Statistical Field Theory for Neural Networks

Mathematical Statistics and Field-Theoretic Frameworks (Ch. 1-5)

Ziliang Shen

2026 年 5 月 25 日

School of Statistics and Data Science
Shanghai University of Finance and Economics

Outline

Introduction & Methodological Bridge

Background: Why Statistical Field Theory (SFT)?

The Challenge

- **Complexity:** Massive non-linearity & quenched disorder.
- **Bottleneck:** Standard integration methods fail in infinite dimensions.

The SFT Shift

- **Old way:** Direct integration.
- **New way:** Algebraic manipulation of Generating Functionals & Topology.

Roadmap: The 0D Bridge

$0 \Rightarrow \infty$ **Complexity** is daunting.

We begin with 0D Fields (equivalent to standard random variables $x \in \mathbb{R}^p$) to build the formal bridge.

Probability Density vs. The Action

Core Concept: We define the relationship between the Probability Density Function (PDF) $p(x)$ and the Field-Theoretic Action $S(x)$ as:

$$p(x) = \frac{1}{Z_0} \exp(S(x)) \quad \Longleftrightarrow \quad S(x) = \ln p(x) + \text{const} \tag{1}$$

The "Likelihood" Connection

- In **Statistics**, $\ln p(x)$ is the Log-Likelihood Function. Maximizing it (MLE) finds the most probable state of the system.
- In **Physics**, $S(x)$ is the Action. Minimizing the negative Action finds the ground state (Principle of Least Action).

The statistical problem of defining distributions is mathematically isomorphic to the physics problem of defining energy potentials.

Probabilities, Moments, and Cumulants

Glossary of Symbols: The Bracket Notation Shift

In statistical field theory, we translate standard statistical operators into Dirac brackets to easily handle multi-dimensional graph derivations:

Single Bracket $\langle \dots \rangle$: Expectation & Raw Moments

The standard ensemble average over the probability density $p(x)$:

- **1-Point:** $\langle x_i \rangle = \mathbb{E}[x_i]$ (Expectation / **期望**)
- **k-Point:** $\langle x_{i_1} \dots x_{i_k} \rangle = \mathbb{E}[x_{i_1} \dots x_{i_k}]$ (Raw Moment / **原点矩**)

Double Bracket $\langle\langle \dots \rangle\rangle$: Covariance & Cumulants

The connected correlation function that filters out independent background noise:

- **2-Point:** $\langle\langle x_i x_j \rangle\rangle = \mathrm{Cov}(x_i, x_j)$ (Covariance / **协方差**)
- **k-Point:** $\langle\langle x_{i_1} \dots x_{i_k} \rangle\rangle = \mathrm{Cumulant}(x_{i_1}, \dots, x_{i_k})$ (Cumulant / **累积量**)

Generators: Laplace Transform and Extractions

By introducing an external source (current) $j \in \mathbb{R}^p$, we can systematically generate these single and double brackets.

The Laplace Transform Definition

We define $Z(j)$ as the Partition Function(**配分函数**), which is mathematically the Laplace Transform of $p(x)$ evaluated at $-j$:

$$Z(j) = \langle \exp(j^\top x) \rangle = \int_\Omega \exp(j^\top x) p(x) dx \tag{2}$$

Extracting Raw Moments and Cumulants via Taylor Expansion

The Taylor expansion coefficients of $Z(j)$ yield the **Raw Moments**:

$$Z(j) = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{i_1 \dots i_k} j_{i_1} \dots j_{i_k} \langle x_{i_1} \dots x_{i_k} \rangle \tag{3}$$

Its logarithm isolates genuine dependencies, defining the Cumulants $\langle\langle \dots \rangle\rangle$:

$$W(j) = \ln Z(j) = \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{i_1 \dots i_k} j_{i_1} \dots j_{i_k} \langle\langle x_{i_1} \dots x_{i_k} \rangle\rangle \tag{4}$$

The structural bridge between raw moments ($\langle \ldots \rangle$) and cumulants ($\langle\langle \ldots \rangle\rangle$) is governed by exact set-partition combinatorics (**集合划分**):

$$\langle x_1 x_2 \ldots x_k \rangle = \sum_{\pi \in \mathcal{P}(\{1,\ldots,k\})} \prod_{B \in \pi} \langle\langle \prod_{i \in B} x_i \rangle\rangle \tag{5}$$

Deconstruction of the Notation:

- $\mathcal{P}(\{1, \ldots, k\})$: The set of all possible ways to partition k elements into non-empty, non-overlapping groups.
- π: One specific partition scheme (e.g., separating $\{1, 2, 3, 4\}$ into $\{1, 2\}$ and $\{3, 4\}$).
- $B \in \pi$: A single cluster or "block" within that partition scheme.

