**Abstract**

Large deviation theory provides a mathematical framework for understanding the probabilities of rare events in stochastic systems. In probability theory, the theory of large deviations concerns the asymptotic behaviour of remote tails of sequences of probability distributions. Roughly speaking, large deviations theory concerns itself with the exponential decline of the probability measures of certain kinds of extreme or tail events. This thesis introduces the basic principles of large deviaition theory with examples about Gaussian Sample Mean and Markov chain and explores the applications of large deviation theory in the context of complex systems and complex networks including Stochastic Climate Model, the Curie-Weiss Model and percolation in complex networks. Also, the thesis introduces the extension of traditional Large Deviation Theory framework which captures the power law decay of rare events through q-exponentials.

**Key Words**: Large deviation theory, complex system, complex network.

**Abstract**

大偏差理论为理解随机系统中罕见事件的概率提供了数学框架。在概率论中，大偏差理论研究的是概率分布序列远尾的渐近行为。通俗而言，它关注某些极端事件或尾事件的概率测度的指数衰减特性。本文介绍了大偏差理论的基本原理，并结合高斯样本均值和马尔可夫链的实例进行说明。随后，本文深入探讨了大偏差理论在复杂系统与复杂网络中的应用，包括随机气候模型、Curie–Weiss 模型以及复杂网络中的渗流过程。此外，论文还介绍了传统大偏差理论框架的扩展，该扩展利用 q 指数刻画罕见事件的幂律衰减行为。

**关键词**: 大偏差理论，复杂系统，复杂网络.

# Contents

# 1 Introduction

Large deviation theory (LDT) is a branch of probability theory that deals with the exponential decay of probabilities associated with rare events. Initially developed in the context of statistical mechanics, it has found applications in diverse areas such as finance, biology, and computer science. Complex systems and networks, characterized by numerous interacting components, often exhibit emergent behaviors that can be analyzed using LDT. This paper focuses on the interplay between LDT and complex networks, highlighting their mutual enrichment in theoretical and practical domains.

The foundation of Large Deviation Theory (LDT) was laid by Cramér in the 1930s, later developed by Donsker and Varadhan, with significant advancements by Freidlin and Wenzell in the 1970s. Remarkably, physicists had been implicitly using LDT concepts a century earlier when calculating entropy and free energy in statistical mechanics. LDT proves indispensable for analyzing both equilibrium and non-equilibrium systems of interacting particles.

LDT is an important tool in many different areas. Large Deviation Theory (LDT) supplies a precise exponential scale for probabilities of atypical outcomes, typically written

$$P(\text{rare event}) \approx e^{-a_n I}.$$

Whereas the Law of Large Numbers or the Central Limit Theorem describe what happens "most of the time," LDT quantifies the likelihood of extreme or non-typical fluctuations and identifies the rate $I$ governing their decay.

In LDT the rate function is often the Legendre–Fenchel transform of a log-moment-generating function. This observation turns a probabilistic tail-estimation problem into a variational or optimisation problem, making both analytical proofs and numerical computation more tractable.

In mathematical finance and statistical inference, LDT provides explicit error exponents and robustness guarantees. Because it yields exact exponential rates, it sharpens traditional confidence intervals, which are often overly conservative when only central-limit heuristics are used.

LDT provides the mathematical foundation for quantifying "improbable probabilities," bridging theoretical probability with practical risk analysis in physics, engineering, and finance. Its power lies in unifying extreme event characterization across scales and disciplines.

There are many Cross-deciplinary applications including Statistical Physics and Thermodynamics, Information Theory and Communications, Machine Learning and Optimization.

Large Deviation Theory offers a rigorous framework for analysing phase transitions and computing free energies. By attaching an explicit rate function to macroscopic observables, it quantifies the exponentially small probability of extreme entropy or energy fluctuations in many-body systems, thereby sharpening classical thermodynamic arguments.

In digital communication channels, LDT shows that the block-error probability decays as

$$P_{\text{error}} \sim e^{-nI}.$$

This exponential law pinpoints how rapidly reliability improves with block length and informs the construction of coding schemes that approach Shannon capacity while remaining robust to noise.

In high-dimensional learning problems, LDT analyses the probability that a trained parameter vector deviates significantly from the true optimum:

$$P\big(\|\theta_{\text{train}} - \theta_{\text{true}}\| > \varepsilon\big) \leq e^{-nI(\varepsilon)}.$$

Such estimates shed light on algorithmic robustness, generalisation error, and the likelihood of overfitting, linking statistical performance to underlying geometric and probabilistic structure.

# 2 Overview of Large Deviation Theory

A basic approximation or scaling law of the form $P_n \approx e^{-nI}$, where $P_n$ is some probability, $n$ a parameter assumed to be large, and $I$ some positive constant, is referred to as a large deviation principle.

## 2.1 Theoretical Significance

The Law of Large Numbers (LLN) tells us "what will happen".

The Central Limit Theorem (CLT) tells us "how things are distributed at the typical fluctuation scale".

Large Deviation Theory (LDT) answers: "If something atypical occurs, exactly how small is its probability, and what is the most likely deviation path?".

These three theories form a pyramid of probability asymptotic theory - from macroscopic averages, to microscopic fluctuations, and finally to extreme tails - progressing sequentially and complementing each other.

| Theory | Focus | Scaling |
|---|---|---|
| Law of Large Numbers | Mean convergence | Almost sure |
| Central Limit Theorem | Small fluctuations | $O(1/\sqrt{n})$ |
| LDT | Extreme deviations | $e^{-nI}$ scaling |

The LLN only concerns "where the average eventually lands", providing an almost sure limit, but says nothing about deviation probabilities. The CLT characterizes the limiting stationary distribution of small deviations at the $\sqrt{n}$ scale, enabling us to construct confidence intervals or perform hypothesis testing. LDT addresses macroscopic-scale ($O(1)$) deviations; it not only shows these events are "rare", but precisely quantifies their exponential rate $I$ - a level of detail that LLN and CLT cannot provide.

## 2.2 Basic Principles

Large deviation theory revolves around the concept of the rate function $I(x)$, which quantifies the likelihood of rare events:

$$P(X_n \approx x) \sim e^{-nI(x)} \quad \text{as } n \to \infty.$$

Here, $X_n$ represents a sequence of random variables, and $I(x)$ is typically a non-negative, lower semi-continuous function.

## 2.3 Example: Gaussian Sample Mean

Consider the sample mean of independent and identically distributed (i.i.d.) Gaussian random variables:

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Assume that $X_i$ follows a normal distribution:

$$p(X_i = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

The expectation and variance of $S_n$ are given by:

$$E[S_n] = \mu, \quad \mathrm{Var}(S_n) = \frac{\sigma^2}{n}.$$

**Claim** A sum of Gaussian random variables is also exactly Gaussian-distributed.

*Proof.* For independent random variables $X$ and $Y$, the distribution $f_Z(z)$ of $Z = X + Y$ equals the convolution of $f_X$ and $f_Y$:

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(z - x) f_X(x)\, dx.$$

Given that $f_X$ and $f_Y$ are normal densities,

$$f_X(x) = \mathcal{N}(x; \mu_X, \sigma_X^2) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}},$$

$$f_Y(y) = \mathcal{N}(y; \mu_Y, \sigma_Y^2) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}.$$

Substituting into the convolution:

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left[-\frac{(z-x-\mu_Y)^2}{2\sigma_Y^2}\right] \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right] dx.$$

Simplifying:

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{2\pi}\sigma_X\sigma_Y} \exp\left[-\frac{\sigma_X^2(z-x-\mu_Y)^2 + \sigma_Y^2(x-\mu_X)^2}{2\sigma_X^2\sigma_Y^2}\right] dx.$$

Expanding and completing the square:

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp\left[-\frac{(z-(\mu_X+\mu_Y))^2}{2\sigma_Z^2}\right] dx,$$

where

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2.$$

Since the expression in the integral is a normal density distribution on $x$, the integral evaluates to 1. The desired result follows:

$$f_Z(z) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp\left[-\frac{(z-(\mu_X+\mu_Y))^2}{2\sigma_Z^2}\right].$$

■

Thus

$$P(S_n = s) = \frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}} e^{-\frac{n(s-\mu)^2}{2\sigma^2}}.$$

Neglecting the term $\sqrt{n}$, which is subdominant with respect to the decaying exponential, thereby obtaining

$$P(S_n = s) \approx e^{-nJ(s)}, J(s) = \frac{(s-\mu)^2}{2\sigma^2}, s \in \mathbb{R}.$$

We can analyse the example with LDT. First we introduce some definitons and theorems [5].

