}{n!}, \tag{A.6}$$

and the second term in right hand side can be written by multiplying the $G(z)$ by z,

$$zG(z) = \sum_{n\geq 0} a_n \frac{z^{n+1}}{n!} = \sum_{n\geq 0} (n+1) a_n \frac{z^{n+1}}{n+1!} = \sum_{n\geq 1} n a_{n-1} \frac{z^n}{n!}. \tag{A.7}$$

Plugging back relations (A.5), (A.6) and (A.7) in (A.4) obtains

$$G'(z) - a_1 = sG(z) - sa_0 + pzG(z). \tag{A.8}$$

The initial condition in equation (2.10) requires

$$a_0 = 1, \qquad a_1 = s, \tag{A.9}$$

and therefore

$$G'(z) - (s + pz)G(z) = 0. \tag{A.10}$$

It is a first-order differential equation. Solving its solution results in

$$G(z) = \exp(sz) \exp(\frac{pz^2}{2}). \tag{A.11}$$

Each factor of $G(z)$ has a power series expansion such as

$$\mathrm{e}^{sz} = \sum_{n\geq 0} \frac{s^n z^n}{n!}, \tag{A.12}$$

and

$$\mathrm{e}^{\frac{pz^2}{2}} = \sum_{n\geq 0} \frac{p^n z^{2n}}{2^n n!}. \tag{A.13}$$

Indeed, we can write them back as a convolution of one by considering the other as the kernel of the convolution

$$G(z) = \sum_{i\geq 0} \frac{s^i z^i}{i!} \sum_{j\geq 0} \frac{p^j z^{2j}}{2^j j!} = \sum_{i,j\geq 0} \frac{s^i (\frac{p}{2})^j}{i! j!} z^{i+2j} =$$

$$\sum_{i,j\geq 0} \frac{s^i (\frac{p}{2})^j (i+2j)!}{i! j!} \times \frac{z^{i+2j}}{(i+2j)!} \tag{A.14}$$

such that

$$G(z) = \sum_{n\geq 0} \left[\sum_k \circ \right] \frac{z^n}{n!} \tag{A.15}$$

where ($\circ$) represents all the coefficients indexed on k such that $i + 2j = n$. There are two cases for even and odd n:

- $n = 2p$ is even and $i + 2j = n$ for consecutive $0 \leq k \leq \frac{n}{2}$:

  1. $i \in \{0, 2, ..., 2p\} = \{0, 2, ..., n\} \equiv 2k$.
  2. $j \in \{p, p-1, ..., 0\} = \{\frac{n}{2}, \frac{n}{2} - 1, ..., 0\} \equiv \frac{n}{2} - k$.
  3. $i - j \in \{-p, 3-p, 6-p, ..., 2p\} = \{-\frac{n}{2}, 3 - \frac{n}{2}, ..., n\} \equiv 3k - \frac{n}{2}$.
  4. $i + 2j \in \{2p, 2p, ..., 2p\} = \{n, n, ..., n\}$.

  So inside of the bracket must be

$$\sum_{0\leq k\leq \frac{n}{2}} n! \frac{s^{2k} (\frac{p}{2})^{\frac{n}{2}-k}}{2k! (\frac{n}{2} - k)!}. \tag{A.16}$$

- $n = 2p + 1$ is odd and $i + 2j = n$:

  1. $i \in \{1, 3, ..., 2p+1\} = \{1, 3, ..., n\} \equiv 2k + 1$.
  2. $j \in \{p, p-1, ..., 0\} = \{\frac{n-1}{2}, \frac{n-3}{2}, ..., 0\} \equiv \frac{n-1}{2} - k$.
  3. $i - j \in \{1-p, 4-p, 7-p, ..., 2p+1\} = \{\frac{3-n}{2}, \frac{9-n}{2}, ..., n\} \equiv 3k + \frac{3}{2} - \frac{n}{2}$.
  4. $i + 2j \in \{2p+1, 2p+1, ..., 2p+1\} = \{n, n, ..., n\}$.

  So inside of the bracket must be

$$\sum_{0\leq k\leq \frac{n-1}{2}} n! \frac{s^{2k+1} (\frac{p}{2})^{\frac{n-1}{2}-k}}{(2k+1)! (\frac{n-1}{2} - k)!}. \tag{A.17}$$

By defining the following functions

$$\mathbf{1}_{\text{odd}}(n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}, \tag{A.18}$$

and

$$\mathbf{1}_{\text{even}}(n) = \begin{cases} 0 & \text{if } n \text{ is even} \\ 1 & \text{if } n \text{ is odd} \end{cases}, \tag{A.19}$$

$G(z)$ becomes

$$G(z) = \sum_{n\geq 0}\left[\mathbf{1}_{\text{odd}}(n)\sum_{0\leq k\leq \frac{n}{2}} n!\frac{s^{2k}(\frac{p}{2})^{\frac{n}{2}-k}}{2k!(\frac{n}{2}-k)!} + \mathbf{1}_{\text{even}}(n)\sum_{0\leq k\leq \frac{n-1}{2}} n!\frac{s^{2k+1}(\frac{p}{2})^{\frac{n-1}{2}-k}}{(2k+1)!(\frac{n-1}{2}-k)!}\right]\frac{z^n}{n!}. \tag{A.20}$$

So, $\Omega_s(N)$ is the coefficient of the above power series expansion:

$$\Omega_s(2N) = (2N)!(\frac{p}{2})^N\sum_{k=0}^{N}\frac{(\frac{2s^2}{p})^k}{2k!(N-k)!} \tag{A.21}$$

and

$$\Omega_s(2N+1) = (2N+1)!(\frac{p}{2})^N\sum_{k=0}^{N}\frac{s(\frac{2s^2}{p})^k}{(2k+1)!(N-k)!}, \tag{A.22}$$

or in general

$$\Omega_s(N) = N!(\frac{p}{2})^{\lfloor N/2\rfloor}\sum_{k=0}^{\lfloor N/2\rfloor}\frac{(\frac{2s^2}{p})^k}{(2k)!(\lfloor\frac{N}{2}\rfloor-k)!}. \tag{A.23}$$

It is possible to write both equations (A.21) and (A.22) differently. We will see, that the next forms are more intuitively related to a combinatorial argument, while the resulted ones are more suitable for finding the asymptotic leading terms in the next section.

To start, let rewrite equation (A.21) as

$$\Omega_s(2N) = \sum_{k=0}^{N}\frac{(2N)!}{2k!(2N-2k)!}\times\frac{(2N-2k)!}{(N-k)!2^{N-k}}\times s^{2k}p^{N-k}$$

$$= \sum_{k=0}^{N}\binom{2N}{2N-2k}(2N-2k-1)!!s^{2k}p^{N-k} \qquad (n_p = N-k)$$

$$= \sum_{n_p=0}^{N} \binom{2N}{2n_p} (2n_p - 1)!! s^{2N-2n_p} p^{n_p}. \tag{A.24}$$

Similarly, equation (A.22) is written as

$$\Omega_s(2N+1) = \sum_{k=0}^{N} \frac{(2N+1)!}{(2k+1)!(2N-2k)!} \times \frac{(2N-2k)!}{(N-k)!2^{N-k}} \times s^{2k+1} p^{N-k}$$

$$= \sum_{k=0}^{N} \binom{2N+1}{2N-2k} (2N-2k-1)!! s^{2k+1} p^{N-k} \qquad (n_p = N-k)$$

$$= \sum_{n_p=0}^{N} \binom{2N+1}{2n_p} (2n_p - 1)!! s^{2N+1-2n_p} p^{n_p}. \tag{A.25}$$

In general, for odd and even N, it writes as

$$\Omega_s(N) = \sum_{n_p=0}^{\lfloor N/2 \rfloor} \binom{N}{2n_p} (2n_p - 1)!! s^{N-2n_p} p^{n_p}. \tag{A.26}$$

Observe that $\binom{N}{2n_p}(2n_p - 1)!!$ is the degeneracy correspond to n_p pairs among N elements. And since $N - 2n_p$ is the number of stand-alone elements, s^{N-2n_p} enumerates the distinct configurations of stand-alone elements, whereas p^{n_p} enumerates pairs.

A.2 Recursive Relation: Second Method

In the previous section, we derived the generating function of $\Omega_s(N)$ in equation (A.11) as

$$G(z) = \exp(sz + \frac{pz^2}{2}), \tag{A.27}$$

and as we shall see, it is possible to write $\Omega_s(N)$ in a different form than equation (A.26). To start, let us use the well-known exponential generating function of the Hermite polynomial [55]

$$\sum_{n=0}^{\infty} H_n(x) \frac{y^n}{n!} = \exp(2xy - y^2), \tag{A.28}$$

where $H_n(x)$ is the Hermite polynomial with degree n. Using the transformation

$$\begin{aligned} y &= \frac{z}{i}\sqrt{\frac{p}{2}}, \\ x &= \frac{is}{\sqrt{2p}}, \end{aligned} \tag{A.29}$$

where $i = \sqrt{-1}$, the right hand side of equation (A.28) becomes

$$\exp(2xy - y^2) = \exp(sz + \frac{pz^2}{2}) = G(z). \tag{A.30}$$

At the same time, the left-hand side of equation (A.28) transforms as

$$\sum_{n=0}^{\infty} H_n(x)\frac{y^n}{n!} = \sum_{n=0}^{\infty} \left(\frac{p}{2}\right)^{n/2} H_n(\frac{is}{\sqrt{2p}})\frac{z^n}{i^n n!}, \tag{A.31}$$

in which the Hermite polynomial's domain is transformed into the imaginary axis. So, the generating function $G(z)$ is

$$G(z) = \sum_{n=0}^{\infty} \left(\frac{p}{2}\right)^{n/2} H_n(\frac{is}{\sqrt{2p}})\frac{z^n}{i^n n!}. \tag{A.32}$$

Considering the identity between the Hermit polynomial and the generalised Laguerre polynomial, denoted by $L_n^{(\alpha)}(x)$, we have [55]

$$\begin{aligned} H_{2n}(x) &= n!(-1)^n 2^{2n} L_n^{(-\frac{1}{2})}(x^2) \\ H_{2n+1}(x) &= n!(-1)^n 2^{2n+1} x L_n^{(\frac{1}{2})}(x^2), \end{aligned} \tag{A.33}$$

and the Hermit polynomial on the imaginary axis can be written as

$$\begin{aligned} H_{2n}(\frac{is}{\sqrt{2p}}) &= n!(-1)^n 2^{2n} L_n^{(-\frac{1}{2})}(\frac{-s^2}{2p}) \\ H_{2n+1}(\frac{is}{\sqrt{2p}}) &= n!(-1)^n 2^{2n+1} \left(\frac{is}{\sqrt{2p}}\right) L_n^{(\frac{1}{2})}(\frac{-s^2}{2p}). \end{aligned} \tag{A.34}$$

Therefore, substituting these results in equation (A.32) obtains the generating function $G(z)$ in terms of the generalised Laguerre polynomial as

$$G(z) = \sum_{2n} \left(\frac{p}{2}\right)^n H_{2n}(\frac{is}{\sqrt{2p}})\frac{z^{2n}}{i^{2n}(2n)!} + \sum_{2n+1} \sqrt{\frac{p}{2}} \left(\frac{p}{2}\right)^n H_{2n+1}(\frac{is}{\sqrt{2p}})\frac{z^{2n+1}}{i^{2n+1}(2n+1)!}$$

$$= \sum_{2n} \left(\frac{p}{2}\right)^n n!(-1)^n 2^{2n} L_n^{(-\frac{1}{2})}(\frac{-s^2}{2p}) \frac{z^{2n}}{(-1)^n(2n)!}$$

$$+ \sum_{2n+1} \sqrt{\frac{p}{2}} \left(\frac{p}{2}\right)^n n!(-1)^n 2^{2n+1} \left(\frac{is}{\sqrt{2p}}\right) L_n^{(\frac{1}{2})}(\frac{-s^2}{2p}) \frac{z^{2n+1}}{i(-1)^n(2n+1)!}$$

$$= \sum_{2n} n!(2p)^n L_n^{(-\frac{1}{2})}(\frac{-s^2}{2p}) \frac{z^{2n}}{(2n)!} + \sum_{2n+1} sn!(2p)^n L_n^{(\frac{1}{2})}(\frac{-s^2}{2p}) \frac{z^{2n+1}}{(2n+1)!}. \tag{A.35}$$

by which, from the definition of $G(z)$ in equation (A.1), we directly find

$$\Omega_s(2n) = n!(2p)^n L_n^{(-\frac{1}{2})}(\frac{-s^2}{2p}), \tag{A.36}$$

$$\Omega_s(2n+1) = sn!(2p)^n L_n^{(\frac{1}{2})}(\frac{-s^2}{2p}). \tag{A.37}$$

A.3 Asymptotic Leading Term of $\Omega_s(N)$

In this section, we calculate the asymptotic leading term of $\Omega_s(N)$ in equation (A.23) which is written as

$$\Omega_s(N) = N!(\frac{p}{2})^{\lfloor N/2 \rfloor} \sum_{0 \le k \le \lfloor N/2 \rfloor} \frac{(\frac{2s^2}{p})^k}{(2k)!(\lfloor \frac{N}{2} \rfloor - k)!}. \tag{A.38}$$

In what follows, first, we numerically study the location of the maximum of the summand to exploit the exponentially fast drop of the other terms around the maximum. Next, using the Stirling approximation of $N!$ the leading terms simplify further. Besides, the numerical results will support the claim. And finally, the Euler-Maclaurin summation formula [59] derives an approximated leading term for the summation.

A.3.1 Numerical investigation

In this section, without loss of the generality, we numerically study $\Omega_2(N)$ instead of the general form, namely, $\Omega_s(N)$. The final finding regarding the maximum of the summand and the exponential decreasing of terms around the maximum do not affect by a different choices of s.

For $\Omega_2(N)$, consider the case for even numbers

$$\Omega_2(2N) = \frac{2N!}{2^N} \sum_{0 \le k \le N} \frac{2^{3k}}{(2k)!(N-k)!}. \tag{A.39}$$

We define

$$t_N(k) = \frac{2^{3k}}{(2k)!(N-k)!}. \tag{A.40}$$

For $N = 200$, figure (A.1) shows $t_N(k)$ where it is normalised and scaled on $[0, 1]$.

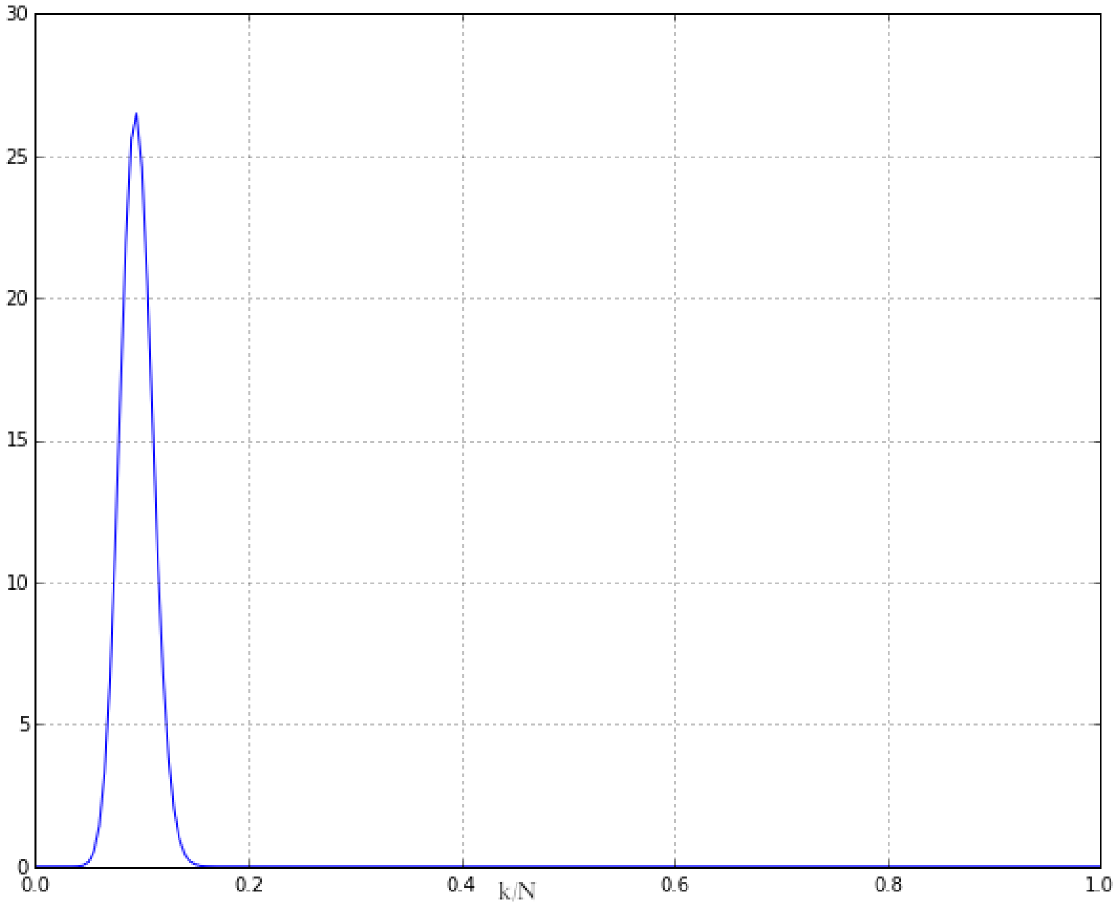


Figure A.1: $t_N(k)$ for $N = 200$

It shows that the maximum value of $t_N(k)$ is for k somewhere between $0 < k < N$. The ratio of two consecutive terms lets us find the maximum point,

$$\frac{t_N(k+1)}{t_N(k)} = \left(\frac{2^{3k+3}}{2^{3k}}\right)\left(\frac{2k!}{(2k+2)!}\right)\left(\frac{(N-k)!}{(N-k-1)!}\right)$$

$$= 2^3 \frac{(N-k)}{(2k+2)(2k+1)} = \frac{4(N-k)}{(k+1)(2k+1)}, \tag{A.41}$$

which $\frac{t_N(2)}{t_N(1)} > 1$ for $N > 2$ and $\frac{t_N(N)}{t_N(N-1)} < 1$ for $N > 1$. Indeed for

$$\frac{t_N(k+1)}{t_N(k)} = 1 \tag{A.42}$$

evaluates k for which $t_N(k)$ is maximum,

$$4(N-k) = (k+1)(2k+1) = 2k^2 + 3k + 1 \implies \tag{A.43}$$

$$2k^2 + 7k + 1 - 4N = 0 \implies \tag{A.44}$$

$$k = \pm\sqrt{2N + \frac{41}{16}} - \frac{7}{4} \tag{A.45}$$

k must be positive and it can be approximated like $k = \lfloor\sqrt{2N}\rfloor$. In figure (A.1) we can see $t_N(k)$ around its maximum decreases fast and becomes exponentially small in comparison to $t_N(k)$ maximum. To prove this claim, first we approximate $t_N(k)$ by using Stirling approximation for factorials

$$t_N(k) = \frac{2^{3k}}{(\frac{2k}{\mathrm{e}})^{2k}(\frac{N-k}{\mathrm{e}})^{N-k}} = \frac{2^k \mathrm{e}^{N+k}}{k^{2k}(N-k)^{N-k}}. \tag{A.46}$$

Thus

$$t_N(\sqrt{2N}) = \frac{2^{\sqrt{2N}}\mathrm{e}^{N+\sqrt{2N}}}{(\sqrt{2N})^{2\sqrt{2N}}(N-\sqrt{2N})^{N-\sqrt{2N}}} =$$

$$\frac{\mathrm{e}^{N+\sqrt{2N}}}{N^{\sqrt{2N}}(N-\sqrt{2N})^{N-\sqrt{2N}}}, \tag{A.47}$$

and

$$t_N(2\sqrt{2N}) = \frac{2^{2\sqrt{2N}}\mathrm{e}^{N+2\sqrt{2N}}}{(2\sqrt{2N})^{4\sqrt{2N}}(N-2\sqrt{2N})^{N-2\sqrt{2N}}}$$

$$= \frac{\mathrm{e}^{N+2\sqrt{2N}}}{16^{\sqrt{2N}}N^{2\sqrt{2N}}(N-\sqrt{2N})^{N-\sqrt{2N}}}, \tag{A.48}$$

Notice that the ratio of these two terms is

$$\frac{t_N(\sqrt{2N})}{t_N(2\sqrt{2N})} = \left(\frac{\mathrm{e}^{N+\sqrt{2N}}}{\mathrm{e}^{N+2\sqrt{2N}}}\right)\left(\frac{16^{\sqrt{2N}}N^{2\sqrt{2N}}(N-2\sqrt{2N})^{N-2\sqrt{2N}}}{N^{\sqrt{2N}}(N-\sqrt{2N})^{N-\sqrt{2N}}}\right)$$

$$= \left(\frac{16}{\mathrm{e}}\right)^{\sqrt{2N}} \left(1 - \frac{\sqrt{2N}}{N - \sqrt{2N}}\right)^{N-\sqrt{2N}} \left(1 - \frac{4}{\sqrt{2N}}\right)^{-\sqrt{2N}} \tag{A.49}$$

Using the following two limits

$$\lim_{N\to\infty} \left(1 - \frac{\sqrt{2N}}{N - \sqrt{2N}}\right)^{N-\sqrt{2N}} = \mathrm{e}^{-\sqrt{2N}}, \tag{A.50}$$

$$\lim_{N\to\infty} \left(1 - \frac{4}{\sqrt{2N}}\right)^{-\sqrt{2N}} = \mathrm{e}^{4}, \tag{A.51}$$

it finds

$$\lim_{N\to\infty} \frac{t_N(2\sqrt{2N})}{t_N(\sqrt{2N})} = \lim_{N\to\infty} \left(\frac{\mathrm{e}}{16}\right)^{\sqrt{2N}} \mathrm{e}^{\sqrt{2N}}\mathrm{e}^{-4} = \lim_{N\to\infty} \left(\frac{\mathrm{e}}{4}\right)^{2\sqrt{2N}} \mathrm{e}^{-4} \to 0. \tag{A.52}$$

So $t_N(k)$ becomes exponentially small when k moves from $\sqrt{2N}$ to $2\sqrt{2N}$. This is same on moving toward $k = 1$, as it is shown in Appendix A. In figure (A.2) and (A.3) exact values and approximated ratios are depicted for $6 \le k \le 1000$.

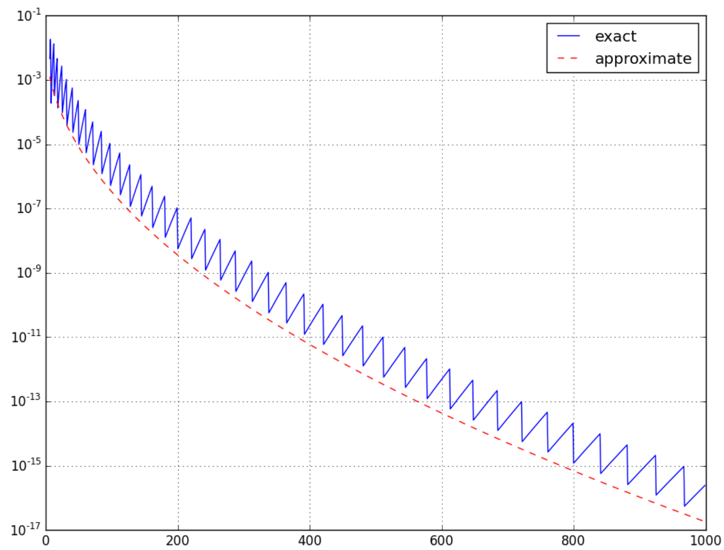


Figure A.2: $\frac{t_N(2\sqrt{2N})}{t_N(\sqrt{2N})}$ for $6 \le N \le 1000$

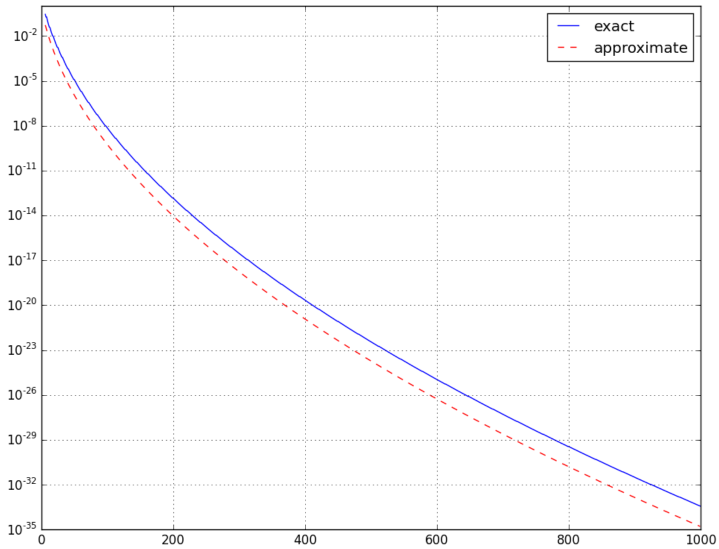


Figure A.3: $\frac{t_N(\frac{\sqrt{2N}}{2})}{t_N(\sqrt{2N})}$ for $6 \leq N \leq 1000$

These results allow us to divide the range of k into three regimes

$$\Omega_2(2N) = \frac{(2N)!}{2^N}\left(\frac{1}{N!} + \sum_{1\leq k\leq 2\sqrt{2N}} \frac{2^{3k}}{(2k)!(N-k)!} + \sum_{2\sqrt{2N}<k\leq N} \frac{2^{3k}}{(2k)!(N-k)!}\right)$$

$$\Omega_2(2N) = \frac{(2N)!}{2^N}\left(\frac{1}{N!} + \sum_{1\leq k\leq 2\sqrt{2N}} \frac{2^{3k}}{(2k)!(N-k)!} + \Delta\right) \tag{A.53}$$

where $\frac{1}{N!}$ and Δ are exponentially small in comparison to the sum in the middle. For $\Omega_s(2N)$ the range of the summation divides as

$$\Omega_s(2N) = \frac{(2N)!}{(2/p)^N}\left(\frac{1}{N!} + \sum_{1\leq k\leq 2\sqrt{s^2N/2p}} \frac{(\frac{2s^2}{p})^k}{(2k)!(N-k)!} + \Delta\right), \tag{A.54}$$

and the maximum is at $k = \lfloor\sqrt{s^2N/2p}\rfloor$.