The Generating Mechanism: Why Calculus Is Isomorphic to Combinatorics

Why does set-partition combinatorics appear here? It stems entirely from the fact that $Z(j) = \exp(W(j))$ and the recursive application of the Chain Rule and Product Rule:

Let's sequentially differentiate $Z(j)$ with respect to j and evaluate at $j = 0$ (where $Z(0) = 1$):

- **1st Derivative:**

$$Z'(j) = W'(j)\exp(W(j)) = W'(j)Z(j) \implies \langle x \rangle = \langle\langle x \rangle\rangle$$

- **2nd Derivative:** (Product Rule)

$$Z''(j) = W''(j)Z(j) + W'(j)Z'(j) = [W'' + (W')^2]Z(j) \implies \langle x^2 \rangle = \langle\langle x^2 \rangle\rangle + \langle\langle x \rangle\rangle^2$$

- **3rd Derivative:**

$$Z'''(j) = [W''' + 3W''W' + (W')^3]Z(j) \implies \langle x^3 \rangle = \langle\langle x^3 \rangle\rangle + 3\langle\langle x^2 \rangle\rangle\langle\langle x \rangle\rangle + \langle\langle x \rangle\rangle^3$$

The Deep Connection (Faà di Bruno's Formula)

Because the derivative of $\exp(W)$ replicates itself, the 4^{th} derivative will naturally generate the exact coefficients (1, 4, 3, 6, 1) as the Leibniz product combinations gather up.

Combinatorial Mapping: The 4th-Order Expansion

Let's expand the 4^{th}-order raw moment $\langle x^4 \rangle$ assuming a zero-mean system ($\langle x \rangle = 0$, which eliminates all terms containing a single $\langle x \rangle$):

$$\langle x^4 \rangle = \underbrace{\langle\langle x^4 \rangle\rangle}_{\text{True 4-body Interaction}} + \underbrace{3\langle\langle x^2 \rangle\rangle^2}_{\text{Disconnected Pairs}}$$

If we do *not* assume a zero mean, the full combinatorics yields:

$$\langle x^4 \rangle = \langle\langle x^4 \rangle\rangle + 4\langle\langle x^3 \rangle\rangle\langle\langle x \rangle\rangle + 3\langle\langle x^2 \rangle\rangle^2 + 6\langle\langle x^2 \rangle\rangle\langle\langle x \rangle\rangle^2 + \langle\langle x \rangle\rangle^4$$

Now that the mathematical machinery of $Z(j)$ and $W(j)$ is defined, we establish the formal rosetta stone between Statistics and Statistical Field Theory (SFT):

Entity	Statistical Identity	Field-Theoretic Intuition
Action $S(x)$	$\ln p(x)$	The Energy Landscape (**势能面**)
Propagator Δ	Covariance Matrix Σ	Fundamental Link (**传播子**)
Partition Func $Z(j)$	$\langle e^{j^\top x} \rangle$	Sum over all topologies
Free Energy $W(j)$	$\ln Z(j)$	Generator of connected processes

Note: The logarithm ($\ln Z$) acts as a mathematical filter that systematically removes disconnected vacuum bubbles, leaving only the connected physical interactions.

Derivation: The Multivariate Gaussian Field (1/2)

Let's observe how the SFT formalism operates on a simple Gaussian measure ($x \sim \mathcal{N}(0, \Delta)$).

Step 1: The Action (Potential)
Given $p(x) \propto \exp(-\frac{1}{2}x^\top \Delta^{-1} x)$, the Action is a quadratic form:

$$S(x) = -\frac{1}{2}x^\top \Delta^{-1} x \tag{6}$$

Step 2: The Partition Function (Integral)
The Laplace transform $Z(j) = \langle \exp(j^\top x) \rangle$ yields:

$$Z(j) = \int \cdots = \exp\left(\frac{1}{2} j^\top \Delta j\right) \tag{7}$$

Step 3: The Free Energy (Connected Essence)

Taking the logarithm extracts the Cumulants (**累积量**):

$$W(j) = \ln Z(j) = \frac{1}{2} \sum_{i,j} j_i \Delta_{ij} j_j \tag{8}$$

Analysis: $W(j)$ contains *only* the quadratic term (Δ_{ij}). All higher-order cumulants $\kappa_n = 0$ for $n \geq 3$.

