Statistical Physics: From Zero to Research Frontiers

A 70-Chapter Technical Roadmap

Shen Han

December 20, 2025

Contents

Part I

Core Language and Minimal Axioms

Chapter 1

Probability Spaces and Minimal Measure Theory

1.1 Motivation: Why Measure-Theoretic Probability?

Core idea. Probability theory is a framework for assigning "sizes" to events and defining averages of observables in a way that is stable under *limits*, *countable operations*, and *continuous state spaces*.

Event & observable.

- **Event**: a yes/no statement, modeled by a set $A \subseteq \Omega$.
- **Observable**: a numerical quantity, modeled by a function $X : \Omega \to R$.

Resolution viewpoint (intuition). A σ-algebra can be viewed as a "resolution" or "information filter": it specifies which events are distinguishable/allowed to be assigned probabilities.

1.2 σ-algebras: the language of measurable events

Let Ω be a set (sample space). A collection $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is a **σ-algebra** if:

1. $\Omega \in \mathcal{F}$;
2. if $A \in \mathcal{F}$ then $A^c = \Omega \setminus A \in \mathcal{F}$;
3. if $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Consequences.

- $\in \mathcal{F}$ since $= \Omega^c$.
- Closed under countable intersections: $\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c \in \mathcal{F}$.
- Finite unions and intersections are included as special cases.

1.3 Measures and probability measures

A function $\mu : \mathcal{F} \to [0, \infty]$ is a **measure** if:

1. $\mu() = 0$;
2. (**countable additivity**) if $A_1, A_2, \ldots$ are pairwise disjoint, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

A **probability measure** P is a measure with $P(\Omega) = 1$.

Monotonicity. If $A \subseteq B$ then $\mu(A) \leq \mu(B)$. *Proof sketch:* write $B = A \cup (B \setminus A)$ disjointly, so $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$.

Bounds for probabilities. For any $A \in \mathcal{F}$,

$$0 \leq P(A) \leq P(\Omega) = 1.$$

1.4 Generating σ-algebras

Given a family of sets $\mathcal{A} \subseteq \mathcal{P}(\Omega)$, define

$$\sigma(\mathcal{A}) = \bigcap \{\mathcal{G} : \mathcal{G} is a$$

σ-algebra on $\Omega,\ \mathcal{A} \subseteq \mathcal{G}\}. This is the$**smallest**$\sigma$-algebra containing $\mathcal{A}$.

Monotonicity of generation. If $\mathcal{A} \subseteq \mathcal{B}$ then $\sigma(\mathcal{A}) \subseteq \sigma(\mathcal{B})$.

1.5 Probability spaces

A **probability space** is a triple $(\Omega, \mathcal{F}, P)$ where:

- Ω is the sample space,
- $\mathcal{F}$ is a σ-algebra of events,
- P is a probability measure on $\mathcal{F}$.

Example (coin toss). $\Omega = \{0, 1\}$, $\mathcal{F} = \mathcal{P}(\Omega) = \{, \{0\}, \{1\}, \{0, 1\}\}$, $P(\{1\}) = p$, $P(\{0\}) = 1 - p$.

1.6 Random variables as measurable functions

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(R, \mathcal{B}(R))$ the real line with its Borel σ-algebra. A function $X : \Omega \to R$ is a **random variable** if it is **measurable**:

$$\forall B \in \mathcal{B}(R), \quad X^{-1}(B) \in \mathcal{F}.$$

Equivalently, it suffices to check $\{X \leq a\} \in \mathcal{F}$ for all $a \in R$.

Inverse image. For $B \subseteq R$,

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}.$$

Measurability guarantees that events like "$X \in B$" have well-defined probabilities.

1.7 Distribution as pushforward measure

The **distribution** (law) of X is the probability measure μ_X on $(R, \mathcal{B}(R))$ defined by

$$\mu_X(B) = P(X \in B) = P(X^{-1}(B)), \qquad B \in \mathcal{B}(R).$$

This is the **pushforward** of P by X, sometimes written $\mu_X = P \circ X^{-1}$.

Coin example. If $X(1) = 1$ and $X(0) = 0$ with $P(1) = p$, then $\mu_X(\{1\}) = p$ and $\mu_X(\{0\}) = 1 - p$.