A.3.2 Approximating $t_N(k)$

For large numbers like N, the Stirling approximation is

$$N! = \sqrt{2\pi N}(\frac{N}{\mathrm{e}})^N(1 + O(\frac{1}{N})). \tag{A.55}$$

Recall that the summand $t_N(k)$ for $\Omega_s(2N)$ is

$$t_N(k) = \frac{(\frac{2s^2}{p})^k}{(2k)!(N-k)!}. \tag{A.56}$$

By using the Stirling approximation to find the leading terms in $t_N(k)$ asymptotic expansion we have

$$t_N(k) = \frac{(\frac{2s^2}{p})^k}{\sqrt{4\pi k}(\frac{2k}{\mathrm{e}})^{2k}\sqrt{2\pi(N-k)}(\frac{N-k}{\mathrm{e}})^{N-k}(1+O(\frac{1}{k}))(1+O(\frac{1}{N-k}))}$$

$$= \frac{1}{2\sqrt{2}\pi\sqrt{k(N-k)}} \times \frac{(\frac{s^2}{2p})^k \mathrm{e}^{N+k}}{k^{2k}(N-k)^{N-k}} \times \frac{1}{(1+O(\frac{1}{k}))(1+O(\frac{1}{N-k}))}. \tag{A.57}$$

The maximum of $t_N(k)$ is at $k = \sqrt{s^2N/2p}$, and therefore for $N \gg 1$, around the maximum we must have $O(\frac{1}{k}) = O(\frac{1}{\sqrt{N}})$. At the same time by using Taylor expansion for small x we get $(1-x)^{-1} = 1 + O(x)$. Hence

$$\frac{1}{N-k} = \frac{1}{N(1-\frac{k}{N})} = \frac{1}{N}(1 + O(\frac{k}{N})) = \frac{1}{N} + O(\frac{k}{N^2}), \tag{A.58}$$

which around $k = \sqrt{s^2N/2p}$ is of the order

$$O(\frac{1}{N-k}) = O(\frac{1}{N} + O(\frac{1}{N^{\frac{3}{2}}})) = O(\frac{1}{N}). \tag{A.59}$$

Afterwards, both asymptotic orders combine as

$$\left[1 + O(\frac{1}{k})\right]\left[1 + O(\frac{1}{N-k})\right] = \left[1 + O(\frac{1}{\sqrt{N}})\right]\left[1 + O(\frac{1}{N})\right] = 1 + O(\frac{1}{\sqrt{N}}). \tag{A.60}$$

And finally,

$$t_N(k) = \frac{1}{2\sqrt{2}\pi\sqrt{k(N-k)}} \times \frac{(\frac{s^2}{2p})^k \mathrm{e}^{N+k}}{k^{2k}(N-k)^{N-k}} \times \left[1 + O(\frac{1}{\sqrt{N}})\right]$$

$$= \frac{1}{2\sqrt{2}\pi\sqrt{k(N-k)}} \times \frac{(\frac{s^2}{2p})^k \mathrm{e}^{N+k}}{k^{2k}N^{N-k}} \times \frac{(1-\frac{k}{N})^k}{(1-\frac{k}{N})^N} \times \left[1+O(\frac{1}{\sqrt{N}})\right]. \tag{A.61}$$

Other fractions can be expanded and written as their asymptotic leading terms. For instance, the first fraction rewrites as

$$\frac{1}{\sqrt{k(N-k)}} = \frac{1}{\sqrt{kN}}\left(1-\frac{k}{N}\right)^{-\frac{1}{2}}, \tag{A.62}$$

and using Taylor expansion for small x as $(1-x)^{-\frac{1}{2}} = 1+O(x)$, the fraction leading term must be

$$\frac{1}{\sqrt{kN}}\left(1+O(\frac{k}{N})\right) = \frac{1}{(\frac{s^2}{2p})^{\frac{1}{4}}N^{\frac{3}{4}}}\left(1+O(\frac{1}{\sqrt{N}})\right). \tag{A.63}$$

Remember that the $t_N(k)$ is maximum at $k=\sqrt{s^2N/2p}$, and we finds its magnitude around its peak.

Defining $k=c\sqrt{N}$ where $c\in\mathbb{R}$ and $0\le c\le 2\sqrt{s^2/2p}$, the next fraction writes as

$$\frac{(1-\frac{k}{N})^k}{(1-\frac{k}{N})^N} = \frac{(1-\frac{c}{\sqrt{N}})^{c\sqrt{N}}}{(1-\frac{c}{\sqrt{N}})^N} = \left(\frac{(1-\frac{c}{\sqrt{N}})^c}{(1-\frac{c}{\sqrt{N}})^{\sqrt{N}}}\right)^{\sqrt{N}}$$

$$= \left(\frac{1-\frac{c^2}{\sqrt{N}}+O(\frac{1}{N})}{\mathrm{e}^{-c}(1-\frac{c^2}{2\sqrt{N}}+O(\frac{1}{N}))}\right)^{\sqrt{N}} \tag{A.64}$$

The numerator is the Taylor expansion at $x=0$ in which we assume $c/\sqrt{N}$ is small. And the denominator is the asymptotic expansion of $(1-\frac{c}{\sqrt{N}})^{\sqrt{N}}$. Remember the below limit when $\sqrt{N}\to\infty$

$$\lim_{\sqrt{N}\to\infty}(1-\frac{c}{\sqrt{N}})^{\sqrt{N}} = \mathrm{e}^{-c}. \tag{A.65}$$

For small x, writing the Taylor expansion of the fraction $\frac{1}{1-x} = 1+x+O(x^2)$ implies the denominator obtains as

$$\mathrm{e}^{c}(1+\frac{c^2}{2\sqrt{N}}+O(\frac{1}{N})), \tag{A.66}$$

and therefore the fraction becomes

$$\left[\mathrm{e}^c(1+\frac{c^2}{2\sqrt{N}}+O(\frac{1}{N}))(1-\frac{c^2}{\sqrt{N}}+O(\frac{1}{N}))\right]^{\sqrt{N}}$$

$$=\mathrm{e}^{c\sqrt{N}}\left[1-\frac{c^2}{2\sqrt{N}}+O(\frac{1}{N})\right]^{\sqrt{N}}=\mathrm{e}^{c\sqrt{N}}\exp\{\sqrt{N}\ln(1-\frac{c^2}{2\sqrt{N}}+O(\frac{1}{N}))\}. \quad \text{(A.67)}$$

Next, the Taylor expansion $\ln(1-x)=-x+O(x^2)$ for logarithm provides

$$\mathrm{e}^{c\sqrt{N}}\exp\{\sqrt{N}(-\frac{c^2}{2\sqrt{N}}+O(\frac{1}{N}))\}=\mathrm{e}^{c\sqrt{N}-\frac{c^2}{2}}\exp\{O(\frac{1}{\sqrt{N}})\}$$

$$=A\mathrm{e}^k(1+O(\frac{1}{\sqrt{N}})). \quad \text{(A.68)}$$

In the last step we used $\mathrm{e}^{O(x)}=1+O(x)$, and defined the constant $A=\mathrm{e}^{-\frac{c^2}{2}}$. For maximum at $k=\sqrt{s^2N/2p}$ it corresponds to $c=\sqrt{s^2/2p}$, and therefore, $A=\mathrm{e}^{-s^2/4p}$. Thus, the fraction's asymptotic leading term must be

$$\frac{(1-\frac{k}{N})^k}{(1-\frac{k}{N})^N}=\mathrm{e}^{-\frac{s^2}{4p}}\mathrm{e}^k(1+O(\frac{1}{\sqrt{N}})). \quad \text{(A.69)}$$

Plugging back both fractions into equation (A.61), it finds

$$t_N(k)=\frac{1}{2\sqrt{2}\pi\mathrm{e}^{\frac{s^2}{4p}}(\frac{s^2}{2p})^{\frac{1}{4}}N^{\frac{3}{4}}}\left[1+O(\frac{1}{\sqrt{N}})\right]\mathrm{e}^k\left[1+O(\frac{1}{\sqrt{N}})\right]\frac{(\frac{s^2}{2p})^k\mathrm{e}^{N+k}}{k^{2k}N^{N-k}}\left[1+O(\frac{1}{\sqrt{N}})\right]. \quad \text{(A.70)}$$

After multiplying all three $\left[1+O(\frac{1}{\sqrt{N}})\right]$ terms and truncating up to the order $O(\frac{1}{\sqrt{N}})$, it results in

$$t_N(k)=\frac{1}{2^{\frac{3}{2}}\pi\mathrm{e}^{\frac{s^2}{4p}}(\frac{s^2}{2p})^{\frac{1}{4}}N^{\frac{3}{4}}}\times\frac{(\frac{s^2}{2p})^kN^k\mathrm{e}^{N+2k}}{k^{2k}N^N}\left[1+O(\frac{1}{\sqrt{N}})\right]$$

$$=\frac{1}{2^{\frac{3}{2}}\pi\mathrm{e}^{\frac{s^2}{4p}}(\frac{s^2}{2p})^{\frac{1}{4}}N^{\frac{3}{4}}}\left(\frac{\mathrm{e}}{N}\right)^N\left(\frac{\sqrt{\frac{s^2N}{2p}}\mathrm{e}}{k}\right)^{2k}\left[1+O(\frac{1}{\sqrt{N}})\right]. \quad \text{(A.71)}$$

Figure (A.4) compares approximated relations and the exact form in log-normal scale for $s=2$, $p=1$, and $N=1000$.

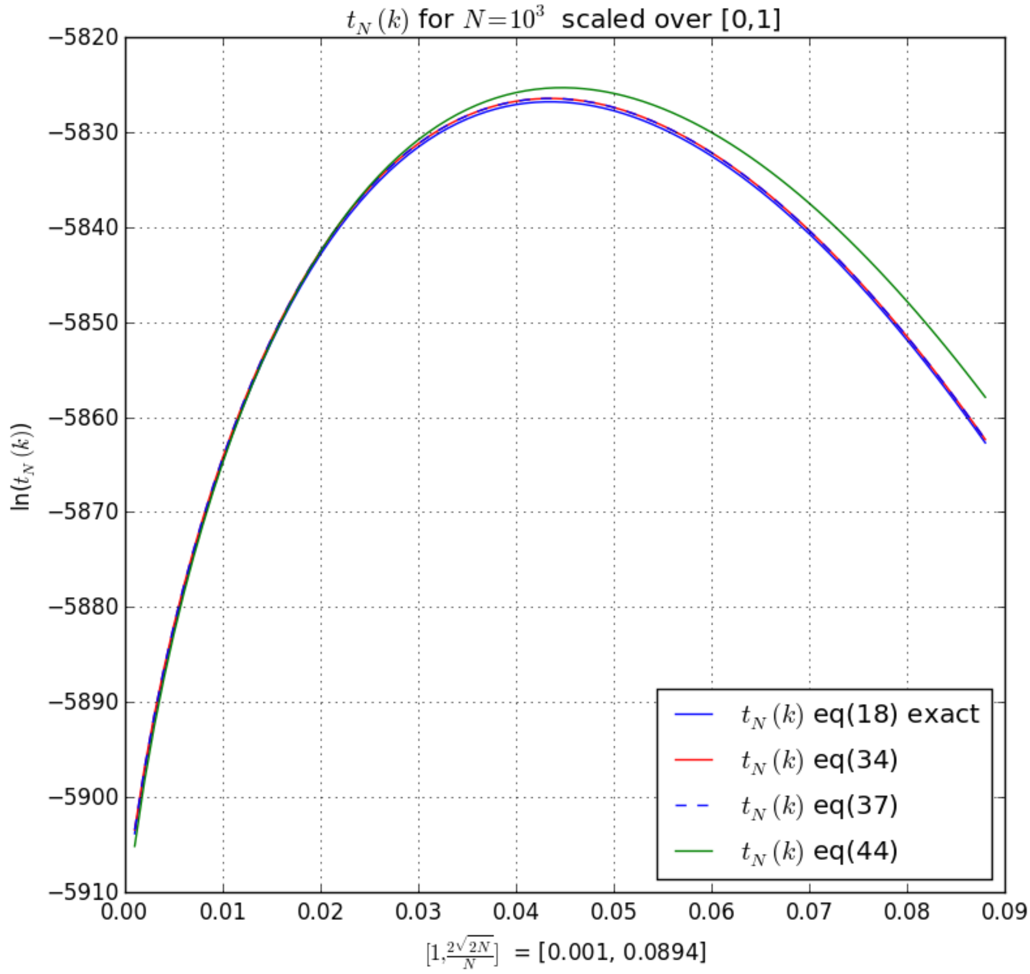


Figure A.4: Asymptotic estimates of $t_N(k)$ for $N = 2 \times 10^3$, $s = 2$, and $p = 1$.

Putting altogether, $\Omega_s(2N)$ becomes

$$\Omega_s(2N) = \frac{(2N)!}{2^{\frac{3}{2}}\pi e^{\frac{s^2}{4p}}(\frac{s^2}{2p})^{\frac{1}{4}}N^{\frac{3}{4}}(2/p)^N}\left(\frac{e}{N}\right)^N \sum_{1\leq k\leq 2\sqrt{s^2N/2p}}\left(\frac{\sqrt{\frac{s^2N}{2p}}e}{k}\right)^{2k}\left[1+O(\frac{1}{\sqrt{N}})\right]. \tag{A.72}$$

When we use the Stirling approximation of $(2N)!$, we find

$$\Omega_s(2N) = \frac{\sqrt{4\pi N}}{2^{\frac{3}{2}}\pi e^{\frac{s^2}{4p}}(\frac{s^2}{2p})^{\frac{1}{4}}N^{\frac{3}{4}}(2/p)^N}\left(\frac{2N}{e}\right)^{2N}\left(\frac{e}{N}\right)^N \sum_{1\leq k\leq 2\sqrt{s^2N/2p}}\left(\frac{\sqrt{\frac{s^2N}{2p}}e}{k}\right)^{2k}\left[1+O(\frac{1}{\sqrt{N}})\right]$$

$$= \frac{p^N}{\sqrt{2\pi}\mathrm{e}^{\frac{s^2}{4p}}(\frac{s^2N}{2p})^{\frac{1}{4}}} \left(\frac{2N}{\mathrm{e}}\right)^N \sum_{1\le k\le 2\sqrt{s^2N/2p}} \left(\frac{\sqrt{\frac{s^2N}{2p}}\mathrm{e}}{k}\right)^{2k} \left[1+O(\frac{1}{\sqrt{N}})\right]. \qquad \text{(A.73)}$$

A.3.3 Approximating the Summation

In this section, we estimate the sum in equation (A.73) to make the $\Omega_s(2N)$ asymptotic leading term more analytically tractable. Let say we are seeking the asymptotic form of $\Omega_s(2N)$ for $N \to \infty$. Hence, in this limiting case, it is possible to approximate the summation by replacing the sum with a suitable asymptomatic form. To do that, by defining $m = \sqrt{s^2N/2p}$ the sum is

$$\sum_{1\le k\le 2\sqrt{s^2N/2p}} \left(\frac{\sqrt{\frac{s^2N}{2p}}\mathrm{e}}{k}\right)^{2k} = \sum_{k=1}^{2m} \left(\frac{m\mathrm{e}}{k}\right)^{2k}, \qquad \text{(A.74)}$$

in which case, the Euler-Maclaurin formulas can address the evaluation of the transformed summation.

Theorem *(Euler-Maclaurin summation formula, first form [59]):* for a function defined on an interval $[a, b]$ with a and b as integers, and suppose that the derivatives $f^{(i)}(x)$ exist and are continuous for $1 \le i \le 2q$, where q is a fixed constant, then

$$\sum_{k=a}^{b} f(k) = \int_a^b f(x)dx + \frac{f(a)+f(b)}{2} + \sum_{i=1}^{q} \frac{B_{2i}}{(2i)!} f^{(2i-1)}(x)\bigg|_a^b + R_q \qquad \text{(A.75)}$$

where B_{2i} is the Bernoulli number and R_q is a reminder term satisfying

$$|R_q| \le \frac{|B_{2q}|}{(2q)!} \int_a^b |f^{(2q)}(x)|dx < \frac{4}{(2\pi)^{2q}} \int_a^b |f^{(2q)}(x)|dx. \qquad \text{(A.76)}$$

□

So for

$$f(x) = \left(\frac{m\mathrm{e}}{x}\right)^{2x}, \qquad \text{(A.77)}$$

we can rewrite the Euler-Maclaurin summation for $f(k)$ like,

$$\sum_{k=1}^{2m} \left(\frac{me}{k}\right)^{2k} = \int_{1}^{2m} \left(\frac{me}{x}\right)^{2x} dx + \frac{f(2m)+f(1)}{2} +$$

$$B_2 \frac{f'(2m)+f'(1)}{2} + R_q, \tag{A.78}$$

and for $x = mt$, the first integral is

$$\int_{1}^{2m} \left(\frac{me}{x}\right)^{2x} dx = m \int_{\frac{1}{m}}^{2} \left(\frac{e}{t}\right)^{2mt} dt = m \int_{\frac{1}{m}}^{2} \exp\left\{2mt - 2mt\ln(t)\right\} dt \tag{A.79}$$

Using *Saddle-Point Method* [46] allows us to approximate the integral. Suppose,

$$g(t) = 2mt - 2mt\ln(t), \tag{A.80}$$

then for first and second derivatives of $g(t)$ we have,

$$g'(t) = -2m\ln(t) = 0 \implies t_0 = 1, \tag{A.81}$$

and

$$g''(t) = -\frac{2m}{t}, \tag{A.82}$$

t_0 is where $g(t)$ is maximum. Taylor expansion of $g(t)$ around t_0 is

$$g(t) \approx 2m - m(t-1)^2. \tag{A.83}$$

The expansion approximates the integral like,

$$m \int_{\frac{1}{m}}^{2} \exp\left\{2mt - 2mt\ln(t)\right\} dt \approx m\exp(2m) \int_{\frac{1}{m}}^{2} \exp\left\{-m(t-1)^2\right\} dt. \tag{A.84}$$

A resulting integral is a Gaussian form and its width is equal to $\frac{1}{\sqrt{2m}}$. It is safe to extend the limits of the integral from $-\infty$ to $+\infty$ and approximate the integral like

$$m\exp(2m)\left(\int_{-\infty}^{\infty} \exp\left\{-m(t-1)^2\right\} dt - \epsilon_m\right)$$

$$\approx \sqrt{\frac{\pi}{m}} m \mathrm{e}^{2m} = \sqrt{\pi m} \mathrm{e}^{2m}, \tag{A.85}$$

where ϵ_m is a constant that depends only on m and exponentially small in comparison to integral.

For other terms in the Euler-Maclaurin formula, the value of $f(x)$ and its derivatives at endpoints are required. we again define $x = mt$ and then

$$f(x) \to F(t) = \left(\frac{\mathrm{e}}{t}\right)^{2mt} \tag{A.86}$$

Therefore,

$$\frac{df(x)}{dx} = \frac{1}{m}\frac{dF(t)}{dt} = \frac{1}{m}\frac{d}{dt}\left[\left(\frac{\mathrm{e}}{t}\right)^{2mt}\right] = \frac{1}{m}\left(-2mF(t)\ln(t)\right)$$

$$= -2F(t)\ln(t) \tag{A.87}$$

and

$$\frac{d^2 f(x)}{dx^2} = \frac{1}{m^2}\frac{d^2 F(t)}{dt^2} = \frac{1}{m^2}\left(2mF(t)(2m\ln^2(t) - \frac{1}{t})\right)$$

$$= 4F(t)\ln^2(t)(1 + O(\frac{1}{m})) \tag{A.88}$$

Plugging back the end points in $f(x)$ and its derivatives,

$$f(x = 2m) = F(t = 2) = \left(\frac{\mathrm{e}}{2}\right)^{4m}, \tag{A.89}$$

$$f(x = 1) = F(t = \frac{1}{m}) = (m\mathrm{e})^2, \tag{A.90}$$

$$f'(x = 2m) = F'(t = 2) = -2\left(\frac{\mathrm{e}}{2}\right)^{4m}\log(2), \tag{A.91}$$

$$f'(x = 1) = F'(t = \frac{1}{m}) = 2(m\mathrm{e})^2\log(m). \tag{A.92}$$

Meanwhile from relation (A.76),

$$|R_q| < O\left(\int_1^{2m} |F''(x)|dx\right) = O\left(\frac{1}{m}\int_{\frac{1}{m}}^{2} |F''(t)|dt\right) = O(\frac{\left(\frac{\mathrm{e}}{2}\right)^{4m}}{m}). \tag{A.93}$$

The last result comes from the fact that from relation (A.87) we have already found the solution of the integral and therefore the order of R_q.

Putting all the results together, the approximate value of the sum is

$$\sum_{k=1}^{2m} f(k) = \sqrt{\pi m}\mathrm{e}^{2m} + \left(\frac{\mathrm{e}}{2}\right)^{4m} + (m\mathrm{e})^2$$

$$-2\left(\frac{\mathrm{e}}{2}\right)^{4m}\log(2) + 2\,(m\mathrm{e})^2\log(m) + O(\frac{\left(\frac{\mathrm{e}}{2}\right)^{4m}}{m})$$

$$= \sqrt{\pi m}\mathrm{e}^{2m} + \left(\frac{\mathrm{e}}{2}\right)^{4m}(1 - 2\log(2)) + O(\frac{\left(\frac{\mathrm{e}}{2}\right)^{4m}}{m}). \tag{A.94}$$

Comparing the first and second terms on the right-hand side brings up this question: between

$$\sqrt{\pi m}\mathrm{e}^{2m} \tag{A.95}$$

and

$$\left(\frac{\mathrm{e}^2}{4}\right)^{2m} \approx (1.8472640)^{2m}, \tag{A.96}$$

which one is the leading term? We can see $m^{\frac{1}{2}}\mathrm{e}^{2m}$ grows faster than $(1.8472640)^{2m}$ (e is greater than $\mathrm{e}^2/4$ and $m^{\frac{1}{2}}$ is N dependent). Therefore,

$$\sum_{k=1}^{2m} f(k) = \sqrt{\pi m}\mathrm{e}^{2m}\left(1 + O(\frac{\left(\frac{\mathrm{e}^2}{4}\right)^{2m}}{\sqrt{m}\mathrm{e}^{2m}})\right)$$

$$= \sqrt{\pi m}\mathrm{e}^{2m}\left(1 + O(\frac{1}{\sqrt{m}(\frac{4}{\mathrm{e}})^{2m}})\right). \tag{A.97}$$

Recall that $m = \sqrt{s^2N/2p}$, hence the sum becomes

$$\sum_{k=1}^{2\sqrt{2N}} f(k) = (\frac{s^2N}{2p})^{\frac{1}{4}}\sqrt{\pi}\mathrm{e}^{2\sqrt{s^2N/2p}}\left(1 + O(\frac{1}{(\frac{s^2N}{2p})^{\frac{1}{4}}(\frac{4}{\mathrm{e}})^{2\sqrt{s^2N/2p}}})\right). \tag{A.98}$$

Indeed finding the sum's approximation transforms the relation (A.73) to

$$\Omega_s(2N) = \frac{p^N}{\sqrt{2\pi}\mathrm{e}^{\frac{s^2}{4p}}(\frac{s^2N}{2p})^{\frac{1}{4}}}\left(\frac{2N}{\mathrm{e}}\right)^N(\frac{s^2N}{2p})^{\frac{1}{4}}\sqrt{\pi}\mathrm{e}^{2\sqrt{s^2N/2p}}\left(1 + O(\frac{1}{(\frac{s^2N}{2p})^{\frac{1}{4}}(\frac{4}{\mathrm{e}})^{2\sqrt{s^2N/2p}}})\right)$$

$$\times\left[1 + O(\frac{1}{\sqrt{N}})\right]$$

$$= \frac{p^N}{\sqrt{2}\mathrm{e}^{\frac{s^2}{4p}}} \left(\frac{2N}{\mathrm{e}}\right)^N \mathrm{e}^{2\sqrt{s^2N/2p}} \left(1 + O(\frac{1}{\sqrt{N}})\right). \tag{A.99}$$

Table (**??**) list ratios between the approximate form and the exact value for $s = 2$, $p = 1$ and different Ns. As we can see, the ratio converge very slowly because of $O(\frac{1}{\sqrt{N}})$.

N	ratio	difference between two ratios
10^0	0.5829109953992809790245939305	-
10^1	0.7945985775377341167964762797	0.7945985775377341167964762797
10^2	0.9231601917908334849576657660	0.1285616142530993681611894863
10^3	0.9743185402276979494843280135	0.0511583484368644645266622475
10^4	0.9917344176841455851197054511	0.0174158774564476356353774376
10^5	0.9973715645660685975096529035	0.0056371468819230123899474524
10^6	0.9991673469451725833525477959	0.0017957823791039858428948924
10^7	0.9997365448979457992519574128	0.0005691979527732158994096169
10^8	0.9999166761297524947566204494	0.0001801312318066955046630366

The

raptionRatio between the approximate form and the exact value of different Ns.

Appendix

B

Probability Distributions

B.1 Finding $c_n(r)$ from Generating Function Method

For even numbers, the power series $f_{2n}(x)$ is defined as

$$f_{2n}(x) = \sum_{i=0}^{n} \binom{2n}{2i} (2i-1)!!\, x^i, \tag{B.1}$$

and the normalisation constant in terms of $f_{2n}(x)$ writes as

$$c_{2n,r} = r^n f_{2n}(\frac{1}{r}). \tag{B.2}$$

summing both sides of equation (B.1), we get the $F_1(Y;x)$, which is a generating function with its Y^{2n} coefficients as $f_{2n}(x)$

$$F_1(Y;x) = \sum_{2n\geq 0} f_{2n}(x)\frac{Y^{2n}}{2n!} = \sum_{2n\geq 0}\sum_{i=0}^{n} \binom{2n}{2i} (2i-1)!!\, x^i \frac{Y^{2n}}{2n!}$$

$$= \sum_{2n\geq 0}\sum_{i=0}^{n} \frac{(\frac{x}{2})^i}{i!}\frac{Y^{2n}}{(2n-2i)!}. \tag{B.3}$$

Looking at table (B.1), you see that we filled the cells that have a value with respect to i and $2n$ in the double sums above. So, it is possible to swap the sums when we start first on columns instead of rows

$$\sum_{2n\geq 0}\sum_{i=0}^{n}(.) = \sum_{i\geq 0}\sum_{2n\geq 2i}(.) \tag{B.4}$$

i	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$	$\dots$
$2n=0$	0					$\dots$
$2n=2$	0	1				$\dots$
$2n=4$	0	1	2			$\dots$
$2n=6$	0	1	2	3		$\dots$
$2n=8$	0	1	2	3	4	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Table B.1: Range of i values for a fixed $2n$.