Conclusion: Because only the 2-point interaction (Propagator Δ) exists, a Gaussian field is defined as a Free Field (**自由场**) with no multiparticle scattering vertices.

Gaussian Distribution and Wick's Theorem

This algebraic property of the Gaussian Taylor series is formalized as Wick's Theorem, which completely reduces calculus integrals to combinatorial topology.

Wick's Theorem

For a Gaussian measure, all odd-order moments are identically zero (as there are no odd powers of j in $Z_0(j)$). Any even-order raw moment decomposes into the sum of all possible pairwise product permutations of the propagator matrix Δ_{ij}:

$$\langle x_{i_1} x_{i_2} \dots x_{i_{2n}} \rangle_0 = \sum_{\text{pairings } P} \prod_{(k,l) \in P} \Delta_{i_k i_l} \tag{9}$$

The Total Number of Pairings: For $2n$ elements, the number of unique ways to contract them into pairs is given by the double factorial:

$$(2n - 1)!! = (2n - 1) \times (2n - 3) \times \cdots \times 3 \times 1 \tag{10}$$

Explicit Example: The 4-Point Function

Let's evaluate the 4-point raw expectation value $\langle x_1 x_2 x_3 x_4 \rangle_0$ using Wick's Theorem. The total number of unique pairings is $(4-1)!! = 3!! = 3$ combinations:

1. **Pairing (1,2) and (3,4):** Connect $1 \leftrightarrow 2$ and $3 \leftrightarrow 4 \implies \Delta_{12}\Delta_{34}$
2. **Pairing (1,3) and (2,4):** Connect $1 \leftrightarrow 3$ and $2 \leftrightarrow 4 \implies \Delta_{13}\Delta_{24}$
3. **Pairing (1,4) and (2,3):** Connect $1 \leftrightarrow 4$ and $2 \leftrightarrow 3 \implies \Delta_{14}\Delta_{23}$

$$\langle x_1 x_2 x_3 x_4 \rangle_0 = \Delta_{12}\Delta_{34} + \Delta_{13}\Delta_{24} + \Delta_{14}\Delta_{23} \tag{11}$$

Beautiful Consistency with Section 2

Recall from our set-partition formula that for a general system:

$$\langle x^4 \rangle = \langle\langle x^4 \rangle\rangle + 3\langle\langle x^2 \rangle\rangle^2$$

Since the Gaussian field is a Free Field ($\langle\langle x^4 \rangle\rangle = 0$), the formula leaves exactly $3\langle\langle x^2 \rangle\rangle^2 = 3\Delta^2$. Wick's theorem arrives at the exact same coefficient 3 by topologically drawing lines!

Perturbation Expansion and Feynman Diagrams

Perturbation Theory: The Gaussian Trick

Gaussian models are analytically tractable.

Real systems often contain nonlinear interactions:

$$S(x) = S_0(x) + \epsilon V(x)$$

The expectation of an observable becomes

$$\langle O \rangle = \frac{1}{Z} \int O(x) e^{-S(x)} dx$$

Move back to Gaussian averages. Define $Z_0 = \int e^{-S_0(x)} dx$. Then

$$\langle O \rangle = \frac{\langle O e^{-\epsilon V} \rangle_0}{\langle e^{-\epsilon V} \rangle_0}$$

Expand

$$e^{-\epsilon V} = 1 - \epsilon V + \frac{\epsilon^2}{2} V^2 + \cdots$$

Now everything reduces to Gaussian polynomial averages.

From Wick Contractions to Feynman Rules

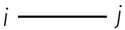
Direct Wick expansion rapidly becomes expensive.

$$10 \text{ variables} \quad \Longrightarrow \quad (10-1)!! = 945.$$

Feynman diagrams organize Wick contractions geometrically.

Propagator

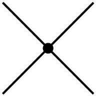
$$\langle x_i x_j \rangle_0 = \Delta_{ij}$$



Gaussian contraction

Interaction Vertex

$$V(x) \sim x^4$$



Quartic interaction

Concrete Example: Connected vs Disconnected

Compute the first-order correction: $\langle x_a x_b V(x) \rangle_0$.
Suppose $V(x) \sim x^4$. Six variables generate $(6-1)!! = 15$ possible pairings.