**Theorem 2.1.** *The Gärtner–Ellis Theorem*

*Consider a real random variable $A_n$ parameterized by the positive integer $n$, and define the scaled cumulant generating function of $A_n$ by the limit:*

$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln\langle e^{nkA_n}\rangle,$$

*where $k \in \mathbb{R}$ and*

$$\langle e^{nkA_n}\rangle = \int_{\mathbb{R}} e^{nka}\, \mathbb{P}(A_n \in da).$$

*The Gärtner–Ellis Theorem states that, if $\lambda(k)$ exists and is differentiable for all $k \in \mathbb{R}$, then $A_n$ satisfies a large deviation principle, i.e.,*

$$\mathbb{P}(A_n \in da) \asymp e^{-nI(a)}\, da,$$

*Here we use the sign $\asymp$ instead of $\approx$ when we treat large deviation principles.*

*with a rate function $I(a)$ given by:*

$$I(a) = \sup_{k\in\mathbb{R}} \{ka - \lambda(k)\}.$$

The Gärtner–Ellis theorem is a cornerstone in the theory of large deviations, providing a framework to establish a Large Deviation Principle (LDP) for sequences of random variables that are not necessarily independent or identically distributed (i.i.d.) [1]. It generalizes classical results like Cramér's theorem by leveraging the analytic properties of the scaled cumulant generating function (SCGF)。

**Definition 2.1.** *Let $\xi$ be a random variable on $\mathbb{R}$. The function*

$$\varphi(\lambda) := \ln \mathbb{E} e^{\lambda \xi}, \quad \lambda \in \mathbb{R},$$

*where infinite values are allowed, is called the logarithmic moment generating function or cumulant generating function associated with $\xi$.*

**Definition 2.2.** *The function*

$$f^*(y) := \sup_{x \in \mathbb{R}} \{yx - f(x)\}$$

*is called the Fenchel-Legendre transform of $f$.*

**Theorem 2.2** (Cramér)**.** *Let $\xi_1, \xi_2, \ldots$ be independent identically distributed random variables with mean $\mu \in \mathbb{R}$ and cumulant generating function $\varphi$. Let also $S_n = \xi_1 + \cdots + \xi_n$. Then, for every $x \geq \mu$*

$$\lim_{n \to \infty} \frac{1}{n} \ln P\left(\frac{1}{n} S_n \geq x\right) = -\varphi^*(x),$$

*where $\varphi^*$ is the Fenchel-Legendre transform of $\varphi$.*

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. The moment generating function is given by:

$$\begin{aligned}
M_X(t) &= \mathbb{E}[e^{tX}] \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(tx - \frac{(x-\mu)^2}{2\sigma^2}\right) dx \quad \text{(Definition)} \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left[(x-\mu)^2 - 2\sigma^2 tx\right]\right) dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left[x^2 - 2(\mu+\sigma^2 t)x + \mu^2\right]\right) dx \\
&= \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}\right) dx \\
&= \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \quad \text{(Gaussian integral = 1)}
\end{aligned}$$

Thus, the cumulant generating function is:

$$\psi(t) := \ln \mathbb{E} e^{tX} = \mu t + \frac{1}{2}\sigma^2 t^2$$

Now, consider the Legendre transform:

$$\psi^*(x) = \sup_{t\in\mathbb{R}}\{tx - \psi(t)\}$$

$$= \sup_{t\in\mathbb{R}}\left\{-\frac{1}{2}\sigma^2 t^2 + t(x-\mu)\right\}$$

Solving for the optimal $t^*$:

$$t^* = \frac{x-\mu}{\sigma^2}$$

$$\psi^*(x) = \frac{(x-\mu)^2}{2\sigma^2}$$

From large deviations theory, we obtain:

$$\lim_{n\to\infty}\frac{1}{n}\ln P\left\{\frac{1}{n}S_n \geq x\right\} = -\psi^*(x), \quad \text{for } x \geq \mu.$$

## 2.4 Example: Large Deviation Theory for Markov Chains

When the sample mean $S_n$ is generated by a Markov chain instead of IID random variables $X_1, \ldots, X_n$, the joint probability density function (pdf) takes the form:

$$p(X_1, \ldots, X_n) = p(X_1)\prod_{i=1}^{n-1}\pi(X_{i+1} \mid X_i)$$

where $p(X_1)$ is the initial distribution of $X_1$, and $\pi(X_{i+1} \mid X_i)$ is the transition probability density governing the state transitions $X_1 \to X_2 \to \cdots \to X_n$ (see [7] for details on Markov chains).

Despite the complexity introduced by Markovian dependence, the Gärtner–Ellis (GE) Theorem remains applicable [6]. The scaled cumulant generating function (SCGF)

$\lambda(k)$ is constructed via the tilted matrix formalism. For a homogeneous and ergodic Markov chain, the SCGF of the sample mean $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is given by:

$$\lambda(k) = \ln \zeta(\tilde{\Pi}_k),$$

where $\zeta(\tilde{\Pi}_k)$ is the dominant eigenvalue (largest in magnitude) of the tilted matrix $\tilde{\Pi}_k$, defined element-wise as:

$$\tilde{\pi}_k(x, x') = \pi(x' \mid x)e^{kx'}.$$

( details of the calculation in Sec. 4.3 of [1])

For finite-state Markov chains, $\lambda(k)$ is analytic and differentiable. By the GE Theorem, $S_n$ satisfies a Large Deviation Principle (LDP) with rate function:

$$I(s) = \sup_{k\in\mathbb{R}} \left\{ ks - \ln \zeta(\tilde{\Pi}_k) \right\}.$$

If the Markov chain has an innite number of states, $\lambda(k)$ may fail to be analytic or even exist (see [8] [9] [10] for examples).

Thus the Gärtner–Ellis Theorem extends LDPs to Markov chains by encoding dependencies into the tilted matrix. The dominant eigenvalue $\zeta(\tilde{\Pi}_k)$ bridges stochastic dynamics and convex optimization, enabling analysis of rare events in non-IID systems. This framework underpins applications ranging from metastability studies in physics to risk management in finance.

## 2.5 The definition of LDP and its application

**Definition 2.3.** *Let $I : X \to [0, \infty]$ be a lower semicontinuous function and $\{r_n\}$ a sequence of positive real constants such that $r_n \uparrow \infty$. A sequence of probability measures $\{\mu_n\} \subset \mathcal{M}_1(X)$ is said to satisfy a large deviation principle with rate function $I$ and normalization $r_n$ if the following inequalities hold:*

$$\limsup_{n\to\infty} \frac{1}{r_n} \log \mu_n(F) \le -\inf_{x\in F} I(x) \quad \textit{for all closed } F \subset X,$$

$$\liminf_{n\to\infty} \frac{1}{r_n} \log \mu_n(G) \geq -\inf_{x\in G} I(x) \quad \textit{for all open } G \subset X.$$

*We will abbreviate LDP($\mu_n, r_n, I$) if all of the above holds. When the sets $\{I \leq c\}$ are compact for all $c \in \mathbb{R}$, we say $I$ is a tight rate function [2].*

The large deviation theory itself constitutes a generalized fluctuation theory. It characterizes how the probability of macroscopic quantities deviating from their typical values is exponentially suppressed as the number of random variables (or the system size) becomes extremely large [1]. Specifically, if a random variable $R_N$ satisfies the large deviation principle, then for any measurable set $B$, we have:

$$-\inf_{r\in B^\circ} I(r) \leq \liminf_{N\to\infty} \frac{1}{N} \ln P(R_N \in B) \leq \limsup_{N\to\infty} \frac{1}{N} \ln P(R_N \in B) \leq -\inf_{r\in \overline{B}} I(r).$$

In the simplest case, when considering a single value $R$, this reduces to:

$$\pi(R) = P(R_N = R) \sim e^{-NI(R)}.$$

LDT is not only applicable to simple independent and identically distributed (IID) scenarios but also extends broadly to diverse physical systems. These include the thermodynamic limit of infinitely many particles, percolation in complex networks, Markov processes, noise-driven dynamical systems, and more. The "exponential" form of large deviations holds universally provided the system satisfies the relevant integrability or regularity conditions.