Therefore

$$\begin{aligned} F_1(Y;x) &= \sum_{2n\geq 0}\sum_{i=0}^{n}\frac{(\frac{x}{2})^i}{i!}\frac{Y^{2n}}{(2n-2i)!} = \sum_{i\geq 0}\frac{(\frac{x}{2})^i}{i!}\sum_{2n\geq 2i}\frac{Y^{2n}}{(2n-2i)!} \\ &= \sum_{i\geq 0}\frac{(\frac{x}{2}Y^2)^i}{i!}\sum_{2n\geq 0}\frac{Y^{2n}}{2n!} = \mathrm{e}^{\frac{x}{2}Y^2}\cosh Y. \end{aligned} \tag{B.5}$$

Similarly, for odd numbers, the generating function is defined as

$$f_{2n+1}(x) = \sum_{i=0}^{n}\binom{2n+1}{2i}(2i-1)!!\, x^i, \tag{B.6}$$

and

$$\begin{aligned} F_2(Y;x) &= \sum_{2n+1\geq 1} f_{2n+1}(x)\frac{Y^{2n+1}}{(2n+1)!} = \sum_{2n+1\geq 1}\sum_{i=0}^{n}\frac{(2n+1)!}{(2n+1-2i)!}\frac{(\frac{x}{2})^i}{i!}\frac{Y^{2n+1}}{(2n+1)!} \\ &= \sum_{i\geq 0}\frac{(\frac{x}{2}Y^2)^i}{i!}\sum_{2n+1\geq 0}\frac{Y^{2n+1}}{(2n+1)!} = \mathrm{e}^{\frac{x}{2}Y^2}\sinh Y. \end{aligned} \tag{B.7}$$

Using both $F_1(Y;x)$ and $F_2(Y;x)$

$$F(Y,x) := F_1(Y;x) + F_2(Y;x) = \sum_{2n\geq 0} f_{2n}(x)\frac{Y^{2n}}{2n!} + \sum_{2n+1\geq 1} f_{2n+1}(x)\frac{Y^{2n+1}}{(2n+1)!}$$

$$= \sum_{2n\geq 0}\left[f_{2n}(x)\frac{Y^{2n}}{2n!} + f_{2n+1}(x)\frac{Y^{2n+1}}{(2n+1)!}\right] = \sum_{n\geq 0} f_n(x)\frac{Y^n}{n!}$$

$$= \mathrm{e}^{\frac{x}{2}Y^2}\left[\cosh Y + \sinh Y\right] = \mathrm{e}^{\frac{x}{2}Y^2+Y}. \tag{B.8}$$

On summing $F_1(Y;x)$ and $F_2(Y;x)$, we wrote $F(Y,x)$ as a power series with odd and even powers of Y. $F(Y,x)$ is an analytic function and is infinitely differentiable, hence, the coefficients of Taylor expansion of $F(Y,x)$ at $Y=0$ results in $f_n(x)$

$$f_n(x) = \left.\frac{d^n F(Y,x)}{dY^n}\right|_{Y=0} := F^{(n)}(0,x). \tag{B.9}$$

Since the first derivative of $F(Y,x)$ is recursively related to itself

$$F'(Y,x) = (xY+1)F(Y,x), \qquad F'(0,x) = 1, \tag{B.10}$$

repeating the derivatives gets the recursive relation for $F^{(n)}(Y,x)$

$$F^{(n+1)}(Y,x) = nxF^{(n-1)}(Y,x) + (xY+1)F^{(n)}(Y,x). \tag{B.11}$$

Table (B.2) shows the procedure.

$n=1$	$F'(Y,x) = (xY+1)F(Y,x)$
$n=2$	$F^{(2)}(Y,x) = (xY+1)F'(Y,x) + xF(Y,x)$
$n=3$	$F^{(3)}(Y,x) = (xY+1)F^{(2)}(Y,x) + 2xF'(Y,x)$
$n=4$	$F^{(4)}(Y,x) = (xY+1)F^{(3)}(Y,x) + 3xF^{(2)}(Y,x)$
$n=5$	$F^{(5)}(Y,x) = (xY+1)F^{(4)}(Y,x) + 4xF^{(3)}(Y,x)$
$\vdots$	$\vdots$
n	$F^{(n+1)}(Y,x) = (xY+1)F^{(n)}(Y,x) + nxF^{(n-1)}(Y,x)$

Table B.2: Recursive relation for $F(Y,x)$ derivatives.

$n = 0$	$F(0, x) = 1$
$n = 1$	$F'(0, x) = 1$
$n = 2$	$F^{(2)}(0, x) = x + 1$
$n = 3$	$F^{(3)}(0, x) = 3x + 1$
$n = 4$	$F^{(4)}(0, x) = 3x^2 + 6x + 1$
$n = 5$	$F^{(5)}(0, x) = 15x^2 + 10x + 1$
$\vdots$	$\vdots$
n	$F^{(n+1)}(0, x) = F^{(n)}(0, x) + nxF^{(n-1)}(0, x)$

Table B.3: Recursive relation for $F^{(n+1)}(0, x)$ derivatives.

Also in table (B.3), we show the $F^{(n+1)}(0, x)$ in terms of its previous values, and therefore

$$F^{(n+1)}(0, x) = F^{(n)}(0, x) + nxF^{(n-1)}(0, x). \tag{B.12}$$

When we plug in equation (B.9) in the last relation it finds

$$f_{n+1}(x) = f_n(x) + nxf_{n-1}(x). \tag{B.13}$$

Remember the normalisation constants relate to $f_{2n}(x)$ and $f_{2n+1}(x)$ as

$$\begin{cases} c_{2n}(r) = r^n f_{2n}(\frac{1}{r}) \\ c_{2n+1}(r) = r^n f_{2n+1}(\frac{1}{r}) \end{cases} . \tag{B.14}$$

So, it is necessary to rewrite $f_n(x)$ for odds and even numbers separately. It means equation (B.13) for odd and even numbers rewrites as

$$\begin{cases} f_{2n}(x) = f_{2n-1}(x) + (2n-1)xf_{2n-2}(x) \\ f_{2n+1}(x) = f_{2n}(x) + 2nxf_{2n-1}(x) \end{cases} . \tag{B.15}$$

Then, we have

$$\begin{cases} r^n f_{2n}(\frac{1}{r}) = r^n f_{2n-1}(\frac{1}{r}) + (2n-1)r^{n-1} f_{2n-2}(\frac{1}{r}) \\ r^n f_{2n+1}(\frac{1}{r}) = r^n f_{2n}(\frac{1}{r}) + 2nr^{n-1} f_{2n-1}(\frac{1}{r}) \end{cases} \implies$$

$$\begin{cases} c_{2n}(r) = rc_{2n-1}(r) + (2n-1)c_{2n-2}(r) \\ c_{2n+1}(r) = c_{2n}(r) + 2nc_{2n-1}(r). \end{cases} \tag{B.16}$$

B.2 Finding the LDP Limit of $P_n(n_p)$

Let us start with $P_n(n_p)$. We have

$$P_n(n_p) = \frac{1}{c_n(r)} \binom{n}{2n_p} (2n_p - 1)!!\; r^{\lfloor n/2 \rfloor - n_p}, \tag{B.17}$$

then, using Sterling's approximation for $\log n!$, we get

$$\begin{aligned} \ln P_n(n_p) = & -\ln c_n(r) + n\ln n - n - n_p \ln 2 - (n-2n_p)\ln(n-2n_p) + (n-2n_p) - n_p \ln n_p + n_p \\ & + (\lfloor n/2 \rfloor - n_p)\ln r \\ = & -\ln c_n(r) + n_p \ln n - \frac{n}{2}\left[\frac{2n_p}{n}\ln\frac{2n_p}{n} + 2(1-\frac{2n_p}{n})\ln(1-\frac{2n_p}{n})\right. \\ & \left. + \frac{2n_p}{n} - (1-\frac{2n_p}{n})\ln r\right]. \end{aligned} \tag{B.18}$$

Equation (3.57) defines $m_n = \frac{2n_p}{n}$, and the last equation becomes

$$\begin{aligned} \ln P_n(m_n) = & -\ln c_n(r) + \frac{nm_n}{2}\ln n - \frac{n}{2}\left[m_n \ln m_n + 2(1-m_n)\ln(1-m_n)\right. \\ & \left. + m_n - (1-m_n)\ln r\right]. \end{aligned} \tag{B.19}$$

To find $\log c_n(r)$ for $1 \ll n$, its asymptotic leading term in equation (3.53) gives

$$\ln c_n(r) = \ln\left(\frac{\mathrm{e}^{\frac{-r}{4}}}{\sqrt{2}}\left(\frac{n}{\mathrm{e}}\right)^{n/2}\mathrm{e}^{\sqrt{rn}}\right) = \frac{n}{2}\ln n - \frac{n}{2} + O(\sqrt{n}). \tag{B.20}$$

Hence

$$\begin{aligned} \ln P_n(m_n) = & -\frac{n}{2}\ln n + \frac{n}{2} + \frac{nm_n}{2}\ln n - \frac{n}{2}\left[m_n \ln m_n + 2(1-m_n)\ln(1-m_n)\right. \\ & \left. + m_n - (1-m_n)\ln r\right] + O(\sqrt{n}) \\ = & -(\frac{1-m_n}{2})n\ln n - \frac{n}{2}\left[m_n \ln m_n + (1-m_n)\ln\frac{(1-m_n)^2}{\mathrm{e}r}\right] + O(\sqrt{n}). \end{aligned} \tag{B.21}$$

B.3 Finding the LDP Limit of $P_n(n_p, n_u)$

The $P_n(n_p, n_u)$ distribution is written as

$$P_n(n_p, n_u) = \frac{1}{c_n(r)} \binom{n}{2n_p} (2n_p - 1)!!\, r^{\lfloor n/2 \rfloor - n_p} \binom{n - 2n_p}{n_u} \rho^{n_u} (1-\rho)^{n-2n_p-n_u}. \quad \text{(B.22)}$$

Let us find the logarithm of $\binom{2-2n_p}{n_u}\rho^{n_u}(1-\rho)^{n-2n_p-n_u}$ terms, which are extra in comparison to $P_n(n_p)$, and then, we can use the results from the previous section to check the large deviation principle.

Recall that $m_n = 2n_p/n$ and $s_n = n_u/(n-2n_p)$. Thus, we have

$$\ln\left[\binom{n-2n_p}{n_u}\rho^{n_u}(1-\rho)^{n-2n_p-n_u}\right] = \ln(n-2n_p)! - \ln n_u! - \ln(n - 2n_p - n_u)!$$

$$= -n(1 - \frac{2n_p}{n})\left[\frac{n_u}{(n-2n_p)}\ln(\frac{\frac{n_u}{(n-2n_p)}}{\rho}) + (1 - \frac{n_u}{n-2n_p})\ln(\frac{1-\frac{n_u}{n-2n_p}}{1-\rho})\right]$$

$$= -n(1-m_n)\left[s_n \ln(\frac{s_n}{\rho}) + (1-s_n)\ln(\frac{1-s_n}{1-\rho})\right]. \quad \text{(B.23)}$$

Combining the last result with equation (3.62) completes the remaining terms

$$\ln P_n(m_n, s_n) = -(\frac{1-m_n}{2})n\ln n + \frac{n}{2}\left[m_n \ln m_n + (1-m_n)\ln\frac{(1-m_n)^2}{er}\right.$$

$$\left. -2(1-m_n)\left(s_n\ln(\frac{s_n}{\rho}) + (1-s_n)\ln(\frac{1-s_n}{1-\rho})\right)\right] + O(\sqrt{n})$$

$$= -(\frac{1-m_n}{2})n\ln n + \frac{n}{2}\left[\tilde{H}_r(m_n)\right.$$

$$\left. -2(1-m_n)\left(s_n\ln(\frac{s_n}{\rho}) + (1-s_n)\ln(\frac{1-s_n}{1-\rho})\right)\right] + O(\sqrt{n}). \quad \text{(B.24)}$$

B.4 Asymptotic Leading Term of $c_n(\epsilon)$

It was mentioned in the text that equation (B.20) is a valid asymptotic leading term when r is kept constant. However, one special case happens when we assume r is increasing with $2n$ such that

$$\lim_{2n\to\infty} \frac{r}{2n} = \epsilon, \quad \text{(B.25)}$$

and the asymptotic term of $\log(c_{2n,\epsilon})$ needs considering this limit. Here, we start from equation (3.51)

$$c_{2n,r} = (2n)! \sum_{k=0}^{n} \frac{r^k 2^{k-n}}{(n-k)!(2k)!},$$

and replace r by $2\epsilon n$

$$\begin{aligned} c_{2n,\epsilon} &= (2n)! \sum_{k=0}^{n} \frac{(2\epsilon n)^k 2^{k-n}}{(n-k)!(2k)!} \\ &= \frac{(2n)!}{2^n} \sum_{k=0}^{n} \frac{(4\epsilon n)^k}{(n-k)!(2k)!}. \end{aligned} \tag{B.26}$$

Without rearranging the order of k (check equation (3.51) to see how the order changes), k^* or where the summand is maximum, approaches n in the limit $2n \to \infty$. Yet, in the above form, k^* is ϵ dependent.

Denoting

$$t_{2n}(k) = \frac{(4\epsilon n)^k}{(n-k)!(2k)!} \tag{B.27}$$

for the summand, and using Sterling's approximation $\log n! = n \log n - n$, we find the $k^*(\epsilon)$ by taking the derivative[1] of the logarithm of $t_{2n}(k)$, since the logarithm is a strictly increasing function, so the maximum of $t_{2n}(k)$ coincides with $\log t_{2n}(k)$

$$\log t_{2n}(k) = k \log(4\epsilon n) - 2k \log(2k) + 2k - (n-k)\log(n-k) + (n-k) \implies$$

$$\begin{aligned} \frac{d \log t_{2n}(k)}{dk} &= \log(4\epsilon n) - 2 - 2\log(2k) + 2 + 1 + \log(n-k) - 1 \\ &= \log(4\epsilon n) + \log \frac{n-k}{4k^2} = 0 \implies \\ & k^2 + \epsilon n k - \epsilon n^2 = 0. \end{aligned} \tag{B.28}$$

The solutions of the above quadratic equation are

$$k = -\frac{\epsilon n}{2} \pm \frac{\epsilon n}{2} \sqrt{1 + \frac{4}{\epsilon}}, \tag{B.29}$$

[1] This is true that we treat a discrete variable, k, as a continuous one. However, in the continuum limit, $2n \to \infty$, the distinction disappears.

We can also use the ratio

$$\frac{t_{2n}(k+1)}{t_{2n}(k)} = 1,$$

to find the maximum where the difference between two consecutive terms is negligible. Doing that, we find the same solution for k^*, round to closest integer.

however, k is non-negative, and therefore, the maximum of $t_{2n}(k)$ is at

$$k^* = \frac{\epsilon n}{2}\left[\sqrt{1+\frac{4}{\epsilon}} - 1\right]. \tag{B.30}$$

To not clutter the notation, we denote the bracket as

$$f(\epsilon) = \sqrt{1+\frac{4}{\epsilon}} - 1, \tag{B.31}$$

and write

$$k^* = \frac{\epsilon n f(\epsilon)}{2}. \tag{B.32}$$

Now that we found the k^*, and knowing the fact that the sum in $c_{2n,\epsilon}$ is concentrated at its maximum, we find the asymptotic leading term of $t_{2n}(k)$ by first simplifying it using the Sterling's approximation, $n! = \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n$, and next evaluating it at k^*

$$\begin{aligned} t_{2n}(k) = \frac{(4\epsilon n)^k}{(n-k)!(2k)!} &= \frac{(4\epsilon n)^k \mathrm{e}^{n+k}}{2\pi\sqrt{2k(n-k)}(2k)^{2k}(n-k)^{n-k}} \\ &= \frac{(\frac{\mathrm{e}\epsilon n}{k^2})^k \mathrm{e}^n}{2\pi\sqrt{2k(n-k)}n^{n-k}(1-\frac{k}{n})^{n-k}} \\ &= \frac{1}{2\sqrt{2\pi}} \times \frac{1}{\sqrt{k(n-k)}} \times \frac{1}{(1-\frac{k}{n})^{n-k}} \times (\frac{\mathrm{e}\epsilon n^2}{k^2})^k \left(\frac{\mathrm{e}}{n}\right)^n. \end{aligned} \tag{B.33}$$

The n-dependence of k^* allows us to exploit the same argument that we used to find $\Omega(n)$ asymptotic leading term. For instance, the first ratio $1/\sqrt{k(n-k)}$ approximates as

$$\frac{1}{\sqrt{k^*(n-k^*)}} \sim \frac{1}{n\sqrt{\frac{\epsilon f(\epsilon)}{2}(1-\frac{\epsilon f(\epsilon)}{2})}}, \tag{B.34}$$

considering the fact that the bulk of the distribution is concentrated around the k^*. Let us define

$$g(\epsilon) = (1-\frac{\epsilon f(\epsilon)}{2}). \tag{B.35}$$

Then

$$\frac{1}{\sqrt{k(n-k)}} \sim \frac{1}{n\sqrt{\frac{\epsilon}{2}f(\epsilon)g(\epsilon)}}, \tag{B.36}$$

and

$$(1-\frac{k}{n})^{n-k} \sim (1-\frac{\epsilon f(\epsilon)}{2})^{n(1-\frac{\epsilon f(\epsilon)}{2})} = g(\epsilon)^{ng(\epsilon)}. \tag{B.37}$$

Therefore, the summand turns to

$$t_{2n}(k) \sim \frac{1}{2\sqrt{2}\pi} \times \frac{1}{n\sqrt{\frac{\epsilon}{2}f(\epsilon)g(\epsilon)}} \times g(\epsilon)^{-ng(\epsilon)} \times (\frac{\mathrm{e}\epsilon n^2}{k^2})^k \left(\frac{\mathrm{e}}{n}\right)^n, \tag{B.38}$$

and $c_{2n,\epsilon}$ becomes

$$c_{2n,\epsilon} \sim \frac{(2n)!}{2^{n+1}\pi} \frac{g(\epsilon)^{-ng(\epsilon)} \left(\frac{\mathrm{e}}{n}\right)^n}{n\sqrt{\epsilon f(\epsilon)g(\epsilon)}} \sum_{k=1}^{n} (\frac{\mathrm{e}\epsilon n^2}{k^2})^k. \tag{B.39}$$

We have to emphasise that we started the sum from $k = 1$ instead of $k = 0$. To justify it, observe that using the equation (B.26), we find the ratio of two consecutive summands for $k = 0$ and $k = 1$

$$\frac{t_{2n}(0)}{t_{2n}(1)} = \frac{1}{4\epsilon n} \times \frac{(n-1)!}{n!} \times \frac{0!}{2!} = \frac{1}{8\epsilon n^2}, \tag{B.40}$$

which means asymptotically $t_{2n}(0) \ll t_{2n}(1)$ for any $\epsilon > 0$ and $n \to \infty$, which justify the exclusion of $k = 0$. It will become clear later that this removal gives a mathematically nicer form to handle.

If we approximate the above $(2n)!$ too, we get

$$\begin{aligned} c_{2n,\epsilon} &\sim \frac{\sqrt{4\pi n}(\frac{2n}{\mathrm{e}})^{2n}\left(\frac{\mathrm{e}}{n}\right)^n}{2^{n+1}\pi} \frac{g(\epsilon)^{-ng(\epsilon)}}{n\sqrt{\epsilon f(\epsilon)g(\epsilon)}} \sum_{k=1}^{n} (\frac{\mathrm{e}\epsilon n^2}{k^2})^k \\ &= \frac{(\frac{2n}{\mathrm{e}})^n g(\epsilon)^{-ng(\epsilon)}}{\sqrt{n\pi\epsilon f(\epsilon)g(\epsilon)}} \sum_{k=1}^{n} (\frac{\mathrm{e}\epsilon n^2}{k^2})^k. \end{aligned} \tag{B.41}$$

Finally, we need to estimate the asymptotic leading term of the sum $\sum_{k=0}^{n}(\frac{\sqrt{\mathrm{e}\epsilon}n}{k})^{2k}$. Resort back to the result for $\Omega(2n)$ asymptotic leading term, the sum can be estimated as an integral. Still, the free parameter ϵ introduces two different behaviours that introduce a subtle consideration in estimating the current sum.

Let us call the summand as

$$d(k) = (\frac{\sqrt{\mathrm{e}\epsilon}n}{k})^{2k}. \tag{B.42}$$

Similar to what we have already done to find k^*, we treat $d(k)$ as a continuous function and find its maximum w.r.t. ϵ for constant n by taking the derivative of $\log d(k)$

$$\frac{d\log d(k)}{dk} = 2\log(\frac{\sqrt{\mathrm{e}\epsilon}n}{k}) - 2 = 0 \implies$$

$$\frac{\sqrt{\mathrm{e}\epsilon n}}{k} = \mathrm{e} \implies$$
$$\frac{k^*}{n} = \sqrt{\frac{\epsilon}{\mathrm{e}}}. \tag{B.43}$$

However, $k^* \leq n$, and for $\epsilon > \mathrm{e}$ the above result implies

$$k^* = n. \tag{B.44}$$

In other words, the maximum of the sum is concentrated at $k^* = n$. The $\epsilon > \mathrm{e}$ bound is what we need to consider when we seek to find the asymptotic leading term of the sum.

Continuing the estimation by an integral, we divide our study into two regions

- $\epsilon > \mathrm{e}$: In this case, the sum can be estimated by its largest term which is $t_{2n}(n)$
$$\sum_{k=1}^{n} (\frac{\mathrm{e}\epsilon n^2}{k^2})^k \sim (\frac{\mathrm{e}\epsilon n^2}{n^2})^n$$
$$\sum_{k=1}^{n} (\frac{\mathrm{e}\epsilon n^2}{k^2})^k \sim (\frac{\mathrm{e}\epsilon n^2}{n^2})^n = (\mathrm{e}\epsilon)^n \,. \tag{B.45}$$
Later, we shall show that the numerical analysis supports the above result, except for $\epsilon \to \mathrm{e}^+$.

- $\epsilon \leq \mathrm{e}$: We have
$$\sum_{k=1}^{n} (\frac{\mathrm{e}\epsilon n^2}{k^2})^k \sim \int_1^n \left(\frac{\sqrt{\mathrm{e}\epsilon n}}{x}\right)^{2x} dx$$
$$= \sqrt{\mathrm{e}\epsilon}\, n \int_{\frac{1}{\sqrt{\mathrm{e}\epsilon n}}}^{\frac{1}{\sqrt{\mathrm{e}\epsilon}}} x^{-2\sqrt{\mathrm{e}\epsilon}\, nx} dx \qquad (x \to \sqrt{\mathrm{e}\epsilon}\, nx)$$
$$= \sqrt{\mathrm{e}\epsilon}\, n \int_{\frac{1}{\sqrt{\mathrm{e}\epsilon n}}}^{\frac{1}{\sqrt{\mathrm{e}\epsilon}}} \mathrm{e}^{-n\left[2\sqrt{\mathrm{e}\epsilon}\, x \ln x\right]} dx. \tag{B.46}$$
We are going to use the steepest descent approximation here. Define
$$h(x) = 2\sqrt{\mathrm{e}\epsilon}\, x \ln x, \tag{B.47}$$

and see that $h'(x^*) = 0$ gives

$$x^* = \mathrm{e}^{-1}. \tag{B.48}$$

So, the Taylor expansion of $h(x)$ around x^*, up and including the quadratic term, is

$$\begin{aligned} h(x) &= h(x^*) + \frac{h''(x^*)}{2}(x - x^*)^2 + O(x^3) \\ &= -2\sqrt{\frac{\epsilon}{\mathrm{e}}} + \sqrt{\mathrm{e}^3\epsilon}(x - \mathrm{e}^{-1})^2 + O(x^3). \end{aligned} \tag{B.49}$$

Then, defining $a^2 := \sqrt{\mathrm{e}^3\epsilon}$, the integral estimate gives

$$\begin{aligned} \sum_{k=1}^{n}(\frac{\mathrm{e}\epsilon n^2}{k^2})^k &\sim \sqrt{\mathrm{e}\epsilon}\; n \int\limits_{\frac{1}{\sqrt{\mathrm{e}\epsilon}n}}^{\frac{1}{\sqrt{\mathrm{e}\epsilon}}} \mathrm{e}^{-n\left[-2\sqrt{\frac{\epsilon}{\mathrm{e}}}+a^2(x-\mathrm{e}^{-1})^2\right]}dx \\ &= \sqrt{\mathrm{e}\epsilon}\; n\mathrm{e}^{2n\sqrt{\frac{\epsilon}{\mathrm{e}}}} \int\limits_{\frac{1}{\sqrt{\mathrm{e}\epsilon}n}}^{\frac{1}{\sqrt{\mathrm{e}\epsilon}}} \mathrm{e}^{-a^2n(x-\mathrm{e}^{-1})^2}dx \\ &= \frac{\sqrt{\mathrm{e}\epsilon}}{a\sqrt{n}}\; n\mathrm{e}^{2n\sqrt{\frac{\epsilon}{\mathrm{e}}}} \int\limits_{a\sqrt{n}(\frac{1}{\sqrt{\mathrm{e}\epsilon}n}-\mathrm{e}^{-1})}^{a\sqrt{\frac{n}{\mathrm{e}}}(\frac{1}{\sqrt{\epsilon}}-\frac{1}{\sqrt{\mathrm{e}}})} \mathrm{e}^{-t^2}dt \qquad (t = a\sqrt{n}(x - \mathrm{e}^{-1})) \\ &= \left(\frac{\epsilon}{\mathrm{e}}\right)^{\frac{1}{4}} \sqrt{n}\; \mathrm{e}^{2n\sqrt{\frac{\epsilon}{\mathrm{e}}}} \int\limits_{-\left[\left(\frac{\epsilon}{\mathrm{e}}\right)^{\frac{1}{4}}-\left(\frac{\mathrm{e}}{\epsilon}\right)^{\frac{1}{4}}\frac{1}{n}\right]\sqrt{n}}^{(\mathrm{e}\epsilon)^{\frac{1}{4}}\sqrt{n}(\frac{1}{\sqrt{\epsilon}}-\frac{1}{\sqrt{\mathrm{e}}})} \mathrm{e}^{-t^2}dt \qquad (a = (\mathrm{e}^3\epsilon)^{\frac{1}{4}}) \\ &= \left(\frac{\epsilon}{\mathrm{e}}\right)^{\frac{1}{4}} \sqrt{n}\; \mathrm{e}^{2n\sqrt{\frac{\epsilon}{\mathrm{e}}}} \int\limits_{-\left[\left(\frac{\epsilon}{\mathrm{e}}\right)^{\frac{1}{4}}-\left(\frac{\mathrm{e}}{\epsilon}\right)^{\frac{1}{4}}\frac{1}{n}\right]\sqrt{n}}^{\left[\left(\frac{\mathrm{e}}{\epsilon}\right)^{\frac{1}{4}}-\left(\frac{\epsilon}{\mathrm{e}}\right)^{\frac{1}{4}}\right]\sqrt{n}} \mathrm{e}^{-t^2}dt. \end{aligned} \tag{B.50}$$