1.8 (continued) Expectation as an integral

For a random variable X, the expectation is the integral

$$E[X] = \int_\Omega X(\omega)\, dP(\omega),$$

whenever the integral is well-defined.

Discrete case. If $\Omega = \{\omega_i\}$ is countable and $P(\{\omega_i\}) = p_i$, then

$$E[X] = \sum_i X(\omega_i)\, p_i.$$

For the coin example with $X \in \{0, 1\}$, $E[X] = p$.

1.9 Independence

Events. Events $A, B \in \mathcal{F}$ are **independent** if

$$P(A \cap B) = P(A)P(B).$$

If $P(B) > 0$, this is equivalent to $P(A \mid B) = P(A)$.

Random variables. Random variables X, Y are independent if for all Borel sets $B, C \in \mathcal{B}(R)$,

$$P(X \in B,\ Y \in C) = P(X \in B)\, P(Y \in C).$$

Equivalently, the σ-algebras $\sigma(X)$ and $\sigma(Y)$ are independent.

Checking independence on generators (idea). Since $\{(-\infty, a]\}_{a\in R}$ generates $\mathcal{B}(R)$, it is often enough to check independence for events of the form $\{X \le a\}$ and $\{Y \le b\}$, then extend to all Borel sets via a closure theorem (e.g. π-λ/monotone class).

1.10 Modes of convergence

Let X_n, X be random variables.

Almost sure (a.s.) convergence. $X_n \to X$ almost surely if

$$P\Big(\Big\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big\}\Big) = 1.$$

Intuition: convergence may fail on a *probability-zero* set, which is "invisible" to P.

Convergence in probability. $X_n \to X$ in probability if for all $\varepsilon > 0$,

$$P(|X_n - X| > \varepsilon) \to 0.$$

L^p convergence. For $p > 0$, $X_n \to X$ in L^p if

$$E\left[|X_n - X|^p\right] \to 0.$$

Special cases: $p = 1$ (mean absolute error), $p = 2$ (mean squared error / MSE).

Strength relations (useful facts).

- $X_n \to X$ a.s. $\Rightarrow X_n \to X$ in probability.
- $X_n \to X$ in $L^p \Rightarrow X_n \to X$ in probability (via Markov's inequality on $|X_n - X|^p$).
- In general, a countable set does *not* automatically have probability 0; this depends on P (discrete vs continuous).

1.11 Chapter 1 Exercises Summary: σ-Algebras, Measurability, Independence, Convergence, and Core Inequalities

1.11.1 σ-Algebras on Finite Ω: "Distinguishability"

Definition. A collection $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is a σ-algebra if:

1. $\Omega \in \mathcal{F}$;
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$;
3. $A_n \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Finite check trick. When Ω is finite and $\mathcal{F}$ is small, "countable unions" reduce to unions among finitely many members. For example, $\mathcal{F} = \{, \Omega, \{1,2\}, \{3,4\}\}$ is a σ-algebra: complements remain inside, and unions produce only $, \{1,2\}, \{3,4\}, \Omega$.

Partition/atoms viewpoint. On a finite Ω, every σ-algebra corresponds to a partition of Ω into *atoms*

$$\Omega = A_1 \sqcup \cdots \sqcup A_m, \quad \mathcal{F} = \Big\{ \bigcup_{k \in I} A_k : \ I \subseteq \{1, \dots, m\} \Big\}.$$

Interpretation: $\mathcal{F}$ encodes what events are distinguishable; atoms are the finest distinguishable units under $\mathcal{F}$.

Generated σ-algebra. Given $\mathcal{A} \subseteq \mathcal{P}(\Omega)$, $\sigma(\mathcal{A})$ is the smallest σ-algebra containing $\mathcal{A}$. On finite Ω, one can find the atom partition by intersecting members of $\mathcal{A}$ and their complements. For two sets A, B:

$$A \cap B, \quad A \cap B^c, \quad A^c \cap B, \quad A^c \cap B^c.$$

If the atoms are singletons, then $\sigma(\mathcal{A}) = \mathcal{P}(\Omega)$.