# 3 LDP in Complex Systems and Networks

Complex systems widely exist in nature and human society.Complex systems research is becoming ever more important in both the natural and social sciences [11]. However, there is no concise definition of a complex system, let alone a definition on which all scientists agree. To characterize a complex system, and consider a core set of features that are widely associated with complex systems in the literature and by those in the field, we summarize dome important features: Nonlinearity, FeedbacK, Spontaneous order, Robustness and lack of central control, Emergence, Hierarchical organisation, Numerosity.(details in [12])

Networked modeling of complex systems is a favorable means of understanding complex systems. It not only represents complex interactions but also reflects essential attributes of complex systems. The complex network theory provides a new modeling method for complex systems. The complex net work is a special kind of network structure that abstracts elements in a complex system as nodes and interactions between elements as edges [13] [14].

## 3.1 Complex Systems

A complex system comprises many interacting components that give rise to collective behaviors not easily inferred from individual components. Examples include ecosystems, economic markets, and the human brain [12].

The behavior of a complex system is intrinsically difficult to model due to the dependencies, competitions, relationships, and other types of interactions between their parts or between a given system and its environment. Systems that are "complex" have distinct properties that arise from these relationships, such as nonlinearity, emergence, spontaneous order, adaptation, and feedback loops, among others.

The traditional approach to dealing with complexity is to reduce or constrain it. Typically, this involves compartmentalization: dividing a large system into separate parts. Organizations, for instance, divide their work into departments that each deal with separate issues. Engineering systems are often designed using modular components. However, modular designs become susceptible to failure when issues arise that bridge the divisions.

Now we introduce LDT approach to analyse complex systems.

### 3.1.1 LDP in Stochastic Climate Models

In the example, LDT quantifies the exponentially small probability that noise causes transitions from one attractor to another, and identifies the most probable pathway (the “instanton”) that realizes such a transition.

Let us assume, for simplicity, that the true evolution equation for the climate system can be written as a system of autonomous ordinary differential equations of the form:

$$\frac{dz}{dt} = G(z)$$

where $z \in \mathbb{R}^N$ and $z(t) \in \mathbb{R}^N$ encapsulates all the degrees of freedom (or “state variables”) needed to describe the climate system [15]. For example, in geophysical fluid dynamics, $z$ might include: Fluid velocities (wind, ocean currents), Temperature fields, Moisture or chemical concentrations, etc. $G(z)$ provides the rate of change of each component of $z$.

Then one partitions $z = (x, y)$, where: $x \in \mathbb{R}^n$ is the resolved (or “slow,” or “of interest”) part; $y \in \mathbb{R}^{N-n}$ is the unresolved (or “fast,” or “other”) part. Typically, $n \ll N$.

The system can be written schematically as:

$$\begin{cases} \frac{dx}{dt} = f(x) + \epsilon f_x(x, y), \\ \frac{dy}{dt} = \frac{1}{\epsilon} g(y) + g_y(x, y), \end{cases}$$

where $f$ and $g$ define the autonomous dynamics of the $x$ and $y$ components, respectively, $\epsilon$ is a constant controlling the intensity of the coupling, and $\tau$ defines the time scale separation between the two sets of variables.

The Mori-Zwanzig theory indicates that one can in general write the dynamics of the $x$ variables in an implicit form as follows:

$$\frac{dx}{dt} = f_{\epsilon,\delta}(x) + \dot{\sigma}_{\epsilon,\delta}(x) + \int \mathrm{d}s\, K_{\epsilon,\delta}(x, t-s)x(s),$$

where the three terms of the right hand side correspond to the deterministic drift, to a noise contribution, and to the memory term. In the weak-coupling limit ($\delta \to 0$), it is possible to derive via a perturbative approach an explicit expression for these three terms that is valid up to order $\delta^2$ [16] [17].

When ($\epsilon \to 0$): The second scenario is when the "fast" variables $y$ truly evolve on a much faster time scale than the "slow" variables $x$. In that limit, it is classical from homogenization theory that the $x$-dynamics simplifies further: the memory term drops out, and one is left with a Markovian Stochastic Differential Equation (SDE).

$$dx_t = F(x_t)\, dt + \Sigma(x_t)\, dW_t, \tag{1}$$

is exactly such an SDE, where: $F(x_t)$ is the *renormalized drift*, meaning the original deterministic part gets "corrected" by the influence of the fast variables $y$. $\Sigma(x_t)$ is the *diffusion matrix*, governing the intensity and possibly the state-dependence of the noise. $W_t$ is an $m$-dimensional Brownian motion.

Equation (1) is probably the most convenient starting point for discussing the use of LDT in geophysical flows, LDT can be introduced also in the context of deterministic chaos.

We focus on stochastic climate models subject to weak noise and apply Freidlin–Wentzell theory to analyze large deviations of the resulting paths [19] [20]. Specifically, for the stochastic differential equation

$$dX_t^{\varepsilon} \;=\; b\big(X_t^{\varepsilon}\big)\, dt \;+\; \sqrt{\varepsilon}\, \sigma\big(X_t^{\varepsilon}\big)\, dW_t, \quad X_0^{\varepsilon} = x,$$

where $b : \mathbb{R}^n \to \mathbb{R}^n$ is a deterministic drift, $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ is a diffusion function, and $W_t$ is an $m$-dimensional Brownian motion, one introduces $\varepsilon > 0$ to control the noise intensity.

Under bounded and Lipschitz assumptions on $b$ and $\sigma$, as $\varepsilon \to 0$ the distribution of $X_t^\varepsilon$ converges to the trajectory of the deterministic ODE $\dot{x}_t = b(x_t)$.

Next, to quantify the probability of observing a given path $f(\cdot)$ that deviates from the deterministic solution, the Freidlin–Wentzell large deviation principle provides a *rate function* (or action functional). In particular, for $T > 0$, the path-space probability scales like

$$\mathbb{P}\Big\{\sup_{0\le t\le T}\big|X_t^\varepsilon - f(t)\big| < \delta\Big\} \;\approx\; \exp\Big(-\tfrac{1}{\varepsilon}\, I_T(f)\Big),$$

where

$$I_T(f) \;=\; \tfrac{1}{2}\int_0^T \Big\langle \dot{f}(t) - b\big(f(t)\big),\, a^{-1}\big(f(t)\big)\,\big[\dot{f}(t) - b\big(f(t)\big)\big]\Big\rangle\, dt$$

and $a(x) = \sigma(x)\sigma(x)^T$ is the noise covariance. The trajectory $f^*$ minimizing $I_T(f)$ is called the *minimum-action path* or *instanton*, and it dominates the probability of rare events in the small-noise limit.

Role of Large Deviation Principle in Stochastic Climate Models is describing Rare Event Transitions. To study how climate variables (e.g., global mean temperature, ice volume) undergo transitions from one stable state to another under small random perturbations. This includes estimating the probability and identifying the pathway of rare and extreme climate events. This shows LDP can analyse the dynamics of rare transitions caused by stochasticity, including tipping points and early-warning signals in climate systems.