For $n \to \infty$ and considering the condition $\epsilon \leq \mathrm{e}$, the limits of the integral become

$$\lim_{n\to\infty} -\left[\left(\frac{\epsilon}{\mathrm{e}}\right)^{\frac{1}{4}} - \left(\frac{\mathrm{e}}{\epsilon}\right)^{\frac{1}{4}}\frac{1}{n}\right]\sqrt{n} \to -\infty, \tag{B.51}$$

and

$$\lim_{n\to\infty} \left[\left(\frac{\mathrm{e}}{\epsilon}\right)^{\frac{1}{4}} - \left(\frac{\epsilon}{\mathrm{e}}\right)^{\frac{1}{4}}\right]\sqrt{n} \to \infty. \tag{B.52}$$

Then

$$\sum_{k=1}^{n}(\frac{\mathrm{e}\epsilon n^2}{k^2})^k \sim \left(\frac{\epsilon}{\mathrm{e}}\right)^{\frac{1}{4}} \sqrt{n}\,\mathrm{e}^{2n\sqrt{\frac{\epsilon}{\mathrm{e}}}} \int_{-\infty}^{\infty} \mathrm{e}^{-t^2} dt = \left(\frac{\epsilon}{\mathrm{e}}\right)^{\frac{1}{4}} \sqrt{\pi n}\,\mathrm{e}^{2n\sqrt{\frac{\epsilon}{\mathrm{e}}}}. \tag{B.53}$$

So, in both cases

$$\sum_{k=0}^{n}(\frac{\mathrm{e}\epsilon n^2}{k^2})^k \sim \begin{cases} \left(\frac{\epsilon}{\mathrm{e}}\right)^{\frac{1}{4}} \sqrt{\pi n}\,\mathrm{e}^{2n\sqrt{\frac{\epsilon}{\mathrm{e}}}} & \epsilon \leq \mathrm{e} \\ (\epsilon \mathrm{e})^n & \epsilon > \mathrm{e} \end{cases} \tag{B.54}$$

Finally, the $c_{2n,\epsilon}$ asymptotic leading term is

$$c_{2n,\epsilon} \sim \frac{(\frac{2n}{\mathrm{e}})^n g(\epsilon)^{-ng(\epsilon)}}{\sqrt{n\pi\epsilon f(\epsilon)g(\epsilon)}} \times \begin{cases} \left(\frac{\epsilon}{\mathrm{e}}\right)^{\frac{1}{4}} \sqrt{\pi n}\,\mathrm{e}^{2n\sqrt{\frac{\epsilon}{\mathrm{e}}}} & \epsilon \leq \mathrm{e} \\ (\epsilon \mathrm{e})^n & \epsilon > \mathrm{e} \end{cases}$$

$$= \begin{cases} \frac{g(\epsilon)^{-ng(\epsilon)} \mathrm{e}^{2n\sqrt{\frac{\epsilon}{\mathrm{e}}}} (\frac{2n}{\mathrm{e}})^n}{\sqrt{\sqrt{\mathrm{e}\epsilon} f(\epsilon) g(\epsilon)}} & \epsilon \leq \mathrm{e} \\ \frac{g(\epsilon)^{-ng(\epsilon)} (2\epsilon n)^n}{\sqrt{n\pi\epsilon f(\epsilon) g(\epsilon)}} & \epsilon > \mathrm{e} \end{cases}. \tag{B.55}$$

B.5 First Moment

The first moment of $P_n(n_p)$ is defined as

$$\langle n_p \rangle_n = \sum_{i=0}^{\lfloor n/2 \rfloor} i P_n(n_p = i) = \frac{1}{c_n(r)} \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{2i} (2i-1)!!\; i\; r^{\lfloor n/2 \rfloor - i}. \tag{B.56}$$

From equation (B.1) for the generating function $f_n(x)$ we have

$$f_n(x) = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{2i} (2i-1)!!\; x^i \implies$$

$$\left[x \frac{d}{dx}\right] f_n(x) = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{2i} (2i-1)!!\; i\; x^i. \tag{B.57}$$

So, combining it with equation (B.14) derives

$$\langle n_p \rangle_n = \frac{r^{\lfloor n/2 \rfloor}}{c_n(r)} \left[x \frac{d}{dx} \right] f_n(x) \Big|_{x=\frac{1}{r}} . \tag{B.58}$$

Furthermore for even numbers, say $2n$, equation (B.9) implies

$$f_{2n}(x) := \frac{d^{2n} F(Y,x)}{dY^{2n}} \Big|_{Y=0} \implies$$

$$\frac{df_{2n}(x)}{dx} = \frac{d}{dx} \frac{d^{2n} F(Y,x)}{dY^{2n}} \Big|_{Y=0} = \frac{d^{2n}}{dY^{2n}} \frac{dF(Y,x)}{dx} \Big|_{Y=0}$$
$$= \frac{d^{2n}}{dY^{2n}} \frac{d}{dx} e^{\frac{xY^2}{2}+Y} \Big|_{Y=0} = \frac{d^{2n}}{dY^{2n}} \left(\frac{Y^2}{2} F(Y,x) \right) \Big|_{Y=0} . \tag{B.59}$$

At this stage to make the notation less cluttered denote $F(Y,x)$ by F and take derivatives with respect to Y repeatedly as follows

$$\frac{d}{dY} \left(\frac{Y^2}{2} F \right) = YF + \frac{Y^2}{2} F',$$

$$\frac{d^2}{dY^2} \left(\frac{Y^2}{2} F \right) = F + 2YF' + \frac{Y^2}{2} F'',$$
$$\frac{d^3}{dY^3} \left(\frac{Y^2}{2} F \right) = 3F' + 3YF'' + \frac{Y^2}{2} F^{(3)},$$
$$\frac{d^4}{dY^4} \left(\frac{Y^2}{2} F \right) = 6F'' + 4YF^{(3)} + \frac{Y^2}{2} F^{(4)},$$
$$\frac{d^5}{dY^5} \left(\frac{Y^2}{2} F \right) = 10F^{(3)} + 5YF^{(4)} + \frac{Y^2}{2} F^{(5)},$$
$$\frac{d^6}{dY^6} \left(\frac{Y^2}{2} F \right) = 15F^{(4)} + 6YF^{(5)} + \frac{Y^2}{2} F^{(6)},$$
$$\ldots,$$
$$\frac{d^{2n}}{dY^{2n}} \left(\frac{Y^2}{2} F \right) = n(2n-1)F^{(2n-2)} + 2nYF^{(2n-1)} + \frac{Y^2}{2} F^{(2n)}, \tag{B.60}$$

which implies

$$\frac{df_{2n}(x)}{dx} = \frac{d^{2n}}{dY^{2n}} \left(\frac{Y^2}{2} F(Y,x) \right) \Big|_{Y=0} = n(2n-1)F^{(2n-2)}(0,x)$$
$$= n(2n-1) f_{2n-2}(x). \tag{B.61}$$

For odd numbers we find

$$\frac{df_{2n+1}(x)}{dx} = n(2n+1)f_{2n-1}(x). \tag{B.62}$$

Therefore, in general for n we obtain

$$\frac{df_n(x)}{dx} = \frac{n(n-1)}{2} f_{n-2}(x). \tag{B.63}$$

We can say the derivative transform a polynomial $f_n(x)$ to $f_{n-2}(x)$ times $n(n-1)/2$. Using the last result and equations (B.58) and (B.14)

$$\langle n_p \rangle_n = \frac{r^{\lfloor n/2 \rfloor}}{c_n(r)} \left[x\frac{d}{dx}\right] f_n(x)\bigg|_{x=\frac{1}{r}} = \frac{r^{\lfloor n/2 \rfloor}}{c_n(r)} \frac{n(n-1)}{2} x f_{n-2}(x)\bigg|_{x=\frac{1}{r}}$$

$$= \frac{r^{\lfloor n/2 \rfloor -1}}{c_n(r)} \frac{n(n-1)}{2} f_{n-2}(\frac{1}{r}) \implies$$

$$\langle n_p \rangle_n = \frac{n(n-1)}{2} \frac{c_{n-2}(r)}{c_n(r)}. \tag{B.64}$$

B.6 The operator $\left[x\frac{d}{dx}\right]$

In order to find a pattern, let us apply k times the operator $\left[x\frac{d}{dx}\right]$ on $f_n(x)$ consecutively

$$\left[x\frac{d}{dx}\right] f_n(x) = \frac{n(n-1)}{2} x f_{n-2}(x) \implies$$

$$\left[x\frac{d}{dx}\right]^2 f_n(x) = \frac{n(n-1)}{2} x f_{n-2}(x) + \frac{n(n-1)(n-2)(n-3)}{4} x^2 f_{n-4}(x)$$

$$= \frac{n!}{(n-2)!}\frac{x}{2} f_{n-2}(x) + \frac{n!}{(n-4)!}\left(\frac{x}{2}\right)^2 f_{n-4}(x),$$

$$\left[x\frac{d}{dx}\right]^3 f_n(x) = \frac{n!}{(n-2)!}\frac{x}{2} f_{n-2}(x) + \frac{3n!}{(n-4)!}\left(\frac{x}{2}\right)^2 f_{n-4}(x) + \frac{n!}{(n-6)!}\left(\frac{x}{2}\right)^3 f_{n-6}(x),$$

$$\left[x\frac{d}{dx}\right]^4 f_n(x) = \frac{n!}{(n-2)!}\frac{x}{2} f_{n-2}(x) + \frac{7n!}{(n-4)!}\left(\frac{x}{2}\right)^2 f_{n-4}(x) + \frac{6n!}{(n-6)!}\left(\frac{x}{2}\right)^3 f_{n-6}(x)$$

$$+ \frac{n!}{(n-8)!}\left(\frac{x}{2}\right)^4 f_{n-8}(x),$$

$$\left[x\frac{d}{dx}\right]^5 f_n(x) = \frac{n!}{(n-2)!}\frac{x}{2}f_{n-2}(x) + \frac{15n!}{(n-4)!}\left(\frac{x}{2}\right)^2 f_{n-4}(x) + \frac{25n!}{(n-6)!}\left(\frac{x}{2}\right)^3 f_{n-6}(x)$$
$$+\frac{10n!}{(n-8)!}\left(\frac{x}{2}\right)^4 f_{n-8}(x) + \frac{n!}{(n-10)!}\left(\frac{x}{2}\right)^5 f_{n-10}(x),$$
$$\cdots$$
$$\left[x\frac{d}{dx}\right]^k f_n(x) = \sum_{i=1}^{k}\frac{n!}{(n-2i)!}a_i^{(k)}\left(\frac{x}{2}\right)^i f_{n-2i}(x), \tag{B.65}$$

where the recursive relation for $a_i^{(k)}$s is

$$a_i^{(k)} = a_{i-1}^{(k-1)} + i a_i^{(k-1)}, \qquad a_1^{(k)} = a_k^{(k)} = 1. \tag{B.66}$$

Using equation (B.14) we write

$$\begin{cases} r^{-i} f_{2n-2i}(\frac{1}{r}) = \frac{c_{2n-2i}}{r^n} \\ r^{-i} f_{2n+1-2i}(\frac{1}{r}) = \frac{c_{2n+1-2i}}{r^n} \end{cases} \Longrightarrow r^{-i} f_{n-2i}(\frac{1}{r}) = \frac{c_{n-2i}}{r^{\lfloor n/2\rfloor}}, \tag{B.67}$$

and then

$$\left[x\frac{d}{dx}\right]^k f_n(x)\Bigg|_{x=\frac{1}{r}} = \sum_{i=1}^{k}\frac{n!}{(n-2i)!}a_i^{(k)}\left(\frac{x}{2}\right)^i f_{n-2i}(x)\Bigg|_{x=\frac{1}{r}} = \sum_{i=1}^{k}\frac{n!}{(n-2i)!}\frac{a_i^{(k)}}{2^i}\frac{c_{n-2i}}{r^{\lfloor n/2\rfloor}}$$
$$= \sum_{i=1}^{k}\frac{b_i^{(k)}}{2^i}\frac{c_{n-2i}}{r^{\lfloor n/2\rfloor}}, \tag{B.68}$$

where

$$b_i^{(k)} \equiv \frac{n!}{(n-2i)!}a_i^{(k)}. \tag{B.69}$$

B.7 Finding $dc_n(r)/dr$ and $d\ln c_n(r)/dr$

To find $\frac{dc_{2n}(r)}{dr}$, we use equations (B.14) and (B.61)

$$\frac{dc_{2n}(r)}{dr} = \frac{d}{dr}\left[r^n f_{2n}(\frac{1}{r})\right] = nr^{n-1}f_{2n}(\frac{1}{r}) + r^n\frac{df_{2n}(\frac{1}{r})}{dr}$$
$$= nr^{n-1}f_{2n}(\frac{1}{r}) - r^{n-2}\frac{df_{2n}(\frac{1}{r})}{d(\frac{1}{r})}$$

$$= nr^{n-1} f_{2n}(\frac{1}{r}) - n(2n-1)r^{n-2} f_{2n-2}(\frac{1}{r}), \tag{B.70}$$

where we used $f'_{2n}(x) = n(2n-1)f_{2n-2}(x)$ in the last step. Using equation (B.14) by which

$$c_{2n}(r) = r^n f_{2n}(\frac{1}{r}), \qquad c_{2n-2}(r) = r^{n-1} f_{2n-2}(\frac{1}{r}), \tag{B.71}$$

hence equation (B.70) rewrites as

$$\frac{dc_{2n}(r)}{dr} = \frac{n}{r} c_{2n}(r) - \frac{n(2n-1)}{r} c_{2n-2}(r), \tag{B.72}$$

Similarly, for odd numbers we find

$$\frac{dc_{2n+1}(r)}{dr} = nr^{n-1} f_{2n+1}(\frac{1}{r}) - n(2n+1)r^{n-2} f_{2n-1}(\frac{1}{r})$$

$$= \frac{n}{r} c_{2n+1}(r) - \frac{n(2n+1)}{r} c_{2n-1}(r). \tag{B.73}$$

Hence, in general, we have

$$\frac{dc_n(r)}{dr} = \frac{\lfloor n/2 \rfloor}{r} c_n(r) - \frac{n(n-1)}{2r} c_{n-2}(r). \tag{B.74}$$

Also, since

$$\frac{d \ln c_n(r)}{dr} = \frac{1}{c_n(r)} \frac{dc_n(r)}{dr}, \tag{B.75}$$

we have

$$\frac{d \ln c_n(r)}{dr} = \frac{\lfloor n/2 \rfloor}{r} - \frac{n(n-1)}{2r} \frac{c_{n-2}(r)}{c_n(r)} = \frac{\lfloor n/2 \rfloor - \langle n_p \rangle_n}{r}, \tag{B.76}$$

where we used equation (B.64) in the last step.

For even numbers, taking the derivative directly on the definition of $c_{2n}(r)$ obtains another result. We have

$$\frac{dc_{2n}(r)}{dr} = \frac{d}{dr} \sum_{i=0}^{n} \binom{2n}{2i} (2i-1)!! r^{n-i}$$

$$= \sum_{i=0}^{n-1} (n-i) \binom{2n}{2i} (2i-1)!! r^{n-i-1}$$

$$= \sum_{i=0}^{n-1} n \binom{2n-1}{2i} (2i-1)!! r^{n-i-1} = nc_{2n-1}(r). \tag{B.77}$$

B.8 Finding kth Moments in terms of Other First Moments

To find the relation between kth moment and the first moments of smaller system sizes, if we use the definition of $b_i^{(k)}$ in equation (3.106), we observe

$$\frac{b_i^{(k)}}{2^i}\frac{c_{n-2i}(r)}{c_n(r)} = \frac{n!}{(n-i)!}\frac{a_i^{(k)}}{2^i}\frac{c_{n-2i}(r)}{c_n(r)}$$

$$= \left[\frac{n(n-1)}{2}\frac{c_{n-2}(r)}{c_n(r)}\right] \times \left[\frac{(n-2)(n-3)}{2}\frac{c_{n-4}(r)}{c_{n-2}(r)}\right] \times \left[\frac{(n-4)(n-5)}{2}\frac{c_{n-6}(r)}{c_{n-4}(r)}\right] \times \dots$$
$$\times \left[\frac{(n-2i+2)(n-2i+1)}{2}\frac{c_{n-2i}(r)}{c_{n-2i+2}(r)}\right] a_i^{(k)}$$
$$= \langle n_p\rangle_n \langle n_p\rangle_{(n-2)} \langle n_p\rangle_{(n-4)} \dots \langle n_p\rangle_{(n-2i+2)} a_i^{(k)}. \tag{B.78}$$

Therefore, using the definition of kth moment for a system size n in equation (3.108) and equation (B.68) finds

$$\langle n_p^k\rangle_n = \sum_{i=1}^{k} a_i^{(k)} \langle n_p\rangle_n \langle n_p\rangle_{(n-2)} \dots \langle n_p\rangle_{(n-2i+2)}. \tag{B.79}$$

B.9 The Asymptotic Form of the kth Moment

For $1 \ll n$ and $k = O(\sqrt{n})$, using equations (3.90)[2] and (3.108) we must have

$$\langle n_p^k\rangle_n = \sum_{i=1}^{k} \frac{a_i^{(k)}}{2^i}\frac{n!}{(n-2i)!}\frac{c_{n-2i}(r)}{c_n(r)} = \sum_{i=1}^{k} \frac{a_i^{(k)}}{2^i}\frac{n!}{(n-2i)!}\frac{c_{n-2}(r)}{c_n(r)}\frac{c_{n-4}(r)}{c_{n-2}(r)}\dots\frac{c_{n-2i}(r)}{c_{n-2i+2}(r)}$$

$$\sim \sum_{i=1}^{k} \frac{a_i^{(k)}}{2^i}\frac{n!}{(n-2i)!}\frac{\mathrm{e}^{-\sqrt{r/n}}}{n-2}\frac{\mathrm{e}^{-\sqrt{r/(n-2)}}}{n-4}\dots\frac{\mathrm{e}^{-\sqrt{r/(n-2i+2)}}}{n-2i}$$

[2]Note that in equation (3.90), the denominator is $n-2$ as

$$\frac{c_{n-2}(r)}{c_n(r)} \sim \frac{\mathrm{e}^{-\sqrt{r/n}}}{n-2}, \tag{B.80}$$

that we used its asymptotic form as n.

$$\sim \sum_{i=1}^{k} \frac{a_i^{(k)}}{2^i} \frac{n(n-1)}{n-2} \frac{(n-2)(n-3)}{n-4} \cdots \frac{(n-2i+2)(n-2i+1)}{n-2i} \mathrm{e}^{-\sqrt{r/n}\left(1+\sqrt{n/n-2}+\cdots+\sqrt{n/n-2(i-1)}\right)}$$

$$\sim \sum_{i=1}^{k} \frac{a_i^{(k)}}{2^i} n(n-2)\ldots(n-2i+2)\mathrm{e}^{-i\sqrt{r/n}}$$

$$(\text{Using } \tfrac{n-2j+1}{n-2j} \sim \tfrac{n}{n-2j} \sim 1, \quad 1 \le j \le k \le \sqrt{n})$$

$$\sim \sum_{i=1}^{k} \frac{a_i^{(k)}}{2^i} n^i (1-\frac{2}{n})\ldots(1-\frac{2(i-1)}{n})\mathrm{e}^{-i\sqrt{r/n}}$$

$$\sim \sum_{i=1}^{k} \frac{a_i^{(k)}}{2^i} n^i \left(1-\frac{i-1}{n}\right)^i \mathrm{e}^{-i\sqrt{r/n}}$$

$$\sim \frac{a_k^{(k)} n^k}{2^k}\left(1-\frac{k-1}{n}\right)^k e^{-k\sqrt{r/n}} + \frac{a_{k-1}^{(k)} n^{k-1}}{2^{k-1}}\left(1-\frac{k-2}{n}\right)^{k-1} \mathrm{e}^{-(k-1)\sqrt{r/n}} + O(n^{k-2}). \tag{B.81}$$

And since $a_k^{(k)} = 1$ and $a_{k-1}^{(k)} = k(k-1)/2$ and assuming $1 - \frac{k-2}{n} \sim 1 - \frac{k-1}{n}$, we get

$$\langle n_p^k \rangle_n \sim \frac{n^k \mathrm{e}^{-k\sqrt{r/n}}}{2^k}\left(1-\frac{k-1}{n}\right)^{k-1}\left(1-\frac{k-1}{n}+\frac{k(k-1)}{n}\mathrm{e}^{\sqrt{r/n}}\right) + O(n^{k-2})$$

$$\sim \frac{n^k \mathrm{e}^{-k\sqrt{r/n}}}{2^k}\left(1-\frac{(k-1)^2}{n}+O(\frac{1}{n^2})\right)\left(1-\frac{k-1}{n}+\frac{k(k-1)}{n}\mathrm{e}^{\sqrt{r/n}}\right) + O(n^{k-2})$$

$$\sim \frac{n^k \mathrm{e}^{-k\sqrt{r/n}}}{2^k}\left(1-\frac{k-1}{n}+\frac{k(k-1)}{n}\mathrm{e}^{\sqrt{r/n}}-\frac{(k-1)^2}{n}+O(\frac{1}{n^2})\right) + O(n^{k-2}) \implies$$

$$\langle n_p^k \rangle_n \sim \frac{n^k \mathrm{e}^{-k\sqrt{r/n}}}{2^k}\left(1+\frac{k(k-1)}{n}(\mathrm{e}^{\sqrt{r/n}}-1)\right) + O(n^{k-2}) \tag{B.82}$$

The resulting asymptotic leading term is based on the assumption $k = O(\sqrt{n})$, especially, the expansion of $(1-(k-1)/n)^{k-1}$ in the second line. Despite that, we see a good estimation even for k as a fraction of n in the numeric investigation in figures (B.1) and (B.2). Notice that the estimate is good enough for as small as $n = 100$.

At the same time, although for larger k the estimate starts to deviate from the actual, $\langle (\frac{2n_p}{n})^k \rangle$ is exponentially smaller for larger ks. Therefore, for a sum that includes $\langle (\frac{2n_p}{n})^k \rangle$, one can truncate the sum over k around $\sqrt{k}$ such that the error is exponentially small.

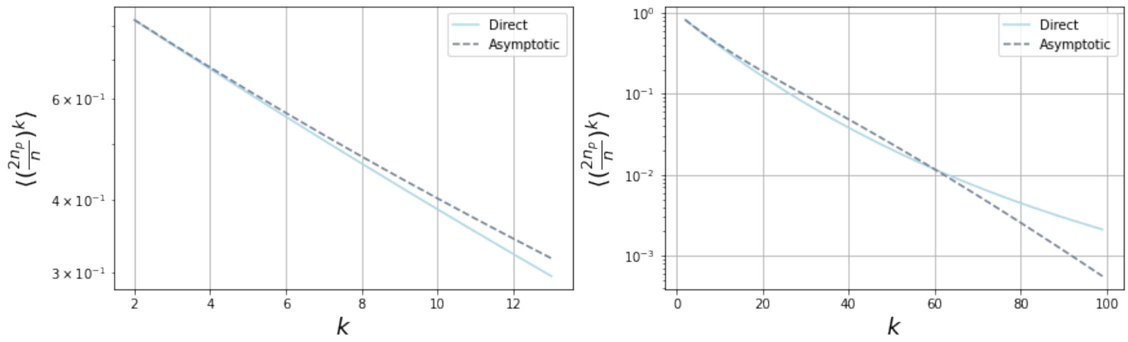


Figure B.1: Plotting $\langle(\frac{2n_p}{n})^k\rangle$ and its asymptotic leading term against k for as small n as $n = 100$. In the left panel, the k axis extends around $\sqrt{n}$. The right panel shows the k axis extends to $k = O(n)$.

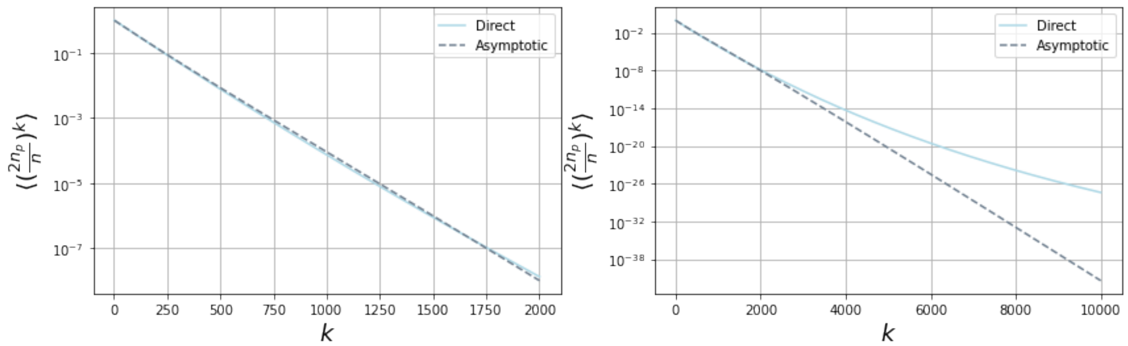


Figure B.2: Plotting $\langle(\frac{2n_p}{n})^k\rangle$ and its asymptotic leading term against k for $n = 10000$. In the left panel, the k axis extends to $n/5$, and it shows a good agreement even beyond $\sqrt{n}$. The right panel shows the k axis extends to $k = O(n)$.