1.11.2 Measurability: Random Variables as Coarse-Graining

Definition. A map $X : (\Omega, \mathcal{F}) \to (R, \mathcal{B}(R))$ is measurable iff

$$\forall a \in R, \quad \{X \le a\} \in \mathcal{F}.$$

(Equivalently: $X^{-1}(B) \in \mathcal{F}$ for all Borel sets B.)

Exercise pattern. If $\mathcal{F}$ only distinguishes atoms A_k, then X is $\mathcal{F}$-measurable iff X is constant on each atom A_k. Otherwise there exists a threshold a such that $\{X \le a\}$ "cuts an atom" and thus is not in $\mathcal{F}$.

Simple-function decomposition on finite $\mathcal{F}$. If $\{A_k\}$ are atoms of $\mathcal{F}$, then any $\mathcal{F}$-measurable X can be written uniquely as

$$X(\omega) = \sum_{k=1}^{m} x_k \, \mathbf{1}_{A_k}(\omega), \quad x_k \in R.$$

Moreover, for any function $f : R \to R$,

$$f(X) = \sum_{k=1}^{m} f(x_k) \, \mathbf{1}_{A_k}.$$

Interpretation: the σ-algebra provides a "basis of distinguishability"; measurable functions are exactly those that respect it.

My viewpoint ("universal"/factorization intuition). A given σ-algebra may admit many different generating families, but there is a canonical "finest" partition into atoms (in the finite case). Any other description can be refined down to this atomic partition, and measurable objects factor through it: first reduce to atomic information, then assemble. This feels like a universal/factorization principle: the atomic partition plays a canonical role among all generating presentations.

1.11.3 Distribution (Pushforward) and Expectation

Pushforward measure (law). Given a probability P on $(\Omega, \mathcal{F})$ and a measurable X, define the distribution (law)

$$\mu_X(B) = P(X \in B) = P(X^{-1}(B)), \quad B \in \mathcal{B}(R),$$

i.e. $\mu_X = P \circ X^{-1}$.

Indicator function bridge. For any event $A \in \mathcal{F}$,

$$\mathbf{1}_A(\omega) = \{ 1 \ , \omega \in A, 0, \omega \notin A, \qquad E[\mathbf{1}_A] = P(A).$$

This is the key bridge turning expectation statements into probability statements.

Tail-sum formula (discrete nonnegative). If Y takes values in $\{0, 1, 2, \dots\}$, then

$$Y = \sum_{k=1}^{\infty} \mathbf{1}_{\{Y \geq k\}}, \qquad E[Y] = \sum_{k=1}^{\infty} P(Y \geq k).$$

1.11.4 Independence: Events, σ-Algebras, Random Variables

Events. A, B are independent iff $P(A \cap B) = P(A)P(B)$.

σ-algebras. $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$ are independent iff

$$\forall A \in \mathcal{A}, \ \forall B \in \mathcal{B}, \quad P(A \cap B) = P(A)P(B).$$

Random variables. X, Y are independent iff $\sigma(X)$ and $\sigma(Y)$ are independent.

A quick non-independence lemma. If $C \subset A$ with $0 < P(C) < 1$ and $P(A) < 1$, then A and $C cannot be independent, since P(C) = P(A \cap C) = P(A)P(C) \Rightarrow P(A) = 1 (contradiction)$.

Independence $\Rightarrow$ factorization of expectation (core idea). If $A \in \sigma(X)$ and $B \in \sigma(Y)$ and $X \perp Y$, then

$$E[\mathbf{1}_A \mathbf{1}_B] = P(A \cap B) = P(A)P(B) = E[\mathbf{1}_A]E[\mathbf{1}_B].$$

By linearity of expectation, this extends to simple functions

$$X = \sum_i a_i \mathbf{1}_{A_i}, \quad Y = \sum_j b_j \mathbf{1}_{B_j} \ \Rightarrow \ E[XY] = E[X]E[Y].$$