### 3.1.2 Example: Statistical Mechanics and the Curie-Weiss Model

In the example, LDP help us understand how macroscopic observables (e.g., magnetization) emerge from microscopic random spin configurations in the thermodynamic limit ($n \to \infty$). LDP is used to explain how the Gibbs measure concentrates on certain macrostates [2].

Statistical mechanics explains macroscopic properties of matter from microscopic models. Microscopic reality is taken as random, and so states of the system are probability measures called Gibbs measures. In the limit of an infinite system, laws of large numbers take hold and yield deterministic values for macroscopic observables.

The Curie-Weiss model begins with $n$ atoms that each have a $\pm 1$ valued spin $\omega_i$, $1 \leq i \leq n$. The space of $n$-spin configurations is $\Omega_n = \{-1, 1\}^n$. The energy of the system is given by the Hamiltonian:

$$H_n(\omega) = -\frac{J}{2n} \sum_{1 \leq i,j \leq n} \omega_i \omega_j - h \sum_{j=1}^{n} \omega_j. \tag{2}$$

A ferromagnet has a positive coupling constant $J > 0$ and $h \in \mathbb{R}$ is the external magnetic field. The Gibbs measure for $n$ spins is given by:

$$\gamma_n(\omega) = \frac{1}{Z_n} e^{-\beta H_n(\omega)} P_n(\omega), \quad \omega \in \Omega_n, \tag{3}$$

where $P_n(\omega) = 2^{-n}$ is the a priori measure under which the spins are oriented entirely randomly as i.i.d. fair coin flips. The parameter $\beta > 0$ is the inverse temperature, and the normalization constant $Z_n = \int e^{-\beta H_n} dP_n$ is called the partition function.

The Gibbs measure captures the competition between the ordering tendency of the energy term $H(\omega)$ and the randomness represented by $P_n$. It is instructive to consider the extremes of $\beta$, with $h = 0$. The zero-temperature limit:

$$\lim_{\beta \to \infty} \gamma_n(\omega) = \frac{1}{2}(\delta_{\omega \equiv 1} + \delta_{\omega \equiv -1}) \tag{4}$$

is concentrated on the two ground states, meaning that at low temperature, energy dominates and complete order reigns.

**Proof of (4):**

*Proof.* For $h > 0$, the Hamiltonian is given by:

$$H_n(w) = -\frac{I}{2n} \left( \sum_{j=1}^{n} w_j \right)^2.$$

To minimize $H_n(w)$, we want to maximize $\left| \sum_{j=1}^{n} w_j \right|$.

Ground states are:

$$w^{(+)} = (1, \dots, 1), \quad w^{(-)} = (-1, \dots, -1).$$

If $H_n(w) > H_n(w^*)$ for some ground state $w^*$, then:

$$\exp\left(-\beta H_n(w)\right) = \exp\left(-\beta(H_n(w^*) + \Delta)\right) = e^{-\beta H_n(w^*)} e^{-\beta\Delta},$$

where $\Delta = H_n(w) - H_n(w^*) > 0$.

As $\beta \to \infty$, $e^{-\beta\Delta} \to 0$, thus:

$$\gamma_n(w) = \frac{e^{-\beta H_n(w)} P_n(w)}{\int e^{-\beta H_n} dP_n}$$

$$= \frac{e^{-\beta H_n(w) 2^{-n}}}{2^{-n}\left(e^{-\beta H_n(w^{(+)})} + e^{-\beta H_n(w^{(-)})}\right) + o(e^{-\beta H_n(w)})}.$$

For ground states:

$$\gamma_n(w) = \frac{1}{2}, \quad \text{when } w = w^{(+)} \text{ or } w^{(-)}.$$

Otherwise:

$$\gamma_n(w) = 0.$$

Thus, we obtain:

$$\gamma_n(w) = \frac{1}{2}\left(\delta_{w=(1,\dots,1)} + \delta_{w=(-1,\dots,-1)}\right).$$

■

By contrast, at high temperature ($\beta \to 0$):

$$e^{-\beta H_n(w)} \to 1.$$

Thus:

$$\lim_{\beta\to 0} \gamma_n(\omega) = P_n,$$

so thermal noise dominates and complete disorder reigns.

A key question is the existence of a phase transition: namely, is there a critical temperature $\beta_c^{-1}$ (Curie point) at which the infinite model undergoes a transition that reflects the order/disorder dichotomy of the finite model?

An excess of + or − spins indicates magnetization, so we define magnetization as the expectation:

$$M_n(\beta, h) = \mathbb{E}_{\gamma_n}[S_n],$$

of the total spin $S_n = \sum_{i=1}^n \omega_i$ under the Gibbs measure $\gamma_n$. If $h = 0$, then $\gamma_n(\omega) = \gamma_n(-\omega)$, implying $M_n(\beta, 0) = 0$. Since this is a probability model on a finite space, $M_n(\beta, h)$ is a continuous function of $(\beta, h)$. To explore phase transitions, we take $n \to \infty$ and show that the limit:

$$m(\beta, h) = \lim_{n\to\infty} \frac{1}{n} M_n(\beta, h)$$

exists, and even that the probability distribution of $S_n/n$ converges.

Large deviation analysis shows that the asymptotic magnetization values are determined by the equation:

$$\frac{1}{2}\log\frac{1+z}{1-z} = J\beta z + \beta h, \quad z \in [-1, 1]. \tag{5}$$

A qualitative change happens at $\beta_c = J^{-1}$, which represents the Curie point, marking a phase transition.

Curie-Weiss is called the mean-field approximation of the Ising model because its Hamiltonian can be expressed as

$$H_n(\omega) = -\frac{J}{2}\sum_{i=1}^n \omega_i \left(\frac{1}{n}\sum_{j=1}^n \omega_j\right) - h\sum_{i=1}^n \omega_i$$

where each spin interacts with the "mean field" $S_n/n$ created by all the spins.

Returning to the Curie-Weiss model, the key to large deviations is that the Hamiltonian is a function of $S_n/n$:

$$-H_n(\omega) = n\left(\frac{1}{2}J(S_n/n)^2 + h(S_n/n)\right).$$

Let $\lambda_n$ be the distribution of $S_n/n$ under $P_n$ and let $\mu_n$ be the distribution of $S_n/n$ under $\gamma_n$. By Cramér's theorem LDP($\lambda_n$, $n$, $I_{\text{BER}}$) holds with rate function

$$I_{\text{BER}}(z) = \begin{cases} \frac{1}{2}(1-z)\log(1-z) + \frac{1}{2}(1+z)\log(1+z) & \text{if } -1 \le z \le 1, \\ \infty & \text{otherwise.} \end{cases} \tag{6}$$

Here we use LDT to derive (6).

*Proof.* Since each spin is independently $\pm 1$ with probability $\frac{1}{2}$, the sum $S_n$ follows a binomial distribution.

The normalized sum $Z_n = S_n/n$ behaves like a sum of i.i.d. Bernoulli variables $X_i$ with:

$$P(X_i = +1) = \frac{1}{2}, \quad P(X_i = -1) = \frac{1}{2}.$$

The moment generating function (MGF) of $X_i$ is:

$$\Lambda(k) = \log \mathbb{E}[e^{kX_i}] = \log\left(\frac{e^k + e^{-k}}{2}\right) = \log\cosh(k).$$

Cramér’s theorem tells us that the rate function $I_{\text{BER}}(z)$ is obtained from the Legendre transform:

$$I_{\text{BER}}(z) = \sup_{k\in\mathbb{R}}\{kz - \log\cosh(k)\}.$$

To find $I_{\text{BER}}(z)$, we solve:

$$\frac{d}{dk}\left(kz - \log\cosh(k)\right) = 0.$$

Using:

$$\frac{d}{dk}\log\cosh(k) = \tanh(k),$$

we set:

$$z - \tanh(k) = 0 \quad \Rightarrow \quad k = \tanh^{-1}(z) = \frac{1}{2}\log\left(\frac{1+z}{1-z}\right).$$