It is useful to also find the asymptotic leading term of the moments of stand-alone elements, $\langle n_s^k\rangle$. So, from the definition of $n - 2n_p$ we get

$$\langle n_s^k\rangle = \langle(n-2n_p)^k\rangle = \langle\sum_{j=0}^{k}\binom{k}{j}n^{k-j}(-2n_p)^j\rangle$$

$$= \sum_{j=0}^{k}\binom{k}{j}n^{k-j}(-2)^j\langle n_p^j\rangle. \tag{B.83}$$

Next, for $1 \ll n$, equation (B.82) obtains

$$\langle n_s^k \rangle \sim \sum_{j=0}^{k} \binom{k}{j} n^{k-j} (-1)^j n^j \mathrm{e}^{-j\sqrt{r/n}} \left(1 + \frac{j(j-1)}{n} (\mathrm{e}^{\sqrt{r/n}} - 1) + O(\frac{1}{n^2})\right)$$

$$\sim n^k \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)^k + n^{k-1} (\mathrm{e}^{\sqrt{r/n}} - 1) \sum_{j=0}^{k} \binom{k}{j} j(j-1) \left(-\mathrm{e}^{-\sqrt{r/n}}\right)^j + O(n^{k-2})$$

$$\sim n^k \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)^k + n^{k-1} k(k-1) (\mathrm{e}^{\sqrt{r/n}} - 1) \sum_{j=2}^{k} \binom{k-2}{j-2} \left(-\mathrm{e}^{-\sqrt{r/n}}\right)^j + O(n^{k-2})$$

$$\sim n^k \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)^k + n^{k-1} k(k-1) (\mathrm{e}^{\sqrt{r/n}} - 1) \sum_{j=0}^{k-2} \binom{k-2}{j} \left(-\mathrm{e}^{-\sqrt{r/n}}\right)^{j+2} + O(n^{k-2})$$

$$\sim n^k \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)^k + \frac{n^k k(k-1)(\mathrm{e}^{\sqrt{r/n}} - 1)\mathrm{e}^{-2\sqrt{r/n}}}{n} \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)^{k-2} + O(n^{k-2}). \tag{B.84}$$

It finds

$$\langle n_s \rangle \sim n \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right) + O(\frac{1}{n}), \tag{B.85}$$

and

$$\langle n_s^k \rangle \sim n^k \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)^k \left[1 + \frac{k(k-1)\mathrm{e}^{-\sqrt{r/n}}}{n\left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)}\right] + O(n^{k-2}). \tag{B.86}$$

B.10 Probability Generating Functions

For constant n, the number of pairs has an upper bound as $n_p \in \{0, 1, \ldots, n/2\}$. Hence, $P_n(n_p > n/2) = 0$. The probability generating function of n_p with respect to $P_n(n_p)$ is

$$G_n(s) = \sum_{n_p=0}^{\infty} P_n(n_p) s^{n_p} = \frac{1}{c_n(r)} \sum_{n_p=0}^{\infty} \binom{n}{2n_p} (2n_p - 1)!!\, r^{\frac{n}{2} - n_p} s^{n_p}$$

$$= \frac{s^{\frac{n}{2}}}{c_n(r)} \sum_{n_p=0}^{\infty} \binom{n}{2n_p} (2n_p - 1)!! \left(\frac{r}{s}\right)^{\frac{n}{2} - n_p}$$

$$= s^{\frac{n}{2}} \frac{c_n(\frac{r}{s})}{c_n(r)}. \tag{B.87}$$

Notice that the normalization constant in the numerator evaluates at (r/s).

Doing the same for $P_n(n_p, n_h)$ one finds

$$G_n(s,u) = \sum_{n_p=0}^{\infty} \sum_{n_h=0}^{n-2n_p} P_n(n_p, n_h) s^{n_p} u^{n_h}$$

$$= \frac{1}{c_n(r)} \sum_{n_p=0}^{\infty} \binom{n}{2n_p} (2n_p-1)!!\, r^{\frac{n}{2}-n_p} s^{n_p} \sum_{n_h=0}^{n-2n_p} \binom{n-2n_p}{n_h} \rho^{n_h} u^{n_h} (1-\rho)^{n-2n_p-n_h}$$

$$= \frac{s^{\frac{n}{2}}}{c_n(r)} \sum_{n_p=0}^{\infty} \binom{n}{2n_p} (2n_p-1)!! \, \left(\frac{r}{s}\right)^{\frac{n}{2}-n_p} (\rho u + 1 - \rho)^{n-2n_p}$$

$$= s^{\frac{n}{2}} \frac{c_n\left(\frac{r(\rho u+1-\rho)^2}{s}\right)}{c_n(r)}. \tag{B.88}$$

The normalization constant in the numerator evaluates at $\left(r\left(\rho u + 1 - \rho\right)^2/s\right)$.

B.11 Finding Identities for $c_n(r)$

In this section, we prove the identities that we use in other sections.

B.11.1 Even Numbers Identities

For even numbers, we will show the following identities are valid:

$$c_{2n}(r) - (r + 2n - 1)c_{2n-1}(r) = -(2n-1)(2n-2)c_{2n-3}(r), \tag{B.89}$$

and

$$c_{2n-1}(r) - (r + 4n - 5)c_{2n-3}(r) = -(2n-3)(2n-4)c_{2n-5}(r). \tag{B.90}$$

We start from the left-hand side of equation (B.89)

$$c_{2n}(r) - (r+2n-1)c_{2n-1}(r) = \sum_{i=0}^{n} \binom{2n}{2i} (2i-1)!!\, r^{n-i} - (r+2n-1) \sum_{i=0}^{n-1} \binom{2n-1}{2i} (2i-1)!!\, r^{n-1-i}$$

$$= (2n-1)!! + \sum_{i=0}^{n-1} \binom{2n}{2i} (2i-1)!! r^{n-i} - \sum_{i=0}^{n-1} \binom{2n-1}{2i} (2i-1)!! r^{n-i}$$

$$-(2n-1)(2n-1)!! - (2n-1) \sum_{i=0}^{n-2} \binom{2n-1}{2i} (2i-1)!!\, r^{n-1-i}$$

$$= -(2n-2)(2n-1)!! + \sum_{i=1}^{n-1} \left[\binom{2n}{2i} - \binom{2n-1}{2i} \right] (2i-1)!! r^{n-i} - (2n-1) \sum_{i=0}^{n-2} \binom{2n-1}{2i} (2i-1)!!\, r^{n-1-i}$$

$$= -(2n-2)(2n-1)!! + \sum_{i=0}^{n-2} \left[\binom{2n}{2i+2} - \binom{2n-1}{2i+2} \right] (2i+1)!! r^{n-1-i} - (2n-1) \sum_{i=0}^{n-2} \binom{2n-1}{2i} (2i-1)!!\, r^{n-}$$

$$= -(2n-2)(2n-1)!! + \sum_{i=0}^{n-2} \left[\frac{2n}{(2i+2)(2n-2-2i)} - \frac{1}{2i+2} - \frac{2n-1}{(2n-1-2i)(2n-2-2i)} \right] \frac{(2n-}{(2i)!(2n-3}$$

$$\times (2i-1)!!\, r^{n-1-i}$$

$$= -(2n-2)(2n-1)!! + \sum_{i=0}^{n-2} \left[\frac{-2i}{(2n-1-2i)(2n-2-2i)} \right] \frac{(2n-1)!}{(2i)!(2n-3-2i)!} (2i-1)!!\, r^{n-1-i}$$

$$= -(2n-2)(2n-1)!! - \sum_{i=1}^{n-2} \frac{(2n-1)!}{(2i-2)!(2n-1-2i)!} (2i-3)!!\, r^{n-1-i}$$

$$= -(2n-2)(2n-1)!! - \sum_{i=0}^{n-3} \frac{(2n-1)!}{(2i)!(2n-3-2i)!} (2i-1)!!\, r^{n-2-i}$$

$$= -(2n-1)(2n-2) \sum_{i=0}^{n-2} \frac{(2n-3)!}{(2i)!(2n-3-2i)!} (2i-1)!!\, r^{n-2-i} = -(2n-1)(2n-2) c_{2n-3}(r), \tag{B.91}$$

and get the left right hand side. Similarly, for equation (B.89)

$$c_{2n-1}(r) - (r+4n-5) c_{2n-3}(r) = \sum_{i=0}^{n-1} \binom{2n-1}{2i} (2i-1)!!\, r^{n-1-i} - (r+4n-5) \sum_{i=0}^{n-2} \binom{2n-3}{2i} (2i-1)!!\, r^{n-2-i}$$

$$= (2n-1)!! + \sum_{i=0}^{n-2} \binom{2n-1}{2i} (2i-1)!! r^{n-1-i} - \sum_{i=0}^{n-2} \binom{2n-3}{2i} (2i-1)!! r^{n-1-i}$$

$$-(4n-5)(2n-3)!! - (4n-5) \sum_{i=0}^{n-3} \binom{2n-3}{2i} (2i-1)!!\, r^{n-2-i}$$

$$= -(2n-4)(2n-3)!! + \sum_{i=1}^{n-2} \left[\binom{2n-1}{2i} - \binom{2n-3}{2i} \right] (2i-1)!! r^{n-1-i} - (4n-5) \sum_{i=0}^{n-3} \binom{2n-3}{2i} (2i-1)!!\, r^{n-}$$

$$= -(2n-4)(2n-3)!! + \sum_{i=0}^{n-3} \left[\binom{2n-1}{2i+2} - \binom{2n-3}{2i+2} \right] (2i+1)!! r^{n-2-i} - (4n-5) \sum_{i=0}^{n-3} \binom{2n-3}{2i} (2i-1)!!\, r^{n-}$$

$$= -(2n-4)(2n-3)!! + \sum_{i=0}^{n-3} \left[\frac{(2n-1)(2n-2)}{(2i+2)(2n-3-2i)(2n-4-2i)} - \frac{1}{(2i+2)} - \frac{(4n-5)}{(2n-3-2i)(2n-4-2} \right.$$

$$\times \frac{(2n-3)!}{(2i)!(2n-5-2i)!} (2i-1)!!\, r^{n-2-i}$$

$$= -(2n-4)(2n-3)!! + \sum_{i=0}^{n-3} \left[\frac{-(2i)(2i+2)}{(2i+2)(2n-3-2i)(2n-4-2i)} \right] \frac{(2n-3)!}{(2i)!(2n-5-2i)!} (2i-1)!!\, r^{n-2-i}$$

$$= -(2n-4)(2n-3)!! - \sum_{i=1}^{n-3} \frac{(2n-3)!}{(2i-1)!(2n-3-2i)!} (2i-1)!!\, r^{n-2-i}$$

$$= -(2n-4)(2n-3)!! - \sum_{i=0}^{n-4} \frac{(2n-3)!}{(2i+1)!(2n-5-2i)!} (2i+1)!!\, r^{n-3-i}$$

$$= -(2n-3)(2n-4) \sum_{i=0}^{n-3} \frac{(2n-5)!}{(2i)!(2n-5-2i)!} (2i-1)!!\, r^{n-3-i} = -(2n-3)(2n-4) c_{2n-5}(r). \tag{B.92}$$

B.11.2 Odd Numbers Identities

For odd numbers, we define

$$h_{2n+1}(r) \equiv \frac{dc_{2n+1}(r)}{dr} = nc_{2n-1}(r) + 2n \frac{dc_{2n-1}(r)}{dr}, \tag{B.93}$$

and show the following identities are valid:

$$c_{2n+1}(r) - \frac{r+2n+1}{n} h_{2n+1}(r) = -2(2n+1) h_{2n-1}(r), \tag{B.94}$$

and

$$h_{2n+1}(r) - \frac{n(r+4n-3)}{n-1} h_{2n-1}(r) = -2n(2n-1) h_{2n-3}(r). \tag{B.95}$$

We start from the left-hand side of equation (B.94)

$$c_{2n+1}(r) - \frac{r+2n+1}{n} h_{2n+1}(r) = \sum_{i=0}^{n} \binom{2n+1}{2i} (2i-1)!!\, r^{n-i}$$

$$
\begin{aligned}
&-\frac{r+2n+1}{n}\sum_{i=0}^{n-1}(n-i)\binom{2n+1}{2i}(2i-1)!!r^{n-1-i}\\
&= (2n+1)!! + \sum_{i=0}^{n-1}\binom{2n+1}{2i}(2i-1)!!\,r^{n-i} - \frac{1}{n}\sum_{i=0}^{n-1}(n-i)\binom{2n+1}{2i}(2i-1)!!r^{n-i}\\
&-\frac{2n+1}{3}(2n+1)!! - \frac{2n+1}{n}\sum_{i=0}^{n-2}(n-i)\binom{2n+1}{2i}(2i-1)!!r^{n-1-i}\\
&= -\frac{(2n-2)}{3}(2n{+}1)!!{+}\sum_{i=1}^{n-1}\frac{i}{n}\binom{2n+1}{2i}(2i{-}1)!!\,r^{n-i}{-}\frac{2n+1}{n}\sum_{i=0}^{n-2}(n{-}i)\binom{2n+1}{2i}(2i{-}1)!!r^{n-1-i}\\
&= -\frac{(2n-2)}{3}(2n{+}1)!!{+}\sum_{i=0}^{n-2}\frac{i+1}{n}\binom{2n+1}{2i+2}(2i{+}1)!!\,r^{n-1-i}{-}\frac{2n+1}{n}\sum_{i=0}^{n-2}(n{-}i)\binom{2n+1}{2i}(2i{-}1)!!r^{n-1-i}\\
&= -\frac{(2n-2)}{3}(2n{+}1)!!{+}\frac{1}{n}\sum_{i=0}^{n-2}\left[\frac{i+1}{2i+2} - \frac{(2n+1)(n-i)}{(2n+1-2i)(2n-2i)}\right]\frac{(2n+1)!}{(2i)!(2n-1-2i)!}(2i{-}1)!!r^{n-1-i}\\
&= -\frac{(2n-2)}{3}(2n+1)!! - \frac{1}{2n}\sum_{i=0}^{n-2}\left[\frac{2i}{(2n+1-2i)}\right]\frac{(2n+1)!}{(2i)!(2n-1-2i)!}(2i-1)!!r^{n-1-i}\\
&= -\frac{(2n-2)}{3}(2n+1)!! - \frac{1}{2n}\sum_{i=1}^{n-2}\frac{(2n-2i)(2n+1)!}{(2i-1)!(2n+1-2i)!}(2i-1)!!r^{n-1-i}\\
&= -\frac{(2n-2)}{3}(2n+1)!! - \frac{1}{n}\sum_{i=0}^{n-3}\frac{(n-1-i)(2n+1)!}{(2i+1)!(2n-1-2i)!}(2i+1)!!r^{n-2-i}\\
&= -\frac{(2n-2)}{3}(2n+1)!! - \frac{2n(2n+1)}{n}\sum_{i=0}^{n-3}(n-1-i)\binom{2n-1}{2i}(2i-1)!!r^{n-2-i}\\
&= -2(2n+1)\sum_{i=0}^{n-2}(n-1-i)\binom{2n-1}{2i}(2i-1)!!r^{n-2-i}\\
&= -2(2n+1)h_{2n-1}(r), \qquad \text{(B.96)}
\end{aligned}
$$

and obtain the right hand side. Similarly, for equation (B.95) we find

$$
\begin{aligned}
h_{2n+1}(r) - \frac{n(r+4n-3)}{n-1}h_{2n-1}(r) &= \sum_{i=0}^{n-1}(n-i)\binom{2n+1}{2i}(2i-1)!!r^{n-1-i}\\
&-\frac{n(r+4n-3)}{n-1}\sum_{i=0}^{n-2}(n-1-i)\binom{2n-1}{2i}(2i-1)!!r^{n-2-i}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{n}{3}(2n+1)!! + \sum_{i=0}^{n-2}(n-i)\binom{2n+1}{2i}(2i-1)!!r^{n-1-i} - \frac{n}{n-1}\sum_{i=0}^{n-2}(n-1-i)\binom{2n-1}{2i}(2i-1)!!r^{n-1-i} \\
&\quad - \frac{n(4n-3)}{3}(2n-1)!! - \frac{n(4n-3)}{n-1}\sum_{i=0}^{n-3}(n-1-i)\binom{2n-1}{2i}(2i-1)!!r^{n-2-i} \\
&= -\frac{2n(n-2)}{3}(2n-1)!! + \sum_{i=1}^{n-2}\left[(n-i)\binom{2n+1}{2i} - (n-\frac{n}{n-1}i)\binom{2n-1}{2i}\right](2i-1)!!r^{n-1-i} \\
&\qquad - \frac{n(4n-3)}{n-1}\sum_{i=0}^{n-3}(n-1-i)\binom{2n-1}{2i}(2i-1)!!r^{n-2-i} \\
&= -\frac{2n(n-2)}{3}(2n-1)!! + \sum_{i=0}^{n-3}\left[(n-1-i)\binom{2n+1}{2i+2} - (n-\frac{n(i+1)}{n-1})\binom{2n-1}{2i+2}\right](2i+1)!!r^{n-2-i} \\
&\qquad - \frac{n(4n-3)}{n-1}\sum_{i=0}^{n-3}(n-1-i)\binom{2n-1}{2i}(2i-1)!!r^{n-2-i} \\
&= -\frac{2n(n-2)}{3}(2n-1)!! + \sum_{i=0}^{n-3}\left[\frac{n(2n+1)}{(2i+2)(2n-1-2i)} - n(\frac{1}{2i+2} - \frac{1}{2(n-1)}) - \frac{n(4n-3)}{2(n-1)(2n-1-2i)}\right] \\
&\qquad \times \frac{(2n-1)!}{(2i)!(2n-3-2i)!}(2i-1)!!r^{n-2-i} \\
&= -\frac{2n(n-2)}{3}(2n-1)!! - \frac{n}{n-1}\sum_{i=1}^{n-3}\frac{i}{(2n-1-2i)}\times\frac{(2n-1)!}{(2i)!(2n-3-2i)!}(2i-1)!!r^{n-2-i} \\
&= -\frac{2n(n-2)}{3}(2n-1)!! - \frac{n}{n-1}\sum_{i=0}^{n-4}\frac{i+1}{(2n-3-2i)}\times\frac{(2n-1)!}{(2i+2)!(2n-5-2i)!}(2i+1)!!r^{n-3-i} \\
&= -\frac{2n(n-2)}{3}(2n-1)!! - 2n(2n-1)\sum_{i=0}^{n-4}(n-2-i)\times\frac{(2n-3)!}{(2i)!(2n-3-2i)!}(2i-1)!!r^{n-3-i} \\
&= -2n(2n-1)\sum_{i=0}^{n-3}(n-2-i)\binom{2n-3}{2i}(2i-1)!!r^{n-3-i} = -2n(2n-1)h_{2n-3}(r).
\end{aligned}
\tag{B.97}
$$

B.12 Joint Probability Distributions

B.12.1 $P_n(X_1, n_p)$

We can use a combinatorial argument to find the joint distribution $P_n(X_1, n_p)$ as follows: given there are n_p pairs, when the element at an arbitrary index is in a pair state ($X_1 = 0$) there must be $(n-1)\binom{n-2}{2n_p-2}(2n_p-3)!! = \binom{n-1}{2n_p-1}(2n_p-1)!!$ distinct configurations, and similarly, if the element is in a stand-alone state ($X_1 = 1$), there are $\binom{2n-1}{2n_p}(2n_p-1)!!$ distinct configurations, and in total there are $\binom{n}{2n_p}(2n_p-1)!!$. Note that the above statement is equivalent to the following identity

$$\binom{n}{2n_p}(2n_p-1)!! = \binom{n-1}{2n_p-1}(2n_p-1)!! + \binom{n-1}{2n_p}(2n_p-1)!!. \tag{B.98}$$

And since the probability of a single configuration with n_p pairs is $r^{\lfloor n/2 \rfloor - n_p}/c_n(r)$, the joint probability must be

$$P_n(X_1, n_p) = \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)} \times \begin{cases} \binom{n-1}{2n_p-1}(2n_p-1)!! & , X_1 = 0 \\ \binom{n-1}{2n_p}(2n_p-1)!! & , X_1 = 1 \end{cases}. \tag{B.99}$$

Also, the conditional distribution can be calculated by dividing $P_n(X_1, n_p)$ by $P_n(n_p)$

$$P_n(X_1|n_p) = \frac{P_n(X_1, n_p)}{P_n(n_p)} = \begin{cases} \frac{2n_p}{n} & , X_1 = 0 \\ \frac{n-2n_p}{n} & , X_1 = 1 \end{cases}. \tag{B.100}$$

B.12.2 $P_n(X_1, X_2, n_p)$

There are four different combinations regarding X_l and X_k, through which, we use combinatorial arguments to find the joint probability distribution:

- For $X_1 = X_2 = 1$, or both balls in stand-alone state, there are $\binom{n-2}{2n_p}(2n_p-1)!!$ distinct configurations, whenever n_p pairs are in the remaining $n-2$ elements. And since the probability of a single configuration with n_p pairs is $r^{\lfloor n/2 \rfloor - n_p}/c_n(r)$, the joint probability must be

$$P_n(X_1 = 1, X_2 = 1, n_p) = \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)} \binom{n-2}{2n_p}(2n_p-1)!!. \tag{B.101}$$

- For $X_1 = 1$ and $X_2 = 0$, or one ball in stand-alone state and the other one in pair state with another one of the $n-2$ remaining elements, there are $(n-2)\binom{n-3}{2n_p-2}(2n_p-3)!!$ distinct configurations. And since the probability of a single configuration with n_p pairs is $r^{\lfloor n/2 \rfloor - n_p}/c_n(r)$, the joint probability must be

$$\begin{aligned} P_n(X_1 = 1, X_2 = 0, n_p) &= \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)}(n-2)\binom{n-3}{2n_p-2}(2n_p-3)!! \\ &= \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)}\binom{n-2}{2n_p-1}(2n_p-1)!!. \end{aligned} \tag{B.102}$$

- The argument for the case $X_1 = 0$ and $X_2 = 1$ is similar to the previous one.

$$P_n(X_1 = 0, X_2 = 1, n_p) = \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)}\binom{n-2}{2n_p-1}(2n_p-1)!!. \tag{B.103}$$

- The case $X_1 = X_2 = 0$ corresponds to two different possibilities. The first combination consists of a pair between the elements at indices l and k and $2n_p - 2$ elements among the remaining part are in pairs with each others. Therefore, there are $\binom{n-2}{2n_p-2}(2n_p-3)!!$ distinct configurations, each with probability $r^{\lfloor n/2 \rfloor - n_p}/c_n(r)$.

  In contrast in the second case, both elements 1 and 2 are in a pair state with one element in the remaining balls. And hence, there are $(n-2)(n-3)\binom{n-4}{2n_p-4}(2n_p-5)!!$ distinct configurations, each with probability $r^{\lfloor n/2 \rfloor - n_p}/c_n(r)$.

  Therefore,

$$\begin{aligned} P_n(X_1 = 0, X_2 = 0, n_p) &= \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)}\binom{n-2}{2n_p-2}(2n_p-3)!! \\ &+ \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)}(n-2)(n-3)\binom{n-4}{2n_p-4}(2n_p-5)!! \implies \\ P_n(X_1 = 0, X_2 = 0, n_p) &= \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)}\binom{n-2}{2n_p-2}(2n_p-1)!!. \end{aligned} \tag{B.104}$$

Putting all the results together, we get

$$P_n(X_1, X_2, n_p) = \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)} \times \begin{cases} \binom{n-2}{2n_p-2}(2n_p-1)!! & , X_1 = X_2 = 0 \\ \binom{n-2}{2n_p-1}(2n_p-1)!! & , X_1 = 1, X_2 = 0 \\ \binom{n-2}{2n_p-1}(2n_p-1)!! & , X_1 = 0, X_2 = 1 \\ \binom{n-2}{2n_p}(2n_p-1)!! & , X_1 = X_2 = 1 \end{cases}. \tag{B.105}$$

The condition distribution finds by dividing the last result by $P_n(n_p)$ as

$$P_n(X_1, X_2|n_p) = \frac{P_n(X_1, X_2, n_p)}{P_n(n_p)} = \begin{cases} \binom{n-2}{2n_p-2}(2n_p-1)!!/\binom{n}{2n_p}(2n_p-1)!! & , X_1 = X_2 = 0 \\ \binom{n-2}{2n_p-1}(2n_p-1)!!/\binom{n}{2n_p}(2n_p-1)!! & , X_1 = 1, X_2 = 0 \\ \binom{n-2}{2n_p-1}(2n_p-1)!!/\binom{n}{2n_p}(2n_p-1)!! & , X_1 = 0, X_2 = 1 \\ \binom{n-2}{2n_p}(2n_p-1)!!/\binom{n}{2n_p}(2n_p-1)!! & , X_1 = X_2 = 1 \end{cases}$$

$$\implies P_n(X_1, X_2|n_p) = \begin{cases} \frac{2n_p(2n_p-1)}{n(n-1)} & , X_1 = X_2 = 0 \\ \frac{2n_p(n-2n_p)}{n(n-1)} & , X_1 = 1, X_2 = 0 \\ \frac{2n_p(n-2n_p)}{n(n-1)} & , X_1 = 0, X_2 = 1 \\ \frac{(n-2n_p)(n-1-2n_p)}{n(n-1)} & , X_1 = X_2 = 1 \end{cases}. \qquad \text{(B.106)}$$

Or by using $n_s = n - 2n_p$ it simplifies to

$$P_n(X_1, X_2|n_p) = \begin{cases} \frac{2n_p(2n_p-1)}{n(n-1)} & , X_1 = X_2 = 0 \\ \frac{2n_p n_s}{n(n-1)} & , X_1 = 1, X_2 = 0 \\ \frac{2n_p n_s}{n(n-1)} & , X_1 = 0, X_2 = 1 \\ \frac{n_s(n_s-1)}{n(n-1)} & , X_1 = X_2 = 1 \end{cases}. \qquad \text{(B.107)}$$

The marginal distribution finds by summing over n_p as

- For $X_1 = X_2 = 0$

$$P_n(X_1 = 0, X_2 = 0) = \sum_{n_p=1}^{\lfloor n/2 \rfloor} \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)} \binom{n-2}{2n_p-2}(2n_p-1)!!$$

$$= \sum_{n_p=0}^{\lfloor n/2 \rfloor} \frac{2n_p(2n_p-1)}{n(n-1)} \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)} \binom{n}{2n_p}(2n_p-1)!!$$

$$= \frac{\langle 2n_p(2n_p-1)\rangle_n}{n(n-1)}. \qquad \text{(B.108)}$$

- For $X_1 = 1, X_2 = 0$ or $X_1 = 0, X_2 = 1$ we obtain

$$P_n(X_1 = 0, X_2 = 1) = P_n(X_1 = 1, X_2 = 0) = \sum_{n_p=1}^{\lfloor n/2 \rfloor} \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)} \binom{n-2}{2n_p-1}(2n_p-1)!!$$

$$= \sum_{n_p=0}^{\lfloor n/2 \rfloor} \frac{2n_p(n-2n_p)}{n(n-1)} \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)} \binom{n}{2n_p}(2n_p-1)!!$$

$$= \frac{\langle 2n_p(n-2n_p)\rangle_n}{n(n-1)} = \frac{\langle 2n_p n_s\rangle_n}{n(n-1)}. \tag{B.109}$$

- For $X_1 = X_2 = 0$

$$P_n(X_1 = 1, X_2 = 1) = \sum_{n_p=0}^{\lfloor n/2\rfloor -1} \frac{r^{\lfloor n/2\rfloor - n_p}}{c_n(r)} \binom{n-2}{2n_p} (2n_p - 1)!!$$

$$= \sum_{n_p=0}^{\lfloor n/2\rfloor} \frac{(n-2n_p)(n-2n_p-1)}{n(n-1)} \frac{r^{\lfloor n/2\rfloor - n_p}}{c_n(r)} \binom{n}{2n_p} (2n_p - 1)!!$$

$$= \frac{\langle (n-2n_p)(n-2n_p-1)\rangle_n}{n(n-1)} = \frac{\langle n_s(n_s-1)\rangle_n}{n(n-1)}. \tag{B.110}$$

Consequently,

$$P_n(X_1, X_2) = \begin{cases} \frac{\langle 2n_p(2n_p-1)\rangle_n}{n(n-1)} & , X_1 = X_2 = 0 \\ \frac{\langle 2n_p n_s\rangle_n}{n(n-1)} & , X_1 = 1, X_2 = 0 \\ \frac{\langle 2n_p n_s\rangle_n}{n(n-1)} & , X_1 = 0, X_2 = 1 \\ \frac{\langle n_s(n_s-1)\rangle_n}{n(n-1)} & , X_1 = X_2 = 1 \end{cases}. \tag{B.111}$$

B.12.3 $P_n(X_1, \ldots, X_k, n_p)$

Let us assume for $P_n(X_1, \ldots, X_k, n_p)$ we have $l = k - \sum_{i=1}^{k} X_i$. In other words, $\sum_{i=1}^{k} X_i$ gives the number of elements in stand-alone state among k ones ($X_i = 1$), and l denotes the number of elements that are in pair state. So, $l \in \{0, 1, 2, \ldots, k\}$.