More generally, for integrable measurable f, g,

$$X \perp Y \ \Rightarrow \ E[f(X)g(Y)] = E[f(X)] \, E[g(Y)].$$

1.11.5 Modes of Convergence

Almost sure convergence. $X_n \to X$ almost surely iff

$$P\big(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\big) = 1.$$

Convergence in probability. $X_n \to X$ in probability iff for all $\varepsilon > 0$,

$$P(|X_n - X| > \varepsilon) \to 0.$$

$\limsup$ **and** $\liminf$ **of events.** For events A_n:

$$\limsup_{n\to\infty} A_n = \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} A_n \quad (``infinitely often''),$$

$$\liminf_{n\to\infty} A_n = \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} A_n \quad (``eventually always'').$$

Sketch: a.s. $\Rightarrow$ in probability. Fix $\varepsilon > 0$, let $A_n = \{|X_n - X| > \varepsilon\}$. If $X_n \to X$ a.s., then $P(\limsup A_n) = 0$. By continuity of probability for the decreasing sets $B_N := \bigcup_{n \geq N} A_n$,

$$P(\limsup A_n) = P\Big(\bigcap_N B_N\Big) = \lim_{N\to\infty} P(B_N) = 0.$$

Since $\sup_{n\geq N} P(A_n) \leq P(B_N)$, it follows that $P(A_n) \to 0$, i.e. $X_n \to X$ in probability.

1.11.6 Three Core Inequalities

Markov's inequality. If $Y \geq 0$ and $a > 0$, then

$$P(Y \geq a) \leq \frac{E[Y]}{a}.$$

Proof idea: $\mathbf{1}_{\{Y\geq a\}} \leq Y/a$, then take expectations.

Chebyshev's inequality. If $E[X^2] < \infty$, then for any $\varepsilon > 0$,

$$P\big(|X - E[X]| \geq \varepsilon\big) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}, \quad \mathrm{Var}(X) = E\big[(X - E[X])^2\big].$$

Proof idea: apply Markov to $Y = (X - E[X])^2$ and $a = \varepsilon^2$.

Jensen's inequality. If φ is convex and $E|X| < \infty$, then

$$\varphi(E[X]) \leq E[\varphi(X)].$$

Special case $\varphi(x) = x^2$: $(E[X])^2 \leq E[X^2]$.

Consequence: $L^p \Rightarrow$ convergence in probability. If $p \geq 1$ and $E|X_n - X|^p \to 0$, then for any $\varepsilon > 0$,

$$P(|X_n - X| \geq \varepsilon) = P(|X_n - X|^p \geq \varepsilon^p) \leq \frac{E|X_n - X|^p}{\varepsilon^p} \to 0,$$

by Markov.

Final synthesis viewpoint.

- A σ-algebra specifies the resolution at which we can distinguish events (atoms are the finest resolution in the finite case).
- Measurable random variables/functions are precisely those that respect this resolution (constant on atoms).
- Independence expresses informational decoupling, leading to factorization of probabilities and expectations.
- Convergence modes and inequalities provide the bridge between pointwise/almost sure statements and probabilistic/mean-type control.

Chapter 2

Generating-Function Techniques for Random Variables

Chapter 3

Conditional Probability and Information Structure

Chapter 4

Usable Versions of LLN and CLT

Chapter 5

An Introduction to Large Deviations

Chapter 6

Convex Analysis and Thermodynamic Duality

Chapter 7

Variational Calculus and Lagrange Multipliers

Chapter 8

Complex Analysis and the Saddle-Point Method (Practical)

Chapter 9

Linear Response and Correlation Functions

Chapter 10

The Objects and States of Statistical Physics

Part II

Thermodynamics and Ensembles: From Definitions to Computation

Chapter 11

Axiomatizing Thermodynamics and Legendre Structure

Chapter 12

The Microcanonical Ensemble and Boltzmann Entropy

Chapter 13

The Canonical Ensemble and the Partition Function Z

Chapter 14

The Grand Canonical Ensemble and Number Fluctuations

Chapter 15

Ensemble Equivalence and Nonequivalence (Prelude to Phase Transitions)

Chapter 16

The Ideal Gas: From Phase Space to Equations of State

Chapter 17

Quantum Statistics: Bose/Fermi Ensemble Techniques

Chapter 18

Thermodynamic Criteria for Phase Transitions

Chapter 19

Stability and Phase Diagram Construction

Chapter 20

Partition-Function Zeros and Phase Transitions (Lee–Yang/Fisher)