Connected Diagram

$$\Delta_{ai}\Delta_{bj}\Delta_{kl}$$

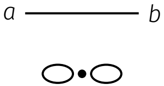
Path exists:

$$a \to b$$

Interaction affects the observable.

Disconnected Diagram

$$\Delta_{ab}\Delta_{ij}\Delta_{kl}$$



Vacuum bubble:
interaction talks only to itself.
Cancels by normalization.

Key Insight

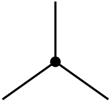
Only connected diagrams contribute to connected correlation functions.

Generalization: Encoding Different Non-linearities

We used ϕ^4 (symmetric) as an example. Different neural network activation functions (or physical potentials) generate different vertex structures.

Asymmetric (ϕ^3 theory)

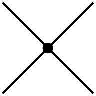
$$V(\phi) = \frac{g}{3!}\phi^3$$



Introduces **skewness**. Three propagators meet. Can generate complex sub-graphs (e.g., joining two ϕ^3 vertices to form a loop).

Symmetric (ϕ^4 theory)

$$V(\phi) = \frac{\lambda}{4!}\phi^4$$



Introduces **kurtosis**. Four propagators meet. (The standard model we just explored).

The Universal Rule

Polynomial order determines vertex legs.

$$\phi^n \implies n\text{-leg interaction vertex}$$

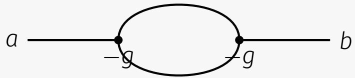
What if our system breaks parity symmetry ($x \to -x$)? We introduce a cubic term, expanding our Feynman toolkit into the $\phi^3 + \phi^4$ Theory:

$$S(x) = -\frac{1}{2}\sigma^{-2}x^2 - \underbrace{\frac{g}{3!}x^3}_{\text{Skewness}} - \underbrace{\frac{\lambda}{4!}x^4}_{\text{Kurtosis}} \tag{12}$$

We now have **two** distinct geometric building blocks.

Diagrammatic Synergy

New composite topologies emerge! A 2-point correlation (self-energy) can now be built by connecting two 3-point vertices:



Computing a Real Observable: The Setup for $\langle x_a x_b \rangle$

Let's compute the actual 2-point correlation $\langle x_a x_b \rangle$ under our $\phi^3 + \phi^4$ theory. By the Linked Cluster Theorem, we drop the denominator and keep **only connected graphs**:

$$\langle x_a x_b \rangle = \left\langle x_a x_b \exp\left(-\frac{\lambda}{4!}x^4 - \frac{g}{3!}x^3\right)\right\rangle_{0,\text{ connected}} \quad (13)$$

Taylor Expanding the Interaction (up to 1-loop order):

$$\langle x_a x_b \rangle_{\text{conn}} = \underbrace{\langle x_a x_b \rangle_0}_{\text{Bare}} - \underbrace{\frac{\lambda}{4!}\langle x_a x_b x^4 \rangle_{0,c}}_{\text{1st Order } \lambda}$$
$$- \underbrace{\frac{g}{3!}\langle x_a x_b x^3 \rangle_{0,c}}_{\text{1st Order } g} + \underbrace{\frac{1}{2!}\left(\frac{g}{3!}\right)^2 \langle x_a x_b x_1^3 x_2^3 \rangle_{0,c}}_{\text{2nd Order } g^2}$$

The Odd-Even Selection Rule

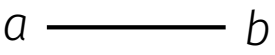
Notice the 1st Order g term: $\langle x_a x_b x^3 \rangle_0$. This contains 5 variables. By Wick's Theorem, it is impossible to perfectly pair an odd number of fields! Therefore, **this term is strictly ZERO.** The cubic skewness (g) only contributes at even orders (e.g., g^2).

Computing a Real Observable: Diagrammatic Evaluation

The final observable is just the sum of these valid connected diagrams:

0th Order (Bare)

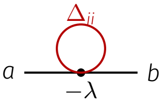
$\langle x_a x_b \rangle_0$



Value: Δ_{ab}

Order λ (ϕ^4 Snail)

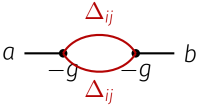
$-\frac{\lambda}{4!}\langle x_a x_b x_i^4 \rangle_{0,c}$



Value: $-\frac{\lambda}{2}\sum_i \Delta_{ai}\Delta_{ii}\Delta_{ib}$

Order g^2 (ϕ^3 Sunset)