Plugging this into the Legendre transform expression:

$$I_{\text{BER}}(z) = kz - \log\cosh(k).$$

Since:

$$\log\cosh(k) = \log\frac{e^k + e^{-k}}{2},$$

we obtain:

$$I_{\mathrm{BER}}(z) = \frac{1}{2}(1+z)\log(1+z) + \frac{1}{2}(1-z)\log(1-z).$$

Thus, for $-1 \le z \le 1$,

$$I_{\mathrm{BER}}(z) = \frac{1}{2}(1+z)\log(1+z) + \frac{1}{2}(1-z)\log(1-z).$$

Outside this range, $I_{\mathrm{BER}}(z) = \infty$ since $P_n(Z_n = z) = 0$ for $|z| > 1$. ■

To take advantage of this, relate $\mu_n$ to $\lambda_n$: for Borel $B \subset [0,1]$,

$$\begin{aligned}
\mu_n(B) = \gamma_n\{S_n/n \in B\} &= \frac{1}{Z_n}\int 1_B(S_n/n)e^{-\beta H_n}\,dP_n \\
&= \frac{1}{Z_n}\int 1_B(S_n/n)e^{n\left(\frac{1}{2}J\beta(S_n/n)^2 + \beta h(S_n/n)\right)}\,dP_n \\
&= \frac{1}{Z_n}\int 1_B(z)e^{n\left(\frac{1}{2}J\beta z^2 + \beta hz\right)}\,\lambda_n(dz)
\end{aligned}$$

LDP($\mu_n$, $n$, $I_{\mathrm{CW}}$) holds with rate function

$$I_{\mathrm{CW}}(z) = I_{\mathrm{BER}}(z) - \frac{1}{2}J\beta z^2 - \beta hz - c, \tag{7}$$

where $c = \inf_z\left\{I_{\mathrm{BER}}(z) - \frac{1}{2}J\beta z^2 - \beta hz\right\}$.

To derive (7), we first need to introduce a theorem.

**Theorem 3.1.** ***Varadhan' s Theorem.*** *Suppose LDP$(\mu_n, r_n, I)$ holds, $f : X \to [-\infty,\infty]$ is a continuous function, and*

$$\lim_{b\to\infty}\lim_{n\to\infty}\frac{1}{r_n}\log\int_{f\ge b} e^{r_n f}\,d\mu_n = -\infty.$$

*Then*

$$\lim_{n\to\infty}\frac{1}{r_n}\log\int e^{r_n f}\,d\mu_n = \sup_{x:f(x)\wedge I(x)<\infty}\{f(x) - I(x)\}.$$

Then we prove a proposition.

**Proposition 1.** *Let $X$ be a topological space, and suppose $\{\mu_n\}$ satisfies an LDP$\big(\mu_n, r_n, I\big)$ on $X$ with rate function $I$. Let $f : X \to \mathbb{R}$ be a bounded continuous function, and define new probability measures $\{\nu_n\}$ by*

$$\nu_n(A) \;=\; \frac{\mathbb{E}_{\mu_n}\big[e^{r_n f}\mathbf{1}_A\big]}{\mathbb{E}_{\mu_n}\big[e^{r_n f}\big]} \;=\; \frac{\int_A e^{r_n f(x)}\,\mu_n(dx)}{\int_X e^{r_n f(x)}\,\mu_n(dx)}.$$

*Then $\{\nu_n\}$ satisfies an LDP$\big(\nu_n, r_n, J\big)$ with rate function*

$$J(x) \;=\; I(x) \;-\; f(x) \;-\; \inf_{y\in X}\big(I(y) - f(y)\big).$$

*Proof.* We want to show that the probability measures $\nu_n$ satisfy a Large Deviation Principle with rate function $J(x)$, meaning that for a measurable set $A$,

$$-\inf_{x\in \mathrm{int}(A)} J(x) \le \liminf_{n\to\infty} \frac{1}{r_n}\log \nu_n(A) \le \limsup_{n\to\infty} \frac{1}{r_n}\log \nu_n(A) \le -\inf_{x\in \overline{A}} J(x). \tag{8}$$

**Upper Bound on a Closed Set**

We wish to prove that for a closed set $F \subset X$,

$$\limsup_{n\to\infty} \frac{1}{r_n}\log \nu_n(F) \le -\inf_{x\in F} J(x),$$

where

$$J(x) = I(x) - f(x) - \inf_{y\in X}[I(y) - f(y)].$$

By definition,

$$\frac{1}{r_n}\log \nu_n(F) = \underbrace{\frac{1}{r_n}\log \int_F e^{r_n f(x)}\,d\mu_n(x)}_{\text{Term (I)}} - \underbrace{\frac{1}{r_n}\log \int_X e^{r_n f(x)}\,d\mu_n(x)}_{\text{Term (II)}}. \tag{9}$$

**Term (I):**

Since $F$ is closed, define the truncated function:

$$\tilde{f}(x) = \begin{cases} f(x), & x \in F, \\ -\infty, & x \notin F. \end{cases}$$

Then

$$\int_F e^{r_n f(x)}\, d\mu_n(x) = \int_X e^{r_n \tilde{f}(x)}\, d\mu_n(x).$$

Applying Varadhan's lemma to $e^{r_n \tilde{f}(x)}$, we obtain

$$\limsup_{n\to\infty} \frac{1}{r_n} \log \int_X e^{r_n \tilde{f}(x)}\, d\mu_n(x) \le \sup_{x\in X}\{\tilde{f}(x) - I(x)\}.$$

Since $\tilde{f}(x) = f(x)$ for $x \in F$ and $\tilde{f}(x) = -\infty$ otherwise,

$$\sup_{x\in X}\{\tilde{f}(x) - I(x)\} = \sup_{x\in F}\{f(x) - I(x)\} = -\inf_{x\in F}\{I(x) - f(x)\}.$$

Thus,

$$\limsup_{n\to\infty} \frac{1}{r_n} \log \int_F e^{r_n f(x)}\, d\mu_n(x) \le -\inf_{x\in F}\{I(x) - f(x)\}.$$

**Term (II):** For the entire space $X$, by Varadhan's lemma we have

$$\lim_{n\to\infty} \frac{1}{r_n} \log \int_X e^{r_n f(x)}\, d\mu_n(x) = \sup_{x\in X}\{f(x) - I(x)\} = -\inf_{x\in X}\{I(x) - f(x)\}.$$

Combine the two estimates:

$$\limsup_{n\to\infty} \frac{1}{r_n} \log \nu_n(F) \le \Big[-\inf_{x\in F}\{I(x) - f(x)\}\Big] - \Big[-\inf_{x\in X}\{I(x) - f(x)\}\Big].$$

we obtain the desired bound:

$$\limsup_{n\to\infty} \frac{1}{r_n} \log \nu_n(F) \le -\inf_{x\in F} J(x).$$

We can prove the lower bound with similar method. ■

By the Theorem above, we get the equality (7). In rate functions $I_{\text{CW}}$ and $I_{\text{BER}}$, CW stands for Curie-Weiss model and BER stands for Bernoulli distribution.

To understand limits of $\frac{S_n}{n}$ under $\gamma_n$, we find the minimizers of $I_{CW}$. Critical points satisfy $I'_{CW}(z) = 0 \iff$ (5) because

$$\frac{d}{dz} I_{\text{CW}}(z) = \frac{1}{2} \log\left(\frac{1+z}{1-z}\right) - \beta J z - \beta h.$$

By direct calculus, when $h \neq 0$, there exists unique solution for (5). Thus $I_{\text{CW}}$ has a unique global minimun and a unique solution $z^*$ of $I'_{CW}(z) = 0$. Then by the construction of $I_{\text{CW}}$, $I_{\text{CW}}(z^*) = 0$. Since $m(\beta, h)$ is a solution of (19), $z^* = m(\beta, h)$ is a unique minimizer of $I_{\text{CW}}$.

By similar argument, If $h = 0$ and $J\beta \leq 1$, then $I_{\text{CW}}$ is uniquely minimized at $z = m(\beta, 0) = 0$.