Part III

Canonical Models and Approximations: Learning to Compute

Chapter 21

The Ising Model and Basic Symmetries

Chapter 22

The 1D Ising Model via the Transfer Matrix (Exact)

Chapter 23

The 2D Ising Model: Duality and the Self-Dual Point

Chapter 24

The Potts Model and Random Clusters (Fortuin–Kasteleyn)

Chapter 25

Continuous Symmetry Models: XY/Heisenberg and Goldstone Modes

Chapter 26

Mean-Field Theory and Landau Free Energy

Chapter 27

The Bethe Approximation and Exactness on Trees

Chapter 28

High-/Low-Temperature Expansions and Graph Counting

Chapter 29

The Variational Principle and the Bogoliubov Inequality

Chapter 30

Computational Statistical Physics I: MCMC Fundamentals

Part IV

Critical Phenomena and Renormalization: The Modern Core Toolkit

Chapter 31

Scaling Hypotheses and Universality

Chapter 32

Correlation Functions and Structure Factors

Chapter 33

Wilson RG: Coarse-Graining to Flow Equations

Chapter 34

ε-Expansion and the ϕ^4 Field-Theory Correspondence

Chapter 35

Conformal Invariance (with a 2D Emphasis)

Chapter 36

The Kosterlitz–Thouless Transition

Chapter 37

Finite-Size Scaling (FSS)

Chapter 38

Computational Statistical Physics II: Cluster Algorithms

Chapter 39

Computational Statistical Physics III: Multihistogram Reweighting

Chapter 40

Transfer Matrices and Tensor-Network Beginnings

Part V

Stochastic Processes and Equilibrium Dynamics: From Noise to Path Integrals

Chapter 41

Markov Chains and Master Equations

Chapter 42

Langevin Equations and Itô/Stratonovich Calculus

Chapter 43

Fokker–Planck Equations and Stationary Distributions

Chapter 44

Dynamic Critical Phenomena (Hohenberg–Halperin Classification)

Chapter 45

MSR/JD Formalism: Stochastic Dynamics to Path Integrals

Chapter 46

The Fluctuation–Dissipation Theorem (FDT)

Chapter 47

Green–Kubo Relations and Transport Coefficients

Chapter 48

Dynamical Renormalization Group (DRG)

Chapter 49

Reaction–Diffusion Systems and the DP Universality Class

Chapter 50

Low-Dimensional Chaos and Statistical Descriptions

Part VI

Nonequilibrium Statistical Physics: The Modern Frontier Arena

Chapter 51

Nonequilibrium Steady States and Probability Currents (NESS)

Chapter 52

Entropy Production and Fluctuation Theorems (FT)

Chapter 53

The Jarzynski Equality and the Crooks Relation

Chapter 54

Large Deviations as the Core Nonequilibrium Language

Chapter 55

Macroscopic Fluctuation Theory (MFT)

Chapter 56

The KPZ Universality Class and Growth Processes

Chapter 57

Active Matter and Nonequilibrium Phase Transitions

Chapter 58

Glassy Dynamics and Slow Relaxation

Chapter 59

Spin Glasses and Energy Landscapes: Replica/Cavity Methods

Chapter 60

Non-Hermitian and Open-System Effective Descriptions

Part VII

Statistical Physics $\times$ Information/Learning/Neural Networks: Cross-Disciplinary Frontiers

Chapter 61

Maximum-Entropy Models and Exponential Families: From Ising to Neural Populations

Chapter 62

Inference Phase Transitions: Detectability and Information-Theoretic Limits

Chapter 63

Message Passing (BP/AMP) and Free Energies

Chapter 64

Random Matrices and High-Dimensional Spectral Statistics

Chapter 65

Neural-Network Mean-Field Theory and the NTK Limit

Chapter 66

Field-Theoretic / Path-Integral Views of Learning Dynamics

Chapter 67

Loss Landscapes Through Glass/Spin-Glass Lenses

Chapter 68

Generative Models and Statistical Physics: EBMs/Diffusion/Score

Chapter 69

Nonequilibrium Information Thermodynamics and Learning

Chapter 70

A Research-Grade Toolbox and Problem Templates