However, among these l elements in the pair state, some make pairs with each other, and some with $n-k$ other elements in the system. Denoting by ω the elements that are paired among l, we must have $\omega \in \{0, 2, \ldots, \lfloor l/2\rfloor\}$. As a result, it asserts we get $\binom{l}{2\omega}(2\omega - 1)!!$ distinct combinations resulted from the pairing among the k element.

In addition, $l - \omega$ of the remaining pairs must have a counter part in $n - k$ other elements, through which, there are $(n-k)^{(l-2\omega)} = (n-k)(n-k-1)\ldots(n-k-l+2\omega+1)$. Furthermore, the $n_p - l$ remaining pairs have $\binom{n-k-(l-2\omega)}{2n_p-2l+2\omega}(2n_p - 2l + 2\omega - 1)!!$ distinct combinations. In total, the number of distinct configurations for a given

$(X_1, \ldots, X_k, n_p)$ is

$$W = \sum_{\omega=0}^{\lfloor l/2 \rfloor} \binom{n-k-(l-2\omega)}{2n_p - 2l + 2\omega} (2n_p - 2l + 2\omega - 1)!! \binom{l}{2\omega} (2\omega - 1)!! (n-k)^{(l-2\omega)}. \tag{B.112}$$

To simplify the sum, observe that

$$\begin{aligned} \binom{n-k-(l-2\omega)}{2n_p - 2l + 2\omega} (n-k)^{(l-2\omega)} &= \frac{(n-k-l+2\omega)!\,(n-k)^{(l-2\omega)}}{(2n_p - 2l + 2\omega)!\,(n-k-2n_p+l)!} \\ &= \frac{(n-k)!}{(2n_p - 2l + 2\omega)!\,(n-k-2n_p+l)!}. \end{aligned} \tag{B.113}$$

and the sum is written as

$$\begin{aligned} W &= \frac{(n-k)!}{(n-k-2n_p+l)!} \sum_{\omega=0}^{\lfloor l/2 \rfloor} \frac{(2n_p - 2l + 2\omega - 1)!!}{(2n_p - 2l + 2\omega)!} \binom{l}{2\omega} (2\omega - 1)!! \\ &= \frac{(n-k)!(2n_p - 2l - 1)!!}{(n-k-2n_p+l)!(2n_p - 2l)!} \sum_{\omega=0}^{\lfloor l/2 \rfloor} \binom{l}{2\omega} (2\omega - 1)!! \\ &\quad \times \frac{(2n_p - 2l + 1)(2n_p - 2l + 3)\ldots(2n_p - 2l + 2\omega - 1)}{(2n_p - 2l + 1)(2n_p - 2l + 2)\ldots(2n_p - 2l + 2\omega)} \\ &= \frac{(n-k)!(2n_p - 2l - 1)!!}{(n-k-2n_p+l)!(2n_p - 2l)!} \sum_{\omega=0}^{\lfloor l/2 \rfloor} \frac{\binom{l}{2\omega}(2\omega - 1)!!}{(2n_p - 2l + 2)(2n_p - 2l + 4)\ldots(2n_p - 2l + 2\omega)} \\ &= \frac{(n-k)!(2n_p - 2l - 1)!!}{(n-k-2n_p+l)!(2n_p - 2l)!} \\ &\quad \times \frac{\sum_{\omega=0}^{\lfloor l/2 \rfloor} \binom{l}{2\omega}(2\omega - 1)!!(2n_p - 2l + 2\omega + 2)\ldots(2n_p - 2l + 2\lfloor l/2 \rfloor)}{(2n_p - 2l + 2)(2n_p - 2l + 4)\ldots(2n_p - 2l + 2\lfloor l/2 \rfloor)} \end{aligned} \tag{B.114}$$

For an even l we have

$$\begin{aligned} W &= \frac{(n-k)!(2n_p - l - 1)!!}{(n-k-2n_p+l)!(2n_p - l)!} \sum_{\omega=0}^{l/2} \binom{l}{2\omega} (2\omega - 1)!!(2n_p - 2l + 2\omega + 2)\ldots(2n_p - l) \\ &= \frac{(n-k)!(2n_p - l - 1)!!}{(n-k-2n_p+l)!(2n_p - l)!} \times (2n_p - l + 1)\ldots(2n_p - 3)(2n_p - 1) \\ &= \frac{(n-k)!(2n_p - 1)!!}{(n-k-2n_p+l)!(2n_p - l)!} = \binom{n-k}{2n_p - l} (2n_p - 1)!!. \end{aligned} \tag{B.115}$$

And for an odd l we have

$$W = \frac{(n-k)!(2n_p-l)!!}{(n-k-2n_p+l)!(2n_p-l)!} \sum_{\omega=0}^{(l-1)/2} \binom{l}{2\omega} (2\omega-1)!!(2n_p-2l+2\omega+2)\dots(2n_p-l-1)$$

$$= \frac{(n-k)!(2n_p-l)!!}{(n-k-2n_p+l)!(2n_p-l)!} \times (2n_p-l+2)\dots(2n_p-3)(2n_p-1)$$

$$= \frac{(n-k)!(2n_p-1)!!}{(n-k-2n_p+l)!(2n_p-l)!} = \binom{n-k}{2n_p-l}(2n_p-1)!!. \tag{B.116}$$

And since the probability of a single configuration with n_p pairs is $r^{\lfloor n/2 \rfloor - n_p}/c_n(r)$, for $l = k - \sum_{i=1}^{k} X_i$ the joint probability must be

$$P_n(X_1,\dots,X_k,n_p) = \frac{r^{\lfloor n/2 \rfloor - n_p}}{c_n(r)} \binom{n-k}{2n_p-l}(2n_p-1)!!. \tag{B.117}$$

In addition, the conditional distribution derives by dividing the joint distribution by $P_n(n_p)$ like

$$P_n(X_1,\dots,X_k|n_p) = \frac{P_n(X_1,\dots,X_k,n_p)}{P_n(n_p)} = \frac{\binom{n-k}{2n_p-l}(2n_p-1)!!}{\binom{n}{2n_p}(2n_p-1)!!}$$

$$= \frac{2n_p(2n_p-1)\dots(2n_p-l+1) \times (n-2n_p)(n-2n_p-1)\dots(n-2n_p-k+l+1)}{n(n-1)\dots(n-k+1)} \Longrightarrow$$

$$P_n(X_1,\dots,X_k|n_p) = \frac{2n_p^{(l)}(n-2n_p)^{(k-l)}}{n^{(k)}}, \tag{B.118}$$

or in terms of $n_s = n - 2n_p$, it writes

$$P_n(X_1,\dots,X_k|n_p) = \frac{2n_p^{(l)} n_s^{(k-l)}}{n^{(k)}}. \tag{B.119}$$

Finally, the marginal finds by summing over n_p

$$P_n(X_1,\dots,X_k) = \sum_{n_p=0}^{\lfloor n/2 \rfloor} P_n(X_1,\dots,X_k,n_p) = \sum_{n_p=0}^{\lfloor n/2 \rfloor} P_n(X_1,\dots,X_k|n_p)P_n(n_p)$$

$$= \sum_{n_p=0}^{\lfloor n/2 \rfloor} \frac{2n_p^{(l)}(n-2n_p)^{(k-l)}}{(n-k)^{(l)}} P_n(n_p) = \langle \frac{2n_p^{(l)}(n-2n_p)^{(k-l)}}{n^{(k)}} \rangle, \tag{B.120}$$

or in terms of n_s

$$P_n(X_1, \ldots, X_k) = \langle \frac{2n_p^{(l)} n_s^{(k-l)}}{n^{(k)}} \rangle. \tag{B.121}$$

B.13 Proving $c_n(r)$ Has $\lfloor n/2 \rfloor$ Distinct Roots

Theorem B.13.1. ***[Sturm' Theorem]*** *[4] Given a and b in $\mathbb{R} \cup \{-\infty, \infty\}$,*

$$V(SS(P, P'), a) - V(SS(P, P'), b)$$

is the number of roots of P in the interval (a, b).

We will use Sturm's Theorem to find the number of distinct roots of $c_n(r)$. And since the even and odd values require different treatment, we will prove them separately.

B.13.0.1 Case one: even numbers

For even number in appendix (B.7), equation (B.77) we show that

$$\frac{d}{dr} c_{2n}(r) = n c_{2n-1}(r), \tag{B.122}$$

and therefore, the first two elements of the Sturm sequence are $P_0 = c_{2n}(r)$ and $P_1 = c_{2n-1}(r)$. In appendix (B.11), the derivations of some useful identities are explained in details. We start with equation (B.89) that finds the Euclidean division of P_0 by P_1 as

$$c_{2n}(r) - (r + 2n - 1)c_{2n-1}(r) = -(2n-1)(2n-2)c_{2n-3}(r). \tag{B.123}$$

We see the quotient polynomial of the Euclidean division is equal to $r + 2n - 1$, and the remainder polynomial is $-(2n-1)(2n-2)c_{2n-3}(r)$. Therefore, the next element of the Sturm sequence is the negative of this reminder polynomial $P_2 = c_{2n-3}(r)$. Note that the positive constant term does not have any effect on the sign of the polynomial and we can safely remove it.

In addition, equation (B.90) finds the reminder of the division of $c_{2n-1}(r)$ by $c_{2n-3}(r)$, or P_1 by P_2, as

$$c_{2n-1}(r) - (r + 4n - 5)c_{2n-3}(r) = -(2n-3)(2n-4)c_{2n-5}(r). \tag{B.124}$$

This result shows that the next element of the Sturm sequence again is the next normalisation constant with an odd degree, say $c_{2n-5}(r)$. If we continue the Euclidean division consecutively, we get a sequence of the normalisation constant with decreasing odd degrees. In other words, all the elements of the Sturm sequence for an even number are

$$SS(c_{2n}(r), c_{2n-1}(r)) = c_{2n}(r), c_{2n-1}(r), c_{2n-3}(r), \ldots, c_3(r), c_1(r), \tag{B.125}$$

for constant $c_1(r) = 1$. Since all the coefficients of $c_n(r)$ are positive

$$\lim_{r \to \infty} c_n(r) > 0 \tag{B.126}$$

for all degrees, and therefore

$$V(SS(c_{2n}(r), c_{2n-1}(r)), \infty) = 0. \tag{B.127}$$

Hence, from the Sturm' Theorem, the number of distinct roots of $c_{2n}(r)$ is equal to $V(SS(c_{2n}(r), c_{2n-1}(r)), -\infty)$.

Observe that the degree of two consecutive polynomials in the Sturm sequence, say $c_{2n-(2i+1)}(r)$ and $c_{2n-(2i-1)}(r)$, are

$$\lfloor \frac{2n-(2i+1)}{2} \rfloor = n - i, \qquad \lfloor \frac{2n-(2i-1)}{2} \rfloor = n - i - 1, \tag{B.128}$$

are even and odd. Hence, in the limit $r \to -\infty$, from one polynomial to the next the sign changes. So, it is straightforward to use the definition of the number of sign changes and start from $c_3(r)$ as the basis of induction and show that

$$V(c_{2n-1}(r), c_{2n-3}(r), \ldots, c_3(r), c_1(r), -\infty) = n - 1. \tag{B.129}$$

The alternating degree happens for the first element of the Sturm sequence too

$$\lfloor \frac{2n}{2} \rfloor = n, \qquad \lfloor \frac{2n-1}{2} \rfloor = n - 1, \tag{B.130}$$

and therefore,

$$\begin{aligned} V(SS(c_{2n}(r), c_{2n-1}(r)), -\infty) &= V(c_{2n}(r), c_{2n-1}(r), c_{2n-3}(r), \ldots, c_1(r), -\infty) \\ &= V(c_{2n-1}(r), c_{2n-3}(r), \ldots, c_1(r), -\infty) + 1 = n. \end{aligned} \tag{B.131}$$

Last result proves that the polynomial $c_{2n}(r)$ has n distinct roots.

B.13.0.2 Case two: odd numbers

The odd case is very similar to the even case. The main difference is in the derivative of the $c_{2n+1}(r)$. We define a polynomial $h_{2n+1}(r)$ as

$$h_{2n+1}(r) \equiv \frac{dc_{2n+1}(r)}{dr} = nc_{2n-1}(r) + 2n\frac{dc_{2n-1}(r)}{dr}, \tag{B.132}$$

which is again a polynomial degree $2n-1$ with positive coefficients. In appendix (B.11), equation (B.94), the Euclidean division of the first two elements of the Sturm sequence obtains as

$$c_{2n+1}(r) - \frac{r+2n+1}{n}h_{2n+1}(r) = -2(2n+1)h_{2n-1}(r), \tag{B.133}$$

where the reminder is $-h_{2n-1}(r)$. Besides, equation (B.95) derives the Euclidean division of the subsequent elements as

$$h_{2n+1}(r) - \frac{n(r+4n-3)}{n-1}h_{2n-1}(r) = -2n(2n-1)h_{2n-3}(r). \tag{B.134}$$

And finally, the Sturm sequence finds as

$$SS(c_{2n+1}(r), h_{2n+1}(r)) = c_{2n+1}(r), h_{2n+1}(r), h_{2n-1}(r), \ldots, h_3(r), \tag{B.135}$$

for constant $h_3(r) = 1$. Because all elements in $SS(c_{2n+1}(r), h_{2n+1}(r))$ are polynomials with positive coefficients, here again no sign changes occur in the limit $r \to \infty$, and thus,

$$V(SS(c_{2n+1}(r), h_{2n+1}(r)), \infty) = 0. \tag{B.136}$$

Notice that from its definition, the degree of $h_{2n-(2i+1)}(r)$ is equal to the degree of $c_{2n-(2i+3)}(r)$, which is $n-i-2$. *e.g.* $h_{2n+1}(r)$ is a polynomial degree $n-1$.

Consequently, the signs of the sequence $h_{2n+1}(r), h_{2n-1}(r), \ldots, h_3(r)$ alternates, and we must have

$$V(h_{2n+1}(r), h_{2n-1}(r), \ldots, h_3(r)) = n-1. \tag{B.137}$$

The alternating sign happens between $c_{2n+1}(r)$ and $h_{2n+1}(r)$ from the fact that the former is a polynomial degree n and the latter is $n-1$. Therefore, we get

$$V(SS(c_{2n}(r), h_{2n-1}(r)), -\infty) = V(c_{2n+1}(r), h_{2n+1}(r), h_{2n-1}(r), \ldots, h_3(r), -\infty)$$
$$= V(h_{2n+1}(r), h_{2n-1}(r), \ldots, h_3(r), -\infty) + 1 = n. \tag{B.138}$$

This result proves that the polynomial $c_{2n+1}(r)$ has n distinct roots.

Appendix

C

Information Theory

C.1 Information Theory Measures

In this part, we will show the steps to derive some of the results in the Information Theory section. Before starting, we define some notations that simplify the notations further in the section.

Recall that the Shannon entropy is a functional that is defined over the space of probability distributions for system size n, denotes by $\mathcal{P}_n$, such that

$$H[P_n] : \mathcal{P}_n \to \mathbb{R}^+ \cup \{0\}. \tag{C.1}$$

At the same time, we define a Shannon *function* $H_2(x) : [0, 1] \to \mathbb{R}^+ \cup \{0\}$ as

$$H_2(x) = -x \ln x - (1 - x) \ln(1 - x), \tag{C.2}$$

for random variables that are in $[0, 1]$. We shall use these definition in the following sections.

C.1.1 The entropy of the Binomial Distribution

For ρ as the probability of getting head, the probability of observing a distinct configuration with n_h head among n coins is

$$\rho^{n_h}(1-\rho)^{n-n_h}, \tag{C.3}$$

and therefore, the Shannon entropy of the Binomial distribution, $Bin_n(n_h)$, derives as

$$H[Bin_n(n_h)] = -\sum_{i=1}^{2^n} P_i \ln P_i = -\sum_{n_h=0}^{n} \binom{n}{n_h} \rho^{n_h}(1-\rho)^{n-n_h} \left(n_h \ln\rho + (n-n_h)\ln(1-\rho)\right)$$

$$= -\langle n_h\rangle \ln\rho - (n - \langle n_h\rangle)\ln(1-\rho) = n\left(-\rho\ln\rho - (1-\rho)\ln(1-\rho)\right), \tag{C.4}$$

where we used the fact that for the Binomial distribution $\langle n_h\rangle = n\rho$. Then, using the definition of the function $H_2(x)$, the Shannon entropy of the Binomial distribution is

$$H[Bin_n(n_h)] = nH_2(\rho). \tag{C.5}$$

Note that since $H_2(\rho)$ is the entropy of a single Bernoulli random variable, the last result is the sum of n independent Bernoulli random variables. In other words, the Shannon entropy is additive in the Cartesian space of binary random variables.

C.1.2 The entropy of the B-model's Probability Distribution

In the B-model the probability of observing a configuration c_i with i pairs obtains as

$$P_n(c_i) = q_i = \frac{r^{\lfloor\frac{n}{2}\rfloor - i}}{c_n(r)}. \tag{C.6}$$

To find the ensemble entropy for the B-model, which has $W(n)$ distinct configurations, one can write

$$H_B[P_n] = -\sum_{i=1}^{W(n)} q_i \ln q_i = -\lfloor\frac{n}{2}\rfloor \ln r \sum_{i=1}^{W(n)} \frac{r^{\lfloor\frac{n}{2}\rfloor-i}}{c_n(r)} + \ln r \sum_{i=1}^{W(n)} \frac{i r^{\lfloor\frac{n}{2}\rfloor-i}}{c_n(r)} + \ln c_n(r) \sum_{i=1}^{W(n)} \frac{r^{\lfloor\frac{n}{2}\rfloor-i}}{c_n(r)}$$

$$= \left[\langle n_p\rangle_n - \lfloor\frac{n}{2}\rfloor\right] \ln r + \ln c_n(r), \tag{C.7}$$

where we used the definition of $\langle n_p \rangle_n$ and the fact that the probability distribution is normalised. In appendix (C.4), equation (C.32) finds $\ln c_n(r)$ as

$$\ln c_n(r) = \lfloor \frac{n}{2} \rfloor \ln r + \sum_{i=1}^{n-1} \ln(1 + \frac{\langle n_s \rangle_i}{r}). \tag{C.8}$$

Notice that, $\langle n_s \rangle_n$ is the expectation of the number of elements in stand-alone state for a system size n ($n_s = n - 2n_p$). Therefore, the Shannon entropy for the B-Model is equal to

$$H_B[P_n] = \langle n_p \rangle_n \ln r + \sum_{i=1}^{n-1} \ln(1 + \frac{\langle n_s \rangle_i}{r}). \tag{C.9}$$

C.1.3 The entropy of the C-model's Probability Distribution

In the C-Model, the probability of a configuration with i pairs and j heads, say c_{ij}, is written as

$$P_n(c_{ij}) = \frac{r^{\lfloor \frac{n}{2} \rfloor - i} \rho^j (1-\rho)^{n-2i-j}}{c_n(r)}, \tag{C.10}$$

and therefore, the ensemble entropy derives as

$$\begin{aligned} H_C[P_n] = -\frac{1}{c_n(r)} \sum_{i=1}^{W(n)} r^{\lfloor \frac{n}{2} \rfloor - i} \rho^j (1-\rho)^{n-2i-j} \Big[(\lfloor \frac{n}{2} \rfloor - i) \ln r + j \ln \rho \\ +(n-2i-j) \ln(1-\rho) - \ln c_n(r) \Big] \\ = -\lfloor \frac{n}{2} \rfloor \ln r + \langle n_p \rangle_n \ln r - \langle n_h \rangle_n \ln \rho - (n - 2\langle n_p \rangle_n - \langle n_h \rangle_n) \ln(1-\rho) + \ln c_n(r) \\ = \langle n_p \rangle_n \ln r + \sum_{i=1}^{n-1} \ln(1 + \frac{\langle n_s \rangle_i}{r}) - \langle n_h \rangle_n \ln \rho - (n - 2\langle n_p \rangle_n - \langle n_h \rangle_n) \ln(1-\rho), \end{aligned} \tag{C.11}$$

where we used the definition of $\langle n_p \rangle_n$, the probability distribution normalisation and equation (C.32) in appendix (C.4). Finally, we write $H_C[P_n]$ in terms of the B-model entropy, $H_B[P_n]$, and the remaining terms

$$H_C[P_n] = H[P_n(n_p)] - \langle n_h \rangle_n \ln \rho - (n - 2\langle n_p \rangle_n - \langle n_h \rangle_n) \ln(1-\rho). \tag{C.12}$$

Furthermore,

$$H_C[P_n] = H_B[P_n] - (n - 2\langle n_p\rangle_n)\left[\frac{\langle n_h\rangle_n}{n - 2\langle n_p\rangle_n}\ln\rho + (1 - \frac{\langle n_h\rangle_n}{n - 2\langle n_p\rangle_n})\ln(1-\rho)\right]$$

$$= H_B[P_n] - (n - 2\langle n_p\rangle_n)\left[(\frac{\langle n_h\rangle_n}{n - 2\langle n_p\rangle_n} - \rho)\ln\rho + (1 - \frac{\langle n_h\rangle_n}{n - 2\langle n_p\rangle_n} - (1-\rho))\ln(1-\rho)\right.$$
$$\left. + \rho\ln\rho + (1-\rho)\ln(1-\rho)\right]$$

$$= H_B[P_n] - (n - 2\langle n_p\rangle_n)\left[\frac{\langle n_h\rangle_n}{n - 2\langle n_p\rangle_n} - \rho\right]\ln\frac{\rho}{1-\rho} + (n - 2\langle n_p\rangle_n)H_2(\rho) \implies$$

$$= H_B[P_n] + (\rho n - \rho\langle 2n_p\rangle_n - \langle n_h\rangle_n)\ln\frac{\rho}{1-\rho} + (n - \langle 2n_p\rangle_n)H_2(\rho). \tag{C.13}$$

However, we will show that the term $\rho n - \rho\langle 2n_p\rangle_n - \langle n_h\rangle_n$ is equal to zero. To do that, first remember that the expectation of the Binomial distribution is $n\rho$ for the system size n. Hence,

$$\sum_{n_h=0}^{n-2n_p}\binom{n-2n_p}{n_h}n_h\rho^{n_h}(1-\rho)^{n-2n_p-n_h} = (n-2n_p)\rho, \tag{C.14}$$

and consequently for the C-model, $\langle n_h\rangle_n$ obtains as

$$\langle n_h\rangle_n = \sum_{n_p=0}^{\lfloor n/2\rfloor}\binom{n}{2n_p}(2n_p-1)!!\frac{r^{\lfloor n/2\rfloor - n_p}}{c_n(r)}\sum_{n_h=0}^{n-2n_p}\binom{n-2n_p}{n_h}n_h\rho^{n_h}(1-\rho)^{n-2n_p-n_h}$$

$$= \sum_{n_p=0}^{\lfloor n/2\rfloor}\binom{n}{2n_p}(2n_p-1)!!\frac{r^{\lfloor n/2\rfloor - n_p}}{c_n(r)}(n-2n_p)\rho = \rho n - \rho\langle 2n_p\rangle_n, \tag{C.15}$$

which assert the claim $\rho n - \rho\langle 2n_p\rangle_n - \langle n_h\rangle_n = 0$. Eventually, $H[P_n(n_p, n_h)]$ writes as

$$H_C[P_n] = H_B[P_n] + (n - \langle 2n_p\rangle_n)H_2(\rho). \tag{C.16}$$

C.1.4 Derivative of the Entropy of the B-model

Taking the derivative of $H_B[P_n]$ in equation (C.7) with respect to r finds

$$\frac{dH_B[P_n]}{dr} = \frac{d}{dr}\left[(\langle n_p\rangle_n - \lfloor\frac{n}{2}\rfloor)\ln r + \ln c_n(r)\right]$$

$$= \frac{d\langle n_p\rangle_n}{dr}\ln r + \frac{\langle n_p\rangle_n - \lfloor\frac{n}{2}\rfloor}{r} + \frac{d\ln c_n(r)}{dr} = \frac{d\langle n_p\rangle_n}{dr}\ln r, \tag{C.17}$$

where we used equation (B.76) in the last step. Continuing from definition $\langle n_p\rangle_n$

$$\frac{d\langle n_p\rangle_n}{dr} = \frac{d}{dr}\sum_{n_p=0}^{\lfloor n/2\rfloor}\binom{n}{2n_p}(2n_p-1)!!\frac{n_p\, r^{\lfloor n/2\rfloor - n_p}}{c_n(r)}$$

$$= \sum_{n_p=0}^{\lfloor n/2\rfloor}\binom{n}{2n_p}(2n_p-1)!!\frac{r^{\lfloor n/2\rfloor - n_p}}{c_n(r)}\left[\frac{n_p\lfloor n/2\rfloor - n_p^2}{r} - n_p\frac{d\ln c_n(r)}{dr}\right]$$

$$= \sum_{n_p=0}^{\lfloor n/2\rfloor}\binom{n}{2n_p}(2n_p-1)!!\frac{r^{\lfloor n/2\rfloor - n_p}}{c_n(r)}\left[\frac{n_p\lfloor n/2\rfloor - n_p^2}{r} - \frac{n_p\lfloor n/2\rfloor - n_p\langle n_p\rangle_n}{r}\right]$$

$$= \langle n_p\rangle_n^2 - \langle n_p^2\rangle_n, \tag{C.18}$$

where again we used equation (B.76) in the step before the last one. Plunging the last result in the derivative of $H_B[P_n]$ we get

$$\frac{dH_B[P_n]}{dr} = (\langle n_p\rangle_n^2 - \langle n_p^2\rangle_n)\ln r. \tag{C.19}$$