If $h = 0$ and $J\beta > 1$, then (5) have 3 solutions: 0, $m(\beta, \pm)$. If $h = 0$ and $J\beta > 1$, then $I_{\text{CW}}$ is minimized by $m(\beta, \pm)$. where

$$\lim_{0<\tilde{h}\searrow 0} m(\beta, \tilde{h}) = m(\beta, +) \quad \text{and} \quad \lim_{0>\tilde{h}\nearrow 0} m(\beta, \tilde{h}) = m(\beta, -).$$

**Theorem 3.2.** *The following holds:*

*(a) Suppose that either $h \neq 0$, or $h = 0$ and $\beta \leq 1/J$. Then $\mu_n \to \delta_{m(\beta,h)}$ weakly. For all $\varepsilon > 0$, $\gamma_n\{|S_n/n - m(\beta, h)| \geq \varepsilon\} \to 0$ exponentially fast.*

*(b) If $h = 0$ and $\beta > 1/J$, then $\mu_n \to \frac{1}{2}(\delta_{m(\beta,+)} + \delta_{m(\beta,-)})$ weakly. If $A$ is a closed set such that $m(\beta, \pm) \notin A$ then $\gamma_n\{S_n/n \in A\} \to 0$ exponentially fast.*

*Proof.* Part (a) is immediate because $I_{\text{CW}}$ has a unique minimizer:

$$\limsup_{n\to\infty} \frac{1}{n} \log \gamma_n\{|S_n/n - m(\beta, h)| \geq \varepsilon\} \leq - \inf_{|z-m(\beta,h)|\geq\varepsilon} I_{\text{CW}}(z) < 0.$$

For part (b) the large deviation upper bound gives similarly

$$\lim_{n\to\infty} \gamma_n\{|S_n/n - m(\beta, -)| < \varepsilon \text{ or } |S_n/n - m(\beta, +)| < \varepsilon\} = 1.$$

From $\gamma_n(\omega) = \gamma_n(-\omega)$ it follows that $\mu_n$ is symmetric and so

$$\lim_{n\to\infty} \gamma_n\{|S_n/n - m(\beta,-)| < \varepsilon\} = \lim_{n\to\infty} \gamma_n\{|S_n/n - m(\beta,+)| < \varepsilon\} = \frac{1}{2}.$$

This shows the weak convergence $\mu_n \to \frac{1}{2}(\delta_{m(\beta,-)} + \delta_{m(\beta,+)})$. ■

**Theorem 3.3.** *Let $0 < \beta, J < \infty$ and $h \in \mathbb{R}$.*

1. *For $h \neq 0$, the limit $m(\beta, h)$ is the unique solution of the equation (14) that has the same sign as $h$.*

2. *Let $h = 0$ and $\beta \leq J^{-1}$. Then $z = 0 = m(\beta, 0)$ is the unique solution of (14), and $m(\beta, h) \to 0$ as $h \to 0$.*

3. *Let $h = 0$ and $\beta > J^{-1}$. Then there exist two nonzero solutions $m(\beta, +) > 0$ and $m(\beta, -) = -m(\beta, +)$, indicating spontaneous magnetization.*

The existence of limit $m(\beta, h)$ and claim (1) of Theorem 4 follow from Theorem 3 because $S_n/n$ is bounded, so a weak limit of its distribution implies the limit of its expectation. The rest is just calculus.

In conclusion, under the Gibbs measure $\gamma_n$, the distribution of $S_n/n$ satisfies an LDP with rate function:

$$I_{\mathrm{CW}}(z) = I_{\mathrm{BER}}(z) - \frac{1}{2}J\beta z^2 - \beta h z - c,$$

where $I_{\mathrm{BER}}(z)$ comes from the a priori Bernoulli measure, and $c$ normalizes the minimum to zero.

Minimizers of $I_{\mathrm{CW}}$ describe the most probable macrostates. Their behavior as $\beta$ varies reveals phase transitions and spontaneous magnetization.

In this case, LDP is used to explain static equilibrium behavior and the emergence of order/disorder in the thermodynamic limit.

### 3.1.3 Extension of Traditional LDT Framework

Traditional Large Deviation Theory (LDT), rooted in Boltzmann-Gibbs (BG) statistical mechanics, is applicable to systems with short-range interactions where the Central

Limit Theorem (CLT) holds. In such cases, the probability of rare events decays exponentially:

$$P \sim e^{-Nr}.$$

However, in complex systems with long-range interactions or strong correlations, this framework may fail. Such systems often exhibit $q$-Gaussian distributions rather than classical Gaussians. Thus we introduce a generalized Large Deviation Theory (q-LDT) based on nonextensive statistical mechanics, utilizing the nonadditive entropy $S_q$ to describe the probability decay of rare events [22].

The nonadditive entropy $S_q$ is defined as:

$$S_q = k\frac{1 - \sum_{i=1}^{W} p_i^q}{q-1}, \quad (q \in \mathbb{R}),$$

which recovers the Boltzmann-Gibbs entropy $S_{BG}$ as $q \to 1$ [22]. Maximizing $S_q$ leads to generalized probability distributions of the form:

$$p \propto e_q^{-\beta_q E},$$

where the $q$-exponential is defined as:

$$e_q^x \equiv [1 + (1-q)x]^{1/(1-q)}, \quad (q \neq 1).$$

For $q = 1$, this reduces to the standard exponential function.

In complex systems, cumulative variables may converge to $q$-Gaussian distributions as $N \to \infty$, governed by the q-CLT. The $q$ parameter relates to the shape parameter $Q$ via a function $q = f(Q)$, with $f(1) = 1$.

In complex systems, the probability of large deviations obeys a $q$-exponential decay:

$$P\left(N; \frac{Y_N}{N} > z\right) \sim e_q^{-r_q(z)N},$$

where $r_q(z)$ is the $q$-rate function, satisfying:

$$\lim_{N\to\infty}\frac{1}{N}\ln_q\left(\frac{P(N;Y_N/N>z)}{C_q(z)}\right)=-r_q(z).$$

When $q=1$, this reduces to the classical form of LDT.

The generalized LPT have applications in several cases including Standard map, Coherent Noise Model , Ehrenfest Dog-Flea Model , Random Walk Avalanche Model, Extensivity of Entropy [23] [22].

| **Property** | **Classical LDT** | **q-Generalized LDT** |
|---|---|---|
| Probability decay | Exponential ($e^{-Nr}$) | $q$-exponential ($e_q^{-Nr_q}$) |
| Distribution attractor | Gaussian | $q$-Gaussian |
| Applicable systems | Short-range / i.i.d. | Long-range / strongly correlated |
| Entropy type | Additive ($S_{BG}$) | Nonadditive ($S_q$) |

Table 1: Comparison of classical and generalized large deviation theory

This work generalizes LDT to a $q$-deformed framework suitable for complex systems. The q-LDT framework captures the power-law decay of rare events through $q$-exponentials and is supported by numerical experiments across diverse models. This provides a dynamic foundation for nonextensive statistical mechanics and opens new avenues for understanding extreme events in nature and engineered systems.

## 3.2 Complex Networks

Complex networks represent relationships or interactions among entities, modeled as graphs $G=(V,E)$, where $V$ denotes vertices and $E$ denotes edges. complex networksare graphs that depart substantially from regular or statis-tically regular graphs [21].

Large Deviation Theory (LDT) provides a quantitative framework for analyzing rare events in complex networks, For rare network events (e.g., when the giant component size $R$ significantly deviates from its typical value $\hat{R}$), LDT characterizes the exponential decay probability through the rate function $I(R)$,

$$\pi(R)\sim e^{-NI(R)},$$

where $I(R)$ measures the "cost" of deviation from typical behavior.

Here we introduce the large deviation approach to percolation.

### 3.2.1 Percolation

Large deviation theory provides tools to analyze rare events in network traffic, such as congestion or overload. For example, the probability of extreme traffic conditions can be estimated using the rate function derived from traffic flow models [3] [4].