$+\frac{g^2}{2(3!)^2}\langle x_a x_b x_i^3 x_j^3 \rangle_{0,c}$



Value: $\frac{g^2}{2}\sum_{ij} \Delta_{ai}\Delta_{ij}^2\Delta_{jb}$

The Final Answer (Matrix Notation)

If we let $\mathbf{D}_{\text{loop}}$ be a diagonal matrix of variances $(\mathbf{D}_{\text{loop}})_{ii} = \Delta_{ii}$, the exact 2-point correlation matrix is simply the sum of matrix products:

$$_{\text{true}} \approx \mathbf{\Delta} - \frac{\lambda}{2}\mathbf{\Delta}\mathbf{D}_{\text{loop}}\mathbf{\Delta} + \frac{g^2}{2}\mathbf{\Delta}(\mathbf{\Delta} \circ \mathbf{\Delta})\mathbf{\Delta} + \dots \quad (14)$$

Conclusion: The non-linear integration is bypassed! We just multiply the bare covariance $\mathbf{\Delta}$ with variance/Hadamard topological matrices.

Linked Cluster Theorem

The Problem: Vacuum Bubbles

Recall from our diagrammatic evaluation: Wick contractions generate both **connected diagrams** and **disconnected vacuum bubbles** (interactions talking only to themselves).

Because disconnected pieces are statistically independent events, their probability amplitudes factorize algebraically:

$$Z(j) \sim (\text{Connected Contributions}) \times (\text{Vacuum Bubbles})$$

Why are they annoying?

If we don't explicitly remove them, the number of disconnected vacuum topologies grows factorially, causing the perturbative series to blow up to infinity. They are mathematical "noise."

Factorization: Cancelling the Vacuum

Let's formalize this factorization:

$$Z(j) = Z_{\mathrm{conn}}(j) \times Z_{\mathrm{vac}}$$

where Z_{vac} collects the infinite sum of all vacuum bubble diagrams.

When we turn off the external source ($j = 0$), all connected parts vanish because there are no external variables to connect to, leaving exactly:

$$Z(0) = Z_{\mathrm{vac}}$$

Therefore, the **normalized** generating functional automatically annihilates the vacuum background:

$$\frac{Z(j)}{Z(0)} = \frac{Z_{\mathrm{conn}}(j) \times Z_{\mathrm{vac}}}{Z_{\mathrm{vac}}} = Z_{\mathrm{conn}}(j)$$

The Logarithm: Isolating Connected Diagrams

To avoid calculating divisions constantly, we switch our focus to the Free Energy operator:

$$W(j) = \ln Z(j)$$

This is the crucial step: $\ln(AB) = \ln A + \ln B$. The logarithm turns the product of independent graphs into a sum, effectively stripping away the exponentiated disconnected structures.

The Linked Cluster Theorem

Taking the logarithm of the partition function isolates **only** the fully connected topologies:

$$W(j) = \sum_{\Gamma \text{ connected}} \mathcal{A}_\Gamma(j)$$

We never have to draw or compute a disconnected graph again.

We have now come full circle to Chapter 2. We already proved that $Z(j)$ generates **Raw Moments** ($\langle\ldots\rangle$), and $W(j) = \ln Z(j)$ generates **Cumulants** ($\langle\langle\ldots\rangle\rangle$).

The Linked Cluster Theorem gives this rigorous statistical relationship a beautiful topological meaning:

Moments ($\langle\ldots\rangle$)	**Cumulants** ($\langle\langle\ldots\rangle\rangle$)
Mixes everything together.	Filters the irreducible core.
All Diagrams (Connected + Vacuum Bubbles)	Connected Diagrams Only

The Ultimate Takeaway

Cumulants are not just abstract statistical variances—they are the **exact mathematical equivalent of physically connected Feynman diagrams**.

Conclusion & Methodological Outlook

Summary: Foundations Laid in Ch. 1-5

1. **Bridge Formed:** Mapped Mathematical Statistics (Likelihood, Laplace Transform, $\mathbb{E}$) to Physics (Action, Partition Function, $\langle \ldots \rangle$).
2. **Calculus to Topology:** Wick's theorem proves that high-dimensional multivariate integration can be replaced by visual graph assembly.
3. **Non-linearities Managed:** Perturbation theory and the Linked Cluster Theorem (via Cumulants and $\ln Z$) provide a rigorous way to evaluate non-Gaussian networks using connected Feynman diagrams.