C.2 Finding the Asymptotic of $H_C[P_n]$

For $1 \ll n$, equation (3.91) gives the asymptotic of $\langle 2n_p\rangle$ and $\langle n_s\rangle$ as

$$\langle 2n_p\rangle_n = n\mathrm{e}^{-\sqrt{\frac{r}{n}}}, \qquad \langle n_s\rangle_n = n(1-\mathrm{e}^{-\sqrt{\frac{r}{n}}}). \tag{C.20}$$

Therefore, we can write the equation (4.24) as

$$H_B[P_n] = \langle n_p\rangle_n \ln r + \sum_{i=1}^{n-1}\ln(1+\frac{\langle n_s\rangle_i}{r}) \sim n\mathrm{e}^{-\sqrt{\frac{r}{n}}}\ln r + \sum_{i=1}^{n-1}\ln(1+\frac{i(1-\mathrm{e}^{-\sqrt{\frac{r}{i}}})}{r})$$

$$\sim n\mathrm{e}^{-\sqrt{\frac{r}{n}}}\ln r + \sum_{i=1}^{n-1}\ln(1+\sqrt{\frac{i}{r}}) \sim n\mathrm{e}^{-\sqrt{\frac{r}{n}}}\ln r + \sum_{i=1}^{n-1}\sum_{k\geq 1}\frac{(-1)^{k+1}}{k}\left(\sqrt{\frac{i}{r}}\right)^k$$

$$\sim n\mathrm{e}^{-\sqrt{\frac{r}{n}}}\ln r + \sum_{k\geq 1}\frac{(-1)^{k+1}}{kr^{k/2}}\sum_{i=1}^{n-1}(i)^{k/2}$$

$$\sim n\mathrm{e}^{-\sqrt{\frac{r}{n}}}\ln r + 2\sum_{k\geq 1}\frac{(-1)^{k+1}}{kr^{k/2}}\frac{n^{\frac{k+2}{2}}}{k+2}, \tag{C.21}$$

where we used Faulhaber's formula to find the asymptotic leading term in the last step. Continuing

$$H_B[P_n] \sim n\ln r \times \mathrm{e}^{-\sqrt{\frac{r}{n}}} + 2n\sum_{k\geq 1}\frac{(-1)^{k+1}}{k(k+2)}\left(\sqrt{\frac{n}{r}}\right)^k$$

$$\sim n\ln r \times \left(1-\sqrt{\frac{r}{n}}+O(\frac{1}{n})\right) + n\sum_{k\geq 1}(-1)^{k+1}\left(\sqrt{\frac{n}{r}}\right)^k(\frac{1}{k}-\frac{1}{k+2})$$

$$\sim n\ln r - \sqrt{rn}\ln r + n\sum_{k\geq 1}\frac{(-1)^{k+1}}{k}\left(\sqrt{\frac{n}{r}}\right)^k - n\sum_{k\geq 1}\frac{(-1)^{k+1}}{k+2}\left(\sqrt{\frac{n}{r}}\right)^k + O(1)$$

$$\sim n\ln r - \sqrt{rn}\ln r + n\ln\sqrt{\frac{n}{r}} - n\sum_{k\geq 3}\frac{(-1)^{k-1}}{k}\left(\sqrt{\frac{n}{r}}\right)^{k-2} + O(1)$$

$$\sim \frac{n}{2}\ln\frac{n}{r} + n\ln r - n\left(\sqrt{\frac{n}{r}}\right)^{-2}\sum_{k\geq 3}\frac{(-1)^{k+1}}{k}\left(\sqrt{\frac{n}{r}}\right)^k + O(\sqrt{n})$$

$$\sim \frac{n}{2}\ln\frac{n}{r} + n\ln r - r\sum_{k\geq 1}\frac{(-1)^{k+1}}{k}\left(\sqrt{\frac{n}{r}}\right)^k + r\left(\sqrt{\frac{n}{r}}-\frac{n}{2r}\right) + O(\sqrt{n})$$

$$\sim \frac{n}{2}\ln\frac{n}{r} + n\ln r - r\ln\sqrt{\frac{n}{r}} - \frac{n}{2} + O(\sqrt{n})$$

$$\sim \frac{n}{2}\ln\frac{n}{r} + n(\ln r - 1/2) + O(\sqrt{n}). \tag{C.22}$$

C.3 Finding the Asymptotic of $H[P_n(X_1,\ldots,X_k)]$

First, using the following identity

$$l\binom{k}{l} = k\binom{k-1}{l-1}, \tag{C.23}$$

and Binomial expansion, we have

$$\sum_{l=0}^{k}\binom{k}{l}lx^l = k\sum_{l=1}^{k}\binom{k-1}{l-1}x^l$$

$$= k \sum_{l=0}^{k-1} \binom{k-1}{l} x^{l+1} = kx(1+x)^{k-1}. \tag{C.24}$$

Next, following equation (4.54), and using the last result, we have

$$H[P_n(X_1, \ldots, X_k)] \sim \sum_{l=0}^{k} \binom{k}{l} l \sqrt{\frac{r}{n}} \mathrm{e}^{-l\sqrt{\frac{r}{n}}} \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)^{k-l}$$

$$- \sum_{l=0}^{k} \binom{k}{l} (k-l) \mathrm{e}^{-l\sqrt{r/n}} \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)^{k-l} \ln\left(1 - \mathrm{e}^{-\sqrt{\frac{r}{n}}}\right)$$

$$= \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)^{k} \left(\sqrt{\frac{r}{n}} + \ln\left(1 - \mathrm{e}^{-\sqrt{\frac{r}{n}}}\right)\right) \sum_{l=0}^{k} \binom{k}{l} l \left(\mathrm{e}^{\sqrt{\frac{r}{n}}} - 1\right)^{-l}$$

$$-k \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)^{k} \ln\left(1 - \mathrm{e}^{-\sqrt{\frac{r}{n}}}\right) \sum_{l=0}^{k} \binom{k}{l} \left(\mathrm{e}^{\sqrt{\frac{r}{n}}} - 1\right)^{-l}$$

$$= \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)^{k} \left(\sqrt{\frac{r}{n}} + \ln\left(1 - \mathrm{e}^{-\sqrt{\frac{r}{n}}}\right)\right) \times k(\mathrm{e}^{\sqrt{\frac{r}{n}}} - 1)^{-1} \left(1 + (\mathrm{e}^{\sqrt{\frac{r}{n}}} - 1)^{-1}\right)^{k-1}$$

$$-k \left(1 - \mathrm{e}^{-\sqrt{r/n}}\right)^{k} \ln\left(1 - \mathrm{e}^{-\sqrt{\frac{r}{n}}}\right) \times \left(1 + (\mathrm{e}^{\sqrt{\frac{r}{n}}} - 1)^{-1}\right)^{k}$$

$$= k\mathrm{e}^{-\sqrt{r/n}} \left(\sqrt{\frac{r}{n}} + \ln\left(1 - \mathrm{e}^{-\sqrt{\frac{r}{n}}}\right)\right) - k \ln\left(1 - \mathrm{e}^{-\sqrt{\frac{r}{n}}}\right)$$

$$= k \left(\sqrt{\frac{r}{n}} \mathrm{e}^{-\sqrt{r/n}} - (1 - \mathrm{e}^{-\sqrt{r/n}}) \ln\left(1 - \mathrm{e}^{-\sqrt{\frac{r}{n}}}\right)\right). \tag{C.25}$$

This is exactly k times the entropy of a single element when we directly use equation (3.130) and write $H[P_n(X_1)]$. Therefore,

$$H[P_n(X_1, \ldots, X_k)] \sim kH_B[P_n(X_1)]. \tag{C.26}$$

Using the asymptotic expansion of the exponential function

$$\mathrm{e}^{-\sqrt{\frac{r}{n}}} = 1 - \sqrt{\frac{r}{n}} + O(\frac{1}{n}), \tag{C.27}$$

simplifies the asymptotic leading term of $H[P_n(X_1, \ldots, X_k)]$

$$H[P_n(X_1, \ldots, X_k)] = k \left(\sqrt{\frac{r}{n}}(1 - \sqrt{\frac{r}{n}}) - \sqrt{\frac{r}{n}} \ln \sqrt{\frac{r}{n}}\right) + O(\frac{1}{n}) \implies$$

$$H[P_n(X_1,\ldots,X_k)] = k\sqrt{\frac{r}{n}}\left[1-\ln\sqrt{\frac{r}{n}}\right]+O(\frac{1}{n}). \tag{C.28}$$

C.4 The Governing Recursive Relation of $\ln c_n(r)$

To find $\ln c_n(r)$ in terms of the average number of pairs for odd and even system sizes, say $\langle n_p\rangle_{2n}$ and $\langle n_p\rangle_{2n-1}$, equation (3.88) finds the dependence of ratios of the normalisation constants to the average number of pairs, while equation (3.46) finds the recursive relation for the normalisation constant. Combining both, we find

$$\begin{cases} r\frac{c_{2n-1}(r)}{c_{2n}(r)} = 1-\frac{2\langle n_p\rangle_{2n}}{2n} \\ \frac{c_{2n-2}(r)}{c_{2n-1}(r)} = 1-\frac{2\langle n_p\rangle_{2n-1}}{2n-1} \end{cases}. \tag{C.29}$$

Recall that the subscript in expectation $\langle . \rangle_n$ represents the system size. Next, we write the logarithm for the normalisation constant by using equation (3.46). For even system sizes and using equation (C.29), we have

$$\begin{aligned} \ln c_{2n}(r) &= \ln\left(rc_{2n-1}(r)+(2n-1)c_{2n-2}(r)\right) \\ &= \ln r+\ln c_{2n-1}(r)+\ln(1+\frac{2n-1}{r}\times\frac{c_{2n-2}(r)}{c_{2n-1}(r)}) \\ &= \ln r+\ln c_{2n-1}(r)+\ln(1+\frac{(2n-1)-\langle 2n_p\rangle_{2n-1}}{r}) \\ &= \ln r+\ln c_{2n-1}(r)+\ln(1+\frac{\langle n_s\rangle_{2n-1}}{r}). \end{aligned} \tag{C.30}$$

Similarly, for odd system sizes, one finds

$$\ln c_{2n-1}(r) = \ln c_{2n-2}(r)+\ln(1+\frac{\langle n_s\rangle_{2n-2}}{r}). \tag{C.31}$$

Hence, there are two different equations for odd and even system sizes. Applying the last two equations iteratively on $\ln c_n(r)$ for $n-1$ times and considering the fact that $c_1(r)=1$, one obtains

$$\begin{aligned} \ln c_n(r) &= \lfloor\frac{n}{2}\rfloor\ln r+\ln c_1(r)+\sum_{i=1}^{n-1}\ln(1+\frac{\langle n_s\rangle_i}{r}) \\ &= \lfloor\frac{n}{2}\rfloor\ln r+\sum_{i=1}^{n-1}\ln(1+\frac{\langle n_s\rangle_i}{r}). \end{aligned} \tag{C.32}$$

C.5 Power Series Expansion for $I_n(S_{n-1}, S_1)$

The power series expansion of $(1-x)\ln(1-x)$ for a random variable $x \in [0,1]$ derives as

$$(1-x)\ln(1-x) = -(1-x)\sum_{k\geq 1}\frac{x^k}{k} = \sum_{k\geq 1}\frac{x^{k+1}-x^k}{k} \implies$$

$$(1-x)\ln(1-x) = -x + \sum_{k\geq 2}\frac{x^k}{k(k-1)}. \tag{C.33}$$

Consequently, taking the expectation with respect to an arbitrary distribution finds

$$\langle(1-x)\ln(1-x)\rangle = -\langle x\rangle + \sum_{k\geq 2}\frac{\langle x^k\rangle}{k(k-1)}. \tag{C.34}$$

Let us define $y = 1-x$. Then, using the last result, $H_2(x)$ drives as

$$\begin{aligned} H_2(x) = -x\ln x - y\ln y &= -(1-y)\ln(1-y) - (1-x)\ln(1-x) \\ &= x + y - \sum_{k\geq 2}\frac{x^k+y^k}{k(k-1)} \qquad \text{(eq. C.33)} \\ &= 1 - \sum_{k\geq 2}\frac{x^k+y^k}{k(k-1)}. \qquad \text{(C.35)} \end{aligned}$$

At the same time,

$$\begin{aligned} \langle H_2(x)\rangle &= \langle -(1-y)\ln(1-y) - (1-x)\ln(1-x)\rangle \\ &= \langle 1 - \sum_{k\geq 2}\frac{x^k+y^k}{k(k-1)}\rangle \\ &= 1 - \sum_{k\geq 2}\frac{\langle x^k\rangle + \langle y^k\rangle}{k(k-1)}. \qquad \text{(C.36)} \end{aligned}$$

Combining the results from equations (C.35) and (C.36), and using result for $I_n(S_{n-1}, S_1)$ in equation (4.66), we derive the power series expansion as

$$I_n(S_{n-1}, S_1) = \sum_{k\geq 2}\frac{\langle(\frac{2n_p}{n})^k\rangle - \langle\frac{2n_p}{n}\rangle^k + \langle(\frac{n_s}{n})^k\rangle - \langle\frac{n_s}{n}\rangle^k}{k(k-1)}. \tag{C.37}$$

Notice that in practice since $x \in [0,1]$ the power series converge very fast, *e.g.*, the terms for $k \geq 4$ are at least one order of magnitude smaller than the first two terms.

When the system size increases, we expect the mutual information approaches zero, since the information about S_1 reduces uncertainty when the whole system is larger. To show that, we need to find the asymptotic form of $I_n(S_{n-1}, S_1)$ from its power series expansion.

Using equation (B.82) one derives

$$\begin{aligned}\langle(\frac{2n_p}{n})^k\rangle - \langle\frac{2n_p}{n}\rangle^k &= \mathrm{e}^{-k\sqrt{r/n}}\left(1+\frac{k(k-1)}{n}(\mathrm{e}^{\sqrt{r/n}}-1)\right) - \mathrm{e}^{-k\sqrt{r/n}} + O(\frac{1}{n^2}) \\ &= \frac{k(k-1)\mathrm{e}^{-k\sqrt{r/n}}}{n}(\mathrm{e}^{\sqrt{r/n}}-1) + O(\frac{1}{n^2}), \qquad \text{(C.38)}\end{aligned}$$

and equation (B.86) finds

$$\langle(\frac{n_s}{n})^k\rangle - \langle\frac{n_s}{n}\rangle^k \sim \frac{k(k-1)(\mathrm{e}^{\sqrt{r/n}}-1)\mathrm{e}^{-2\sqrt{r/n}}}{n}\left(1-\mathrm{e}^{-\sqrt{r/n}}\right)^{k-1} + O(\frac{1}{n^2}). \quad \text{(C.39)}$$

Putting all together,

$$\begin{aligned}I_n(S_{n-1}, S_1) &\sim \sum_{k\geq 2}\frac{\frac{k(k-1)\mathrm{e}^{-k\sqrt{r/n}}}{n}(\mathrm{e}^{\sqrt{r/n}}-1) + \frac{k(k-1)(\mathrm{e}^{\sqrt{r/n}}-1)\mathrm{e}^{-2\sqrt{r/n}}}{n}\left(1-\mathrm{e}^{-\sqrt{r/n}}\right)^{k-1}}{k(k-1)} + O(\frac{1}{n^2}) \\ &\sim \frac{(\mathrm{e}^{\sqrt{r/n}}-1)}{n}\left[\sum_{k\geq 2}\mathrm{e}^{-k\sqrt{r/n}} + \mathrm{e}^{-2\sqrt{r/n}}\sum_{k\geq 1}\left(1-\mathrm{e}^{-\sqrt{r/n}}\right)^k\right] + O(\frac{1}{n^2}) \\ &\sim \frac{(\mathrm{e}^{\sqrt{r/n}}-1)}{n}\left[\frac{\mathrm{e}^{-2\sqrt{r/n}}}{1-\mathrm{e}^{-\sqrt{r/n}}} + \frac{\mathrm{e}^{-2\sqrt{r/n}}\left(1-\mathrm{e}^{-\sqrt{r/n}}\right)}{\mathrm{e}^{-\sqrt{r/n}}}\right] + O(\frac{1}{n^2}) \\ &\sim \frac{1}{n}\left[1+\mathrm{e}^{-2\sqrt{r/n}} - \mathrm{e}^{-\sqrt{r/n}}\right] + O(\frac{1}{n^2}) \sim \frac{1}{n} + O(\frac{1}{n^{3/2}}). \qquad \text{(C.40)}\end{aligned}$$

Note that, as it was explained in section (B.9), numerical estimation of asymptotic leading term of $\langle(\frac{2n_p}{n})^k\rangle$ decreases exponentially fast.

C.6 Rényi Entropy

For B-Model, Rényi Entropy writes as

$$H_\alpha[P_n] = -\frac{1}{1-\alpha}\ln\left(\sum_{i=1}^{W(n)} q_i^\alpha\right) = -\frac{1}{1-\alpha}\ln\left(\sum_{n_p=0}^{n}\frac{\binom{n}{2n_p}(2n_p-1)!!r^{\alpha(\lfloor\frac{n}{2}\rfloor-n_p)}}{c_n^\alpha(r)}\right)$$

$$= -\frac{1}{1-\alpha}\ln\left(\frac{1}{c_n^\alpha(r)}\sum_{n_p=0}^{n}\binom{n}{2n_p}(2n_p-1)!!(r^\alpha)^{\lfloor\frac{n}{2}\rfloor-n_p}\right) = -\frac{1}{1-\alpha}\ln\left(\frac{c_n(r^\alpha)}{c_n^\alpha(r)}\right)$$

$$= -\frac{\ln c_n(r^\alpha) - \alpha\ln c_n(r)}{1-\alpha}. \tag{C.41}$$

C.7 Tsallis Entropy

For the B-Model, Tsallis Entropy writes as

$$H_q[P_n] = \frac{1}{q-1}\left(1-\sum_{i=1}^{W(n)} p_i^q\right) = \frac{1}{q-1}\left(1-\sum_{n_p=0}^{n}\frac{\binom{n}{2n_p}(2n_p-1)!!r^{q(\lfloor\frac{n}{2}\rfloor-n_p)}}{c_n^q(r)}\right)$$

$$= \frac{1}{q-1}\left(1-\frac{1}{c_n^q(r)}\sum_{n_p=0}^{n}\binom{n}{2n_p}(2n_p-1)!!(r^q)^{\lfloor\frac{n}{2}\rfloor-n_p}\right)$$

$$= \frac{1}{q-1}\left(1-\frac{c_n(r^q)}{c_n^q(r)}\right). \tag{C.42}$$

Similarly, for the C-model, we have

$$H_q[P_n] = \frac{1}{q-1}\left(1-\sum_{n_p=0}^{n}\frac{\binom{n}{2n_p}(2n_p-1)!!r^{q(\lfloor\frac{n}{2}\rfloor-n_p)}}{c_n^q(r)}\sum_{n_h=0}^{n-2n_p}\binom{n-2n_p}{n_h}\rho^{qn_h}(1-\rho)^{q(n-2n_p-n_h)}\right)$$

$$= \frac{1}{q-1}\left(1-\frac{1}{c_n^q(r)}\sum_{n_p=0}^{n}\binom{n}{2n_p}(2n_p-1)!!(r^q)^{\lfloor\frac{n}{2}\rfloor-n_p}\sum_{n_h=0}^{n-2n_p}\binom{n-2n_p}{n_h}(\rho^q)^{n_h}((1-\rho)^q)^{n-2n_p-n_h}\right). \tag{C.43}$$

The second sum in the last line is written as

$$\sum_{n_h=0}^{n-2n_p} \binom{n-2n_p}{n_h} (\rho^q)^{n_h} ((1-\rho)^q)^{n-2n_p-n_h} = [\rho^q + (1-\rho)^q]^{n-2n_p} . \tag{C.44}$$

So

$$H_q[P_n] = \frac{1}{q-1} \left(1 - \sum_{n_p=0}^{n} \frac{\binom{n}{2n_p}(2n_p-1)!! r^{q(\lfloor \frac{n}{2} \rfloor - n_p)} \left[\rho^q + (1-\rho)^q\right]^{n-2n_p}}{c_n^q(r)} \right)$$

$$= \frac{1}{q-1} \left(1 - \frac{1}{c_n^q(r)} \sum_{n_p=0}^{n} \binom{n}{2n_p} (2n_p-1)!! (r^q \left[\rho^q + (1-\rho)^q\right]^2)^{\lfloor \frac{n}{2} \rfloor - n_p} \right)$$

$$= \frac{1}{q-1} \left(1 - \frac{c_n(r^q \left[\rho^q + (1-\rho)^q\right]^2)}{c_n^q(r)} \right) . \tag{C.45}$$

C.8 Finding the Shannon Entropy of Pairing Time Series

We write the Shannon entropy over an ensemble of strings with length n as

$$H_L(\boldsymbol{X}_n) = -\sum_{\boldsymbol{X}_n \in \Lambda_n} P(\boldsymbol{X}_n) \log P(\boldsymbol{X}_n)$$

$$= -\sum_{\boldsymbol{X}_n \in \Lambda_n^1} P(\boldsymbol{X}_n) \log P(\boldsymbol{X}_n) - \sum_{\boldsymbol{X}_n \in \Lambda_n^2} P(\boldsymbol{X}_n) \log P(\boldsymbol{X}_n) \qquad (\Lambda_n = \Lambda_n^1 \cup \Lambda_n^2)$$

$$= -\sum_{X_n \in \overleftarrow{X_n}} P(X_n) \sum_{\boldsymbol{X}_{n-1} \in \Gamma_{n-1}^1} P_E(\boldsymbol{X}_{n-1}|X_n) \log \left[P(X_n) P_E(\boldsymbol{X}_{n-1}|X_n)\right]$$

$$- \sum_{X_n \in \overrightarrow{X_n}} P(X_n) \sum_{\boldsymbol{X}_{n-1} \in \Gamma_{n-1}^1} P_E(\boldsymbol{X}_{n-1}|X_n) \log \left[P(X_n) P_E(\boldsymbol{X}_{n-1}|X_n)\right] , \tag{C.46}$$

where in the last step, we used the extension of the joint probability in equations (4.105) and (4.114) to switch to Cartesian spaces. Continuing,

$$H_L(\boldsymbol{X}_n) = -\sum_{X_n \in \overleftarrow{X_n}} \sum_{\boldsymbol{X}_{n-1} \in \Gamma_{n-1}^1} P(X_n) P_E(\boldsymbol{X}_{n-1}|X_n) \left[\log P(X_n) + \log P_E(\boldsymbol{X}_{n-1}|X_n)\right]$$

$$
-\sum_{X_n\in\overrightarrow{X_n}}\sum_{\boldsymbol{X}_{n-1}\in\ \Gamma^2_{n-1}} P(X_n)P_E(\boldsymbol{X}_{n-1}|X_n)\left[\log P(X_n)+\log P_E(\boldsymbol{X}_{n-1}|X_n)\right]
$$

$$
\begin{aligned}
&= -\sum_{X_n\in\overleftarrow{X_n}} P(X_n)\log P(X_n)\sum_{\boldsymbol{X}_{n-1}\in\ \Gamma^1_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n)\\
&-\sum_{X_n\in\overleftarrow{X_n}} P(X_n)\sum_{\boldsymbol{X}_{n-1}\in\ \Gamma^1_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n)\log P_E(\boldsymbol{X}_{n-1}|X_n)\\
&-\sum_{X_n\in\overrightarrow{X_n}} P(X_n)\log P(X_n)\sum_{\boldsymbol{X}_{n-1}\in\ \Gamma^2_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n)\\
&-\sum_{X_n\in\overrightarrow{X_n}} P(X_n)\sum_{\boldsymbol{X}_{n-1}\in\Gamma^2_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n)\log P_E(\boldsymbol{X}_{n-1}|X_n). \qquad \text{(C.47)}
\end{aligned}
$$

The normalisation condition requires that the sum over conditionals must be equal to one. Therefore

$$
\begin{aligned}
H_L(\boldsymbol{X}_n) &= -\sum_{X_n\in\overleftarrow{X_n}} P(X_n)\log P(X_n) - \sum_{X_n\in\overrightarrow{X_n}} P(X_n)\log P(X_n)\\
&-\sum_{X_n\in\overleftarrow{X_n}} P(X_n)\sum_{\boldsymbol{X}_{n-1}\in\ \Gamma^1_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n)\log P_E(\boldsymbol{X}_{n-1}|X_n)\\
&-\sum_{X_n\in\overrightarrow{X_n}} P(X_n)\sum_{\boldsymbol{X}_{n-1}\in\Gamma^2_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n)\log P_E(\boldsymbol{X}_{n-1}|X_n)\\
&= -\sum_{X_n\in\overleftarrow{X_n}\cup\overrightarrow{X_n}} P(X_n)\log P(X_n)\\
&-\sum_{X_n\in\overleftarrow{X_n}} P(X_n)\sum_{\boldsymbol{X}_{n-1}\in\ \Gamma^1_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n)\log P_E(\boldsymbol{X}_{n-1}|X_n)\\
&-\sum_{X_n\in\overrightarrow{X_n}} P(X_n)\sum_{\boldsymbol{X}_{n-1}\in\Gamma^2_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n)\log P_E(\boldsymbol{X}_{n-1}|X_n). \qquad \text{(C.48)}
\end{aligned}
$$