Here we make a brief introduction to Percolation. Pick each node (site-percolation) or edge (bond-percolation) independently with probability $p$. As $p$ increases, clusters of occupied vertices begin to merge. At a critical threshold $p_c$, a *giant connected component* (GCC) emerges: a cluster whose size scales linearly with the total number of nodes $N$.

In the thermodynamic limit $N \to \infty$, the fraction of nodes in the GCC,

$$P_\infty(p) = \frac{\text{size of GCC}}{N},$$

jumps from 0 to strictly positive ($> 0$) at $p = p_c$, and thus serves as the *order parameter* of the system.

Percolation in Complex Networks refers to the study of how large-scale connectivity or "percolation" emerges or disappears in a network when nodes or edges are randomly occupied or removed.

In node percolation, each node is independently present (occupied) with probability $p$, and absent (removed) with probability $1 - p$. The goal is to understand whether a large connected cluster of occupied nodes emerges.

In edge percolation, nodes remain fixed, but each edge is independently present (occupied) with probability $p$. Again, the key question is if a giant connected component arises as we vary $p$.

Percolation has many applications. Percolation helps model how diseases or information spread across contact networks. A giant connected component can indicate a large-scale outbreak. It can also help to analyse the robustness and resilience of complex systems. By studying how removing nodes or edges affects connectivity, one can assess how fragile or robust various infrastructure networks (power grids, transportation, the Internet) are to random failures or targeted attacks.

Percolation on complex networks is the canonical lens through which we study *network robustness*, *epidemic thresholds*, and *rare-event cascades*. It captures how structural connectivity arises or collapses under randomness, and informs the design and protection of real-world infrastructure and communication systems.

Then we introduce message passing algorithm on single realization of damage. Let us consider a given locally tree-like network of $N$ nodes where each node $i = 1, 2, \ldots, N$ is either damaged ($x_i = 0$) or not ($x_i = 1$). In this case it is well known that the following message passing algorithm is able to determine whether a node belongs ($\rho_i = 1$) or not ($\rho_i = 0$) to the giant component. Specifically, the message passing algorithm consists of a set of recursive equations written for the messages $\sigma_{i\to j}$ that each node $i$ sends to a neighbor node $j$ of the network. (Note that for each interaction between node $i$ and node $j$ there are two distinct messages $\sigma_{i\to j}$ and $\sigma_{j\to i}$.) The message passing equations read:

$$\sigma_{i\to j} = x_i \left[ 1 - \prod_{\ell \in \mathcal{N}(i)\setminus j} (1 - \sigma_{\ell\to i}) \right]$$

where $\mathcal{N}(i)$ indicates the set of neighbors of node $i$.

The messages $\sigma_{i\to j} \in \{0, 1\}$ take the value $\sigma_{i\to j} = 1$ if the node $i$ is not initially damaged ($x_i = 1$) and node $i$ receives at least one positive message from neighbor nodes different from node $j$

The messages $\sigma_{i\to j}$ determine $\rho_i$ which is given by:

$$\rho_i = x_i \left[ 1 - \prod_{j \in \mathcal{N}(i)} (1 - \sigma_{j\to i}) \right].$$

In particular, a node $i$ belongs to the GCC if the node $i$ is not initially damaged ($x_i = 1$) and receives at least one positive message ($\exists j \in \mathcal{N}(i)$ such that $\sigma_{j\to i} = 1$)

Finally, the size of the giant component of the network $R$, resulting after the inflicted initial damage $\{x_i\}_{i=1,2,\ldots,N}$, is given by:

$$\mathcal{R} = \sum_{i=1}^{N} \rho_i.$$

Therefore, different realizations of the initial damage can yield, in general, giant components of different sizes.

Now we consider the random realization of the damage and typical behavior. Let us consider initial damage configurations $\mathbf{x}$ obtained by damaging each node with probability $1-p$, i.e., each configuration $\mathbf{x}$ is drawn from a distribution:

$$\tilde{P}(\mathbf{x}) = \prod_{i=1}^{N} p^{x_i}(1-p)^{1-x_i}.$$

In the mean-field theory of percolation, the expected size $\hat{R}$ of the GCC given by

$$\hat{R} = \sum_{\mathbf{x}} \tilde{P}(\mathbf{x})\mathcal{R},$$

which is obtained by averaging the original message passing algorithm over the distribution $\tilde{P}(\mathbf{x})$.

Given the locally tree-like structure of the network, this procedure generates a novel message passing algorithm determined by the set of messages:

$$\hat{\sigma}_{i\to j} = \sum_{\mathbf{x}} \tilde{P}(\mathbf{x})\sigma_{i\to j},$$

satisfying:

$$\hat{\sigma}_{i\to j} = p\left[1 - \prod_{\ell\in\mathcal{N}(i)\setminus j}(1-\hat{\sigma}_{\ell\to i})\right].$$

These messages determine the probability:

$$\hat{\rho}_i = \sum_{\mathbf{x}} \tilde{P}(\mathbf{x})\rho_i$$

that node $i$ is in the giant component, which is given by:

$$\hat{\rho}_i = p\left[1 - \prod_{\ell\in\mathcal{N}(i)}(1-\hat{\sigma}_{\ell\to i})\right].$$

Finally, the expected size of the giant component $\hat{R}$ is given by:

$$\hat{R} = \sum_{i=1}^{N}\hat{\rho}_i.$$

Here we are interested in going beyond the typical scenario by characterizing the probability $\pi(R)$ that a given configuration of the initial damage yields a giant component of size $R$, i.e.

$$\pi(R) = \sum_{\boldsymbol{x}}\tilde{P}(\boldsymbol{x})\delta(\mathcal{R}, R),$$

where $\delta(m, n)$ is the Kronecker delta. For any given value of $p$, and for large network sizes $N \gg 1$, the probability $\pi(R)$ will follow the large deviation scaling

$$\pi(R) \sim e^{-NI(R)},$$

where $I(R) \geq 0$ is called the *rate function* [1].

In order to find $I(R)$ let us introduce the partition function $Z = Z(\omega)$

$$Z = \sum_{x}\tilde{P}(x)e^{-\omega\mathcal{R}}.$$

We can insert a summation over all possible values of $R$ using the Kronecker delta function $\delta(\mathcal{R}, R)$, which selects the configurations where $\mathcal{R} = R$:

$$Z = \sum_{x}\tilde{P}(x)e^{-\omega\mathcal{R}} = \sum_{x}\tilde{P}(x)e^{-\omega\mathcal{R}}\sum_{R}\delta(\mathcal{R}, R).$$

Reordering the summations gives:

$$Z = \sum_{R}\sum_{x}\tilde{P}(x)\delta(\mathcal{R}, R)e^{-\omega R}.$$

The inner sum is simply the definition of $\pi(R)$, which is the probability that the giant component has size $R$:

$$\sum_x \tilde{P}(x)\delta(\mathcal{R}, R) = \pi(R).$$

Therefore:

$$Z = \sum_R \pi(R)e^{-\omega R}.$$

Let

$$\omega F = \omega N f(\omega) = -\ln Z(\omega),$$

where: $\omega$is typically interpreted as a "bias" or "coupling" parameter; $F$ is free energy; $f(\omega)$ is free energy density, obtained by normalizing the total free energy $F$ per "particle" or "node."

This defines the physical concept of free energy. The term $-\ln Z(\omega)$ reflects the dominant statistical weight of the system under the parameter $\omega$. Assuming $\pi(R) \approx e^{-NI(R)}$, the partition function becomes:

$$Z(\omega) \approx \sum_R \exp\left[-NI(R) - \omega R\right].$$

For large $N$, the saddle-point approximation (or Laplace' s principle) applies. Specifically, the dominant contribution to the logarithmic summation arises from minimizing the exponent $\inf_R\{I(R) + \omega R/N\}$. Thus,

$$\ln Z(\omega) \approx -N \inf_R \left[I(R) + \frac{\omega R}{N}\right].$$

Consequently,

$$-\frac{\ln Z(\omega)}{N} = \frac{1}{N}\ln\frac{1}{Z(\omega)} \approx \inf_R \left[I(R) + \frac{\omega R}{N}\right].$$

Comparing this with $\omega f(\omega) = -\ln Z(\omega)/N$, we immediately derive:

$$\omega f(\omega) = \inf_R \left[I(R) + \frac{\omega R}{N}\right].$$

By the Gärtner–Ellis theorem, as long as $\omega f(\omega)$ is differentiable, the Legendre-Fenchel transform of $\omega f(\omega)$ fully determines the rate function $I(R)$, given by the convex function:

$$I(R) = \sup_{\omega} \left( \omega f(\omega) - \omega \frac{R}{N} \right).$$

Therefore, as long as $\omega f(\omega)$ is differentiable, by studying the free energy $\omega f(\omega)$ of the percolation problem, the large deviation properties of the size $R$ of the giant component can be fully established, and the rate function $I(R)$ is convex.

However, when $I(R)$ is non-convex:

$\omega f(\omega)$ becomes non-differentiable because when $I(R)$ is non-convex, the minimization problem

$$\omega f(\omega) = \inf_{R} \left\{ I(R) + \frac{\omega R}{N} \right\}$$

can exhibit discontinuities as $\omega$ varies, with the solution jumping between different local minima of the function $I(R) + \frac{\omega R}{N}$. and the Legendre-Fenchel transform of $\omega f(\omega)$ only provides the convex envelope of $I(R)$: $(I^*)^*(R)$.

In conclusion, we compute via mean-field methods (message-passing algorithm) to estimate the expected giant component size $\hat{R}$. LDT analyzes extreme damage configurations (e.g., removal of critical hubs) by calculating the rate function $I(R)$.

The connection between thermodynamic quantities and the rate function is established through:

(1) Partition function $Z(\omega)$ and free energy density $f(\omega)$.

(2) Legendre transform relates the rate function to thermodynamic variables:

$$I(R) = \sup_{\omega} \left[ \omega f(\omega) - \omega \frac{R}{N} \right],$$

revealing critical behaviors and non-equilibrium statistical properties of network phase transitions.

# 4 Conclusion and Outlook

Large Deviation Theory (LDT) has become a fundamental theoretical framework for characterizing the probabilities of rare events in stochastic systems. It has found widespread applications across diverse domains such as statistical physics, information theory, financial engineering, biological systems, and machine learning.

LDT is now undergoing a paradigm shift—from the classical paradigm of *exponential decay for independent variables* to a more general framework characterized by *power-law decay and topological dependence* in complex systems. This transition reflects the need to move beyond assumptions of independence, short-range interactions, and Gaussian statistics.

Future research in this field must integrate tools from mathematical physics, data science, and engineering applications to develop multiscale theoretical frameworks that are both analytically powerful and practically applicable. Such a unification will not only deepen our understanding of critical behaviors in complex systems but may also provide a quantitative foundation for addressing global challenges such as climate change, financial instability, and public health crises.

LDT is gradually evolving from a traditional tool of mathematical physics into a core methodology for interpreting, predicting, and controlling rare behaviors in complex systems. By integrating ideas from *nonextensive statistical mechanics*, *dynamical systems theory*, and *modern AI*, the future of LDT promises to enhance our comprehension of extreme-event mechanisms and contribute to the development of intelligent, risk-aware system design and decision-making frameworks.

# References

[1] Touchette H. The large deviation approach to statistical mechanics[J]. Physics Reports, 2009, 478(1-3): 1–69.

[2] Rassoul-Agha F, Seppäläinen T. A course on large deviations with an introduction to Gibbs measures[M]. Salt Lake City, Madison: Self-published by the authors, 2014: 35–48.

[3] Bianconi G. Rare events and discontinuous percolation transitions[J]. Physical Review E, 2018, 97: 022314.

[4] Bianconi G. Large deviation theory of percolation on interdependent multiplex networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2019: 023405.

[5] Konarovskyi V. An introduction to large deviations[R]. Hamburg: Universität Hamburg, 2021.

[6] Touchette H. A basic introduction to large deviations: theory, applications, simulations[C]// Leidl R, Hartmann A K, eds. Modern Computational Science 11: Lecture Notes from the 3rd International Oldenburg Summer School. Oldenburg: BIS-Verlag der Carl von Ossietzky Universität Oldenburg, 2011.

[7] Grimmett G, Stirzaker D. Probability and random processes[M]. Oxford: Oxford University Press, 2001:213-296.

[8] Harris R J, Rakos A, Schutz G M. Breakdown of Gallavotti-Cohen symmetry for stochastic dynamics[J]. Europhys. Lett., 2006, 75: 227–233.

[9] Harris R J, Schutz G M. Fluctuation theorems for stochastic dynamics[J]. Journal of Statistical Mechanics: Theory and Experiment, 2007(07): P07020.

[10] Rakos A, Harris R J. On the range of validity of the fluctuation theorem for stochastic Markovian dynamics[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008(05): P05005.

[11] Yang K, Li J, Liu M, Lei T, Xu X, Wu H, Cao J, Qi G. Complex systems and network science: a survey[J]. Journal of Systems Engineering and Electronics, 2023, 34(3): 543–573.

[12] Ladyman J, Lambert J, Wiesner K. What is a complex system?[J]. European Journal for Philosophy of Science, 2013, 3: 33–67.

[13] Barabási A L. Network science[M]. Cambridge: Cambridge University Press, 2013.

[14] Wang X F, Li X, Chen G R. Theory and application of complex networks[M]. Beijing: Tsinghua University Press, 2006. (in Chinese)

[15] Galfi V M, Lucarini V, Ragone F, Wouters J. Applications of large deviation theory in geophysical fluid dynamics and climate science[EB/OL]. 2021, arXiv:2106.13546.

[16] Wouters J, Lucarini V. Disentangling multi-level systems: averaging, correlations and memory[J]. Journal of Statistical Mechanics: Theory and Experiment, 2012(03): P03003.

[17] Wouters J, Lucarini V. Multi-level dynamical systems: connecting the Ruelle response theory and the Mori-Zwanzig approach[J]. Journal of Statistical Physics, 2013, 151(5): 1-11.

[18] Pavliotis G A, Stuart A M. Multiscale methods:Averaging and Homogenization[M]. New York: Springer, 2008:1-10.

[19] Varadhan S R S. Large deviations and applications[M]. Philadelphia: Society for Industrial and Applied Mathematics, 1984:1-60.

[20] Wentzell A D, Freidlin M I. Random perturbations of dynamical systems[M]. 2nd ed. New York: Springer, 1998:192-244.

[21] Costa L F. What is a complex network?[EB/OL]. 2006, arXiv:cs/0606060.

[22] Tirnakli U, Tsallis C, Ay N. Approaching a large deviation theory for complex systems[J]. Entropy, 2021, 23(10): 1306.

[23] Chirikov B V. A universal instability of many-dimensional oscillator systems[J]. Physics Reports, 1979, 52: 263-379.

# 5 Acknowledgements

As I approach the end of my undergraduate journey, I would like to express my heartfelt gratitude to all those who have supported me along the way.

First and foremost, I extend my sincere thanks to Professor Cunquan Qu . Professor Qu not only helped me choose the topic of my thesis but also offered patient and insightful guidance throughout the writing process. His academic rigor, clear explanations, and constant encouragement have greatly inspired me and reinforced my determination to pursue further study.

I am also deeply grateful to Fanpeng Song . His careful analysis and constructive suggestions during the refinement of my thesis helped me overcome many obstacles. His patience and willingness to assist were truly invaluable during moments of uncertainty.

Above all, I wish to thank my parents. Your unwavering support and understanding have been my greatest source of strength. In times of doubt, your encouragement lifted me up; in moments of progress, your quiet pride gave me comfort and motivation. Your love and sacrifices are the foundation on which all of this stands.

To everyone who has offered guidance, kindness, or encouragement during my undergraduate years—thank you. This thesis is dedicated to you, with my deepest appreciation.