Using equations (4.107) and (4.116), we extend the range of X_n in the second and the third sums

$$
\begin{aligned}
H_L(\boldsymbol{X}_n) &= -\sum_{X_n\in\overleftarrow{X_n}\cup\overrightarrow{X_n}} P(X_n)\log P(X_n)\\
&-\sum_{X_n\in\overleftarrow{X_n}\cup\overrightarrow{X_n}} P(X_n)\sum_{\boldsymbol{X}_{n-1}\in\ \Gamma^1_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n)\log P_E(\boldsymbol{X}_{n-1}|X_n)\\
&-\sum_{X_n\in\overleftarrow{X_n}\cup\overrightarrow{X_n}} P(X_n)\sum_{\boldsymbol{X}_{n-1}\in\Gamma^2_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n)\log P_E(\boldsymbol{X}_{n-1}|X_n)
\end{aligned}
$$

$$= - \sum_{X_n \in \overleftarrow{X_n} \cup \overrightarrow{X_n}} P(X_n) \log P(X_n)$$

$$- \sum_{X_n \in \overleftarrow{X_n} \cup \overrightarrow{X_n}} P(X_n) \sum_{\boldsymbol{X}_{n-1} \in \Gamma^2_{n-1} \cup \Gamma^2_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n) \log P_E(\boldsymbol{X}_{n-1}|X_n). \tag{C.49}$$

Theorem (4.2.2) showed that $\Gamma^1_{n-1} \cup \Gamma^2_{n-1}$ is equal to Λ_{n-1}, so

$$H_L(\boldsymbol{X}_n) = - \sum_{X_n \in \overleftarrow{X_n} \cup \overrightarrow{X_n}} P(X_n) \log P(X_n)$$

$$- \sum_{X_n \in \overleftarrow{X_n} \cup \overrightarrow{X_n}} P(X_n) \sum_{\boldsymbol{X}_{n-1} \in \Lambda_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n) \log P_E(\boldsymbol{X}_{n-1}|X_n). \tag{C.50}$$

Calling

$$H_L(X_n) \equiv - \sum_{X_n \in \overleftarrow{X_n} \cup \overrightarrow{X_n}} P(X_n) \log P(X_n), \tag{C.51}$$

as *the present entropy*, and

$$H_L(\boldsymbol{X}_{n-1}|X_n) \equiv - \sum_{X_n \in \overleftarrow{X_n} \cup \overrightarrow{X_n}} P(X_n) \sum_{\boldsymbol{X}_{n-1} \in \Lambda_{n-1}} P_E(\boldsymbol{X}_{n-1}|X_n) \log P_E(\boldsymbol{X}_{n-1}|X_n), \tag{C.52}$$

as *the past conditional entropy* on present, we get

$$H_L(\boldsymbol{X}_n) = H_L(X_n) + H_L(\boldsymbol{X}_{n-1}|X_n). \tag{C.53}$$

C.9 Finding the Ratios of $\Omega_2(L)$

Clearly, we could use the asymptotic leading term of $\Omega_2(L)$ to simplify the above ratios. Instead, we use an elementary method that is easier to handle and is applicable to any other recursive relation that we may see in future. Let us start with the recursive relation for $\Omega_2(L)$

$$\Omega_2(L) = 2\Omega_2(L-1) + (L-1)\Omega_2(L-2) \implies$$

$$1 = 2\frac{\Omega_2(L-1)}{\Omega_2(L)} + (L-1)\frac{\Omega_2(L-2)}{2\Omega_2(L-1) + (L-1)\Omega_2(L-2)} \implies$$

$$1 = 2\frac{\Omega_2(L-1)}{\Omega_2(L)} + \frac{L-1}{2\frac{\Omega_2(L-1)}{\Omega_2(L-2)} + (L-1)} \tag{C.54}$$

Defining

$$\rho_L = \frac{\Omega_2(L-1)}{\Omega_2(L)}, \qquad \rho_{L-1} = \frac{\Omega_2(L-2)}{\Omega_2(L-1)}, \tag{C.55}$$

equation (C.54) writes as

$$1 = 2\rho_L + \frac{L-1}{2/\rho_{L-1} + (L-1)}. \tag{C.56}$$

Assuming for increasing L the sequence ρ_L has a limit, the fix-point of the above relation requires $\rho_L \sim \rho_{L-1}$, we get the following quadratic equation

$$(L-1)\rho_L^2 + 2\rho_L - 1 = 0, \tag{C.57}$$

such that its positive solution for $1 \ll L$ is

$$\rho_L = \frac{\sqrt{L}-1}{L-1} \sim \frac{1}{\sqrt{L}}. \tag{C.58}$$

When we reorder the terms that define ρ_L as

$$\frac{\Omega_2(L-1)}{\Omega_2(L)} = \rho_L \sim \frac{1}{\sqrt{L}}, \quad (L-1)\frac{\Omega_2(L-2)}{\Omega_2(L)} = 1 - 2\rho_L \sim 1 - \frac{2}{\sqrt{L}}. \tag{C.59}$$

Appendix

D

Statistical Mechanics

D.1 Finding the Microcanonical Maximum

For $1 << N$, taking the logarithm of $\Omega_2(E,N,k)$ in equation (5.5) and using the Stirling approximation we derive

$$\ln \Omega_2(E,N,k) = N \ln N - N - k \ln 2 - k \ln k - \left(\frac{N-E}{2}\right) \ln \left(\frac{N-E}{2} - k\right)$$

$$- \left(\frac{N+E}{2}\right) \ln \left(\frac{N+E}{2} - k\right) + k \ln \left[(\frac{N}{2} - k)^2 - \frac{E^2}{4} \right] + N - 2k \implies \quad \text{(D.1)}$$

To find the maximum, taking the derivative of $\ln \Omega_2(E,N,k)$ with respect to k finds

$$\frac{d \ln \Omega_2(E,N,k)}{dk} = -\ln 2 - 1 - \ln k - 1 + \frac{\frac{N-E}{2}}{\frac{N-E}{2} - k} + \frac{\frac{N+E}{2}}{\frac{N+E}{2} - k}$$

$$+ \ln \left[(\frac{N}{2} - k)^2 - \frac{E^2}{4} \right] + \frac{2k(k - \frac{N}{2})}{(\frac{N}{2} - k)^2 - \frac{E^2}{4}}$$

$$= -\ln 2k - 2 + \ln \left[(\frac{N}{2} - k)^2 - \frac{E^2}{4} \right] + \frac{\frac{N^2 - E^2}{2} - kN + 2k^2 - kN}{(\frac{N}{2} - k)^2 - \frac{E^2}{4}}$$

$$= -\ln 2k - 2 + \ln\left[(\frac{N}{2} - k)^2 - \frac{E^2}{4}\right] + \frac{2\left((\frac{N}{2} - k)^2 - \frac{E^2}{4}\right)}{(\frac{N}{2} - k)^2 - \frac{E^2}{4}}$$

$$= -\ln 2k + \ln\left[(\frac{N}{2} - k)^2 - \frac{E^2}{4}\right] = 0. \tag{D.2}$$

Next, equating the last equation to zero to find its zero results in the following quadratic equation

$$k^{*2} - (N+2)k^* + \frac{N^2 - E^2}{4} = 0, \tag{D.3}$$

with solutions as

$$k^* = \frac{N}{2} + 1 \pm \frac{1}{2}\sqrt{E^2 + 4N + 4}. \tag{D.4}$$

Taking $u = \frac{E}{N}$ as a constant and considering the fact that $1 \ll N$, the asymptotic leading term of k^* obtains as

$$k^* = \frac{N}{2}(1 \pm u). \tag{D.5}$$

Between these two solutions, one of them can be the acceptable one, depending on the sign of total energy or u. We can divide energy values into three regions

1. $u > 0$: For $u = 1$, $k^* = N$ or 0. From a combinatorial argument we are sure that for $u = 1$ all the elements must be in *head* state. So the only possible solution corresponds to $k^* = 0$. It means that for $u > 0$, it must be $k^* = \frac{N}{2}(1 - u)$.

2. $u < 0$: Similar to the positive case, $k^* = 0$ is the only possible solution for $u = -1$ (all in *tail* state). It means that for $u < 0$, $k^* = \frac{N}{2}(1+u)$ is the correct solution.

3. $u = 0$: This is correspond to $E = 0$ which implies

$$k^* = \frac{N}{2} \pm \sqrt{N+4} + 1 \sim \frac{N}{2}.$$

Finally, we need to show that the bulk of the sum

$$\Omega_2(E, N) = \sum_{k=0}^{\lfloor N/2 \rfloor} \Omega_2(E, N, k), \tag{D.6}$$

is concentrated around $\Omega_2(E, N, k^*)$. In other words,

$$\Omega_2(E, N, k^*) \approx \sum_{k=0}^{\lfloor N/2 \rfloor} \Omega_2(E, N, k) = \Omega_2(E, N). \tag{D.7}$$

To do that, for a fixed N and E, we must ask what proportion of configuration numbers does belong to k^* in comparison to the others? Since

$$k^* = \begin{cases} \frac{N}{2}(1-u), u \geq 0 \\ \\ \frac{N}{2}(1+u), u < 0 \end{cases} = \begin{cases} \frac{N-E}{2}, E \geq 0 \\ \\ \frac{N+E}{2}, E < 0 \end{cases}, \tag{D.8}$$

plugging back k^* into $\Omega_2(N, E, k)$, we get

$$\Omega_2(E_\pm, N, k^*) = \begin{cases} \frac{N!}{2^{\frac{N-E_+}{2}}(\frac{N-E_+}{2})!(E_+)!}, E_+ \geq 0 \\ \\ \frac{N!}{2^{\frac{N+E_-}{2}}(\frac{N+E_-}{2})!(-E_-)!}, E_- < 0 \end{cases} \tag{D.9}$$

Defining $E_* = E_+ = -|E_-|$, makes $k^* = \frac{N-E_*}{2}$ and

$$\Omega_2(E_*, N, k^*) = \frac{N!}{2^{\frac{N-E_*}{2}}(\frac{N-E_*}{2})!E_*!}. \tag{D.10}$$

where $u_* = \frac{E_*}{N} \geq 0$. Deviating from k^* for both negative and positive E corresponds to $k = k^* - \Delta k = \frac{N-E_*}{2} - \Delta k$. Surly $\Delta k \geq 0$, otherwise k is outside of the valid range, say $2k \pm E \leq N$. Knowing that, we rewrite Ω_2 as

$$\Omega_2(E_*, N, k^* - \Delta k) = \frac{N!}{2^{\frac{N-E_*}{2}-\Delta k}(\frac{N-E_*}{2} - \Delta k)!(\Delta k)!(E_* + \Delta k)!}. \tag{D.11}$$

Therefore the ratio of volumes for $k = k^* - \Delta k$ to k^* is

$$\frac{\Omega_2(E_*, N, k^* - \Delta k)}{\Omega_2(E_*, N, k^*)} = \frac{1}{2^{-\Delta k}(\Delta k)!} \times \frac{(\frac{N-E_*}{2})!}{(\frac{N-E_*}{2} - \Delta k)!} \times \frac{E_*!}{(E_* + \Delta k)!}. \tag{D.12}$$

Using the Stirling approximation lets us simplify it as

$$\frac{\Omega_2(E_*, N, k^* - \Delta k)}{\Omega_2(E_*, N, k^*)}$$

$$= \frac{1}{2^{-\Delta k}(\Delta k)!} \times \frac{(\frac{1-u_*}{2})^{\frac{N}{2}(1-u_*)}}{(\frac{1-u_*}{2} - \frac{\Delta k}{N})^{\frac{N}{2}(1-u_*)-\Delta k}} \times \frac{1}{(1+\frac{\Delta k}{u_* N})^{u_* N}(u_* + \frac{\Delta k}{N})^{\Delta k}}. \tag{D.13}$$

We assume Δk is small enough such that $\frac{\Delta k}{N}$ approaches zero for large N (e.g. $\Delta k = \sqrt{N}$). Thus

$$\frac{\Omega_2(E_*, N, k_{max} - \Delta k)}{\Omega_2(E_*, N, k_{max})} = \frac{\mathrm{e}^{\Delta k}}{(\Delta k)^{\Delta k}} \times (\frac{1-u_*}{u_*})^{\Delta k} \times \frac{1}{(1+\frac{\Delta k}{u_* N})^{u_* N}}. \tag{D.14}$$

For non-zero u_*, the asymptotic $(1+\frac{\Delta k}{u_* N})^{u_* N} = \mathrm{e}^{\Delta k}$ implies

$$\frac{\Omega_2(E_*, N, k_{max} - \Delta k)}{\Omega_2(E_*, N, k_{max})} = \frac{1}{(\Delta k)^{\Delta k}} \times (\frac{1-u_*}{u_*})^{\Delta k} \tag{D.15}$$

E.g. for $\Delta k = \sqrt{N}$ the ratio is

$$\frac{\Omega(E_*, N, k^* - \Delta k)}{\Omega(E_*, N, k^*)} = \frac{1}{(\sqrt{N})^{\sqrt{N}}} \times (\frac{1-u_*}{u_*})^{\sqrt{N}} \tag{D.16}$$

which except for $u_* = 0$, in the limit of $N \to \infty$, the ratio approaches zero. The final step is to compare the width of the deviation, Δk, to the k^* for $1 \ll N$

$$\lim_{\substack{N \to \infty \\ E \to \infty \\ u_* = const.}} \frac{\Delta k}{k_{max}} = \lim_{\substack{N \to \infty \\ E \to \infty \\ u_* = const.}} \frac{\sqrt{N}}{N(\frac{1-u_*}{2})} = \lim_{\substack{N \to \infty \\ E \to \infty \\ u_* = const.}} \frac{1}{\sqrt{N}(\frac{1-u_*}{2})} = 0. \tag{D.17}$$

So it means almost all of the configurations for $\Omega_2(E, N, k)$ concentrate on k_{max} with deviation equal to $\sqrt{N}$. Therefore, we can safely conclude

$$\Omega_{"}(E, N, k^*) \approx \sum_k \Omega_{"}(E, N, k) = \Omega_2(E, N). \tag{D.18}$$

D.2 The Microcanonical Specific Entropy

Let start with negative energy, or $E_- < 0$, as in equation (5.7). We define

$$E = |E_-| = -E_-, \qquad E > 0, \tag{D.19}$$

The sharp peak around the maximum implies negligible number of elements in head states, or say $n_3 \approx 0$. Putting $n_3 = 0$ and using (5.1) and (5.4),

$$n_1 = -E_- = E, \qquad n_2 = 2p = N - E, \qquad n_3 = 0, \tag{D.20}$$

Consequently (5.2) becomes,

$$\Omega(E, N) = \frac{N!}{2^{\frac{N-E}{2}}(\frac{N-E}{2})!E!}. \tag{D.21}$$

Taking logarithm from both sides and using the Stirling approximation, we derive

$$\frac{\log \Omega}{N} = \frac{1}{2}(1 - \frac{E}{N}) \log N - \frac{E}{N} \log \frac{E}{N} - \frac{1}{2}(1 - \frac{E}{N}) \log(1 - \frac{E}{N}) - \frac{1}{2}(1 - \frac{E}{N}). \tag{D.22}$$

Because E and N are positive numbers and $\frac{E}{N} \leq 1$, there exists u such that

$$\lim_{N \to \infty} \frac{E}{N} = u, \qquad 0 < u \leq 1 \tag{D.23}$$

Then

$$\lim_{N \to \infty} \frac{\log \Omega(E, N)}{N} = \frac{1}{2}(1 - u) \lim_{N \to \infty} (\log N) - u \log u - \frac{1}{2}(1 - u) \log(1 - u)$$
$$- \frac{1}{2}(1 - u) \to \infty. \tag{D.24}$$

The l.h.s is molar entropy (if it is multiplied by Boltzmann constant) and diverging. Therefore *Entropy* is non-extensive.

For positive energies $n_1 = 0$. System's total energy $E_+ > 0$ and like negative case,

$$E = E_+ \qquad E > 0, \tag{D.25}$$

$$n_3 = E_+ = E, \qquad n_2 = 2p = N - E, \qquad n_3 = 0, \tag{D.26}$$

Consequently (5.2) becomes,

$$\Omega(E, N) = \frac{N!}{2^{\frac{N-E}{2}}(\frac{N-E}{2})!E!}. \tag{D.27}$$

which is exactly similar to the negative case. Consequently, it results in the same molar entropy

$$\lim_{N\to\infty} \frac{\log \Omega(E, N)}{N} \to \infty \qquad (0 < u < 1), \tag{D.28}$$

where

$$\lim_{N\to\infty} \frac{E}{N} = u, \qquad 0 < u \leq 1. \tag{D.29}$$

D.3 The Partition Function of 1-D Pairing Ising Model

We defined a partition function of an Ising model without nearest-neighbour interaction and M elements in head or tail states ($\sigma_i \in \{-1, 1\}$) as follow

$$Z_M = \sum_{\{\boldsymbol{\sigma_i}\}} \mathrm{e}^{-\beta B \sum_{i=1}^{M} \sigma_{ij}}. \tag{D.30}$$

Note that for each configuration, $\boldsymbol{\sigma_i}$, we removed elements in the pair state, and therefore, the sum for the Hamiltonian runs over M instead of N.

A transformation matrix between two neighbours can be defined by separating each factor like $\mathrm{e}^{-\beta B(\sigma_i + \sigma_{i+1})/2}$. So,

$$\mathbf{T}_{i,i+1} = \begin{bmatrix} \mathrm{e}^{-\beta B} & 1 \\ 1 & \mathrm{e}^{\beta B} \end{bmatrix}. \tag{D.31}$$

Eigenvalues of the matrix $\mathbf{T}$ are

$$\lambda_1 = 2\cosh(\beta B), \qquad \lambda_2 = 0. \tag{D.32}$$

After diagonalizing and considering periodic boundary condition such as $\sigma_1 = \sigma_{M+1}$, the partition function is equal to,

$$Z_M = Tr(\mathbf{T}^M) = 2^M \cosh^M(\beta B). \tag{D.33}$$

D.4 Evaluating the Partition Function's Summand

Let's define the elements of the sum in equation (5.22) like

$$t_{N,2k} = 2^{N-2k}(2k-1)!! \begin{pmatrix} N \\ 2k \end{pmatrix} \cosh^{N-2k}(\beta B). \tag{D.34}$$

We have,

$$t_{N,2k} = \frac{2^{N-3k}N!}{k!(N-2k)!} \cosh^{N-2k}(\beta B). \tag{D.35}$$

Then by taking the logarithm of the term and using Stirling approximation for $\log N!$ we have,

$$\log t_{N,2k} = (N-3k)\log 2 - k - N\log(1-\frac{2k}{N}) + k\log\frac{N^2}{k} + 2k\log(1-\frac{2k}{N})$$
$$+ (N-2k)\log(\cosh(\beta B)). \tag{D.36}$$

Dividing it by N,

$$\frac{1}{N}\log t_{N,2k} = (1-\frac{3k}{N})\log 2 - \frac{k}{N} - \log(1-\frac{2k}{N}) + \frac{k}{N}\log\frac{N}{\frac{k}{N}} + \frac{2k}{N}\log(1-\frac{2k}{N})$$
$$+ (1-\frac{2k}{N})\log(\cosh(\beta B)). \tag{D.37}$$

Since $0 \leq k \leq N/2$, for $\exists k \in \mathbb{N}$ such that the limit $\lim_{N\to\infty}\frac{k}{N} = \epsilon$ exists for $\epsilon > 0$. Thus

$$\lim_{N\to\infty}\frac{1}{N}\log t_{N,2k} = (1-3\epsilon)\log 2 - \epsilon - \log(1-2\epsilon) + 2\epsilon\log(1-2\epsilon)$$
$$+ (1-2\epsilon)\log(\cosh(\beta B)) + \lim_{N\to\infty}\epsilon\log\frac{N}{\epsilon} \to \infty. \tag{D.38}$$

Appendix

E

Generalised Compounding Mechanism

E.1 Generalised Compounding Mechanism

Case I:

- Mechanism:

$$iA \leftrightarrow A_i. \tag{E.1}$$

- Numbers conservation:

$$n_s + in_i = n. \tag{E.2}$$

- Recursive relation:

$$\Omega_s(n+1) = s\Omega_s(n) + p_i \binom{n}{i} \Omega_s(n-i). \tag{E.3}$$

- Degeneracy:

$$D_n(n_i) = \binom{n}{in_i} (in_i - 1)!^{(i)}, \tag{E.4}$$

where $n!^{(i)}$ is a *multifactorial* defined recursively as

$$n!^{(i)} = \begin{cases} 1 & -i < n \leq 0 \\ n & 0 < n \leq i \\ n \times (n-i)!^{(i)} & n > ik \end{cases} . \tag{E.5}$$

- Probability distribution:

$$P_n(n_i) = \frac{D_n(n_i) r_i^{\lfloor \frac{n-n_i}{i} \rfloor}}{c_n(r_i)} = \frac{\binom{n}{in_i}(in_i - 1)!^{(i)} r_i^{\lfloor \frac{n_s}{i} \rfloor}}{c_n(r_i)}, \tag{E.6}$$

such that r_i is the ratio of abundance of stand-alone elements to the i-tet compound.

- Normalisation constant:

$$c_n(r_i) = \sum_{n_s + in_i = n} D_n(n_i) r_i^{\lfloor n_s/i \rfloor} = \sum_{n_i=0}^{\lfloor n/i \rfloor} \binom{n}{in_i} (in_i - 1)!^{(i)} r_i^{\lfloor \frac{n-n_i}{i} \rfloor}. \tag{E.7}$$

- Large deviation probability:

$$P_{\epsilon_i}(m_i) \propto \mathrm{e}^{n\tilde{H}_{\epsilon_i}(m_i)}, \tag{E.8}$$

for

$$\tilde{H}_{\epsilon_i}(m_i) = -m_i \ln m_i - (1 - m_i) \ln \frac{\mathrm{e}(\frac{1-m_i}{\mathrm{e}})^i}{\epsilon_i}, \tag{E.9}$$

where e is Euler constant, and

$$\lim_{n \to \infty} \frac{r_i}{n} = \epsilon_i. \tag{E.10}$$

Case II:

- Mechanism:

$$(n_{i_1} + n_{i_2} + \cdots + n_{i_k})A \leftrightarrow A_{i_1} + A_{i_2} + \cdots + A_{i_k}, \tag{E.11}$$

- Numbers conservation:

$$n_s + \sum_{i \in \mathcal{A}_k} in_i = n. \tag{E.12}$$

- Recursive relation:

$$\Omega_s(n+1) = s\Omega_s(n) + \sum_{i=1}^{n} p_i \binom{n}{i} \Omega_s(n-i). \tag{E.13}$$

- Degeneracy:

$$D_n(n_{i_1}, \ldots, n_{i_k}) = \binom{n}{i_1 n_{i_1},\ i_2 n_{i_2},\ \ldots,\ i_k n_{i_k},\ n_s} \prod_{i \in \mathcal{A}_k} (i n_i - 1)!^{(i)}, \tag{E.14}$$

where the factor on the right side of the equation is the multinomial coefficient.

- Probability distribution:

$$P_n(n_{i_1}, \ldots, n_{i_k}) = \frac{D_n(n_{i_1}, \ldots, n_{i_k}) \prod_{i \in \mathcal{A}_k} r_i^{\lfloor \frac{n_{(i)} - i n_i}{i} \rfloor}}{c_n(r_{i_1}, \ldots, r_{i_k})}. \tag{E.15}$$

- Normalisation constant:

$$c_n(r_{i_1}, \ldots, r_{i_k}) = \sum_{n_s + \sum_{i \in \mathcal{A}_k} i n_i = n} D_n(n_{i_1}, \ldots, n_{i_k}) \prod_{i \in \mathcal{A}_k} r_i^{\lfloor \frac{n_{(i)} - i n_i}{i} \rfloor}. \tag{E.16}$$

- Large deviation probability:

$$P(m_{i_1}, \ldots, m_{i_k}; \epsilon_{i_1}, \ldots, \epsilon_{i_k}) \propto \mathrm{e}^{n\tilde{H}(m_{i_1}, \ldots, m_{i_k}; \epsilon_{i_1}, \ldots, \epsilon_{i_k})}, \tag{E.17}$$

for

$$\tilde{H}(m_{i_1}, \ldots, m_{i_k}; \epsilon_{i_1}, \ldots, \epsilon_{i_k}) = -m_{i_1} \ln m_{i_1} - \sum_{i=2}^{k} (1 - \sum_{j=1}^{i-1} m_{i_j}) \ln \frac{\mathrm{e}(\frac{1 - \sum_{j=1}^{i-1} m_{i_j}}{\mathrm{e}})^{i_i}}{\epsilon_{i_i}} \tag{E.18}$$

where e is Euler constant, and

$$\lim_{n \to \infty} \frac{r_{i_i}}{n} = \epsilon_{i_i}. \tag{E.19}$$

Appendix

F

Copyright Permissions

F.1 Figure (1.1)

This Agreement between Mr. Roozbeh Pazuki (”You”) and Elsevier (”Elsevier”) consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center. License Number: 5190771097333

License Number: 5190771097333

License date: Nov 16, 2021

Licensed Content Publisher: Elsevier

Licensed Content Publication: Toxicon

F.2 Figure (2.5)

This is a License Agreement between Roozbeh H. Pazuki (”User”) and Copyright Clearance Center, Inc. (”CCC”) on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the CCC Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the CCC Terms and Conditions below.

Order Date: 19-Nov-2021

Order License ID: 1162858-1

ISSN: 1751-8121

Type of Use: Republish in a thesis/dissertation

Publisher: IOP Publishing

Portion: Image/photo/illustration

F.3 Permission from publisher for [33]

This is a License Agreement between Roozbeh H. Pazuki (”User”) and Copyright Clearance Center, Inc. (”CCC”) on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the CCC Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the CCC Terms and Conditions below.

Order Date: 01-Dec-2021

Order License ID: 1165327-1

ISSN: 1751-8121

Type of Use: Republish in a thesis/dissertation

Publisher: IOP Publishing

Portion: Page

F.4 Permission from co-authors for [33]

Dear all,

I hope you are all fine, especially in this hectic Covid time.

I am using some of the results in our joint paper in my dissertation, and I wondered whether I could have your approval for this matter.

Best wishes,

Roozbeh

- **Jensen Henrik Jeldtoft**: Hi Roozbeh,

  Yes – that is certainly fine with me.

  Best wishes,

  Henrik

- **Pruessner, Gunnar**: Of course!

  Cheers,

  G

- **Piergiulio Tempesta**: Fine from my side!

  Best wishes,

  Piergiulio