

An Introduction to Information Field Theory

Joe Gillett, Uli Raudales
Department of Physics, Whitman College

July 2024

1 Information Field Theory Basics

1.1 Introduction

Throughout physics, astronomy and engineering, countless problems exist which can be expressed in the form of continuous random fields. Such fields model systems which exhibit spatial and temporal variation. Many of these fields follow a given distribution, namely the Gaussian distribution, with properties of the field being determined by inferring the parameters of the first two moments (mean and variance) of the distribution. Some systems, however, fail to follow a known distribution, and are thus more difficult to solve for.

Irrespective of the form of the model, inference over a continuous field is typically required, as data is collected at discrete points on the domain. Bayes theorem can be used for this purpose, given by the following equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

While this approach works well for parameterized random fields, it becomes infeasible in the case of a non-parameterized, non-Gaussian distribution. Calculation of the evidence $P(B)$ requires integrating the joint probability $P(B|A)P(A)$ over all possible values of the field, which is defined on a continuous underlying space. Since this is impractical computationally, another approach must be found. Conventionally, Markov Chain Monte Carlo (MCMC) or related numerical techniques are used in order to circumvent this issue and infer the moments. While this procedure certainly works, it becomes increasingly computationally intensive as the number of parameters increases.

Presented is an alternative method of calculating the moments using Bayes-derived techniques from Information Field Theory. Diagrammatic expansion, one such technique, characterizes the posterior moments as an infinite sum of diagrams, each of which corresponds to an equation dependant on the diagram's structure. While this technique does require discretization of the field, the result approaches the continuous limit given sufficient resolution. Additionally, direct calculation of the evidence using a similar technique involving

closed diagram expansion is also explored, alongside a potential set of basis functions for more efficient calculations. It should be noted that these techniques require the posterior random field to be perturbatively non-Gaussian, or in other words deviate only slightly from the Gaussian distribution. It is possible that these techniques may work on other named distributions, but to date this has not been tested.

1.2 The Diagrammatic Expansion Technique

The motivation behind many of information field theoretic techniques stem from the mathematical similarity between Bayesian inference and that of probabilistic treatment of physical fields. Similar to the Bayes Formula (Eq(1)) the probability of a physical field being in a certain configuration has been expressed in the following form, often called the Gibbs measure,

$$P(\mathbf{x}) = \frac{e^{H(\mathbf{x})}}{Z(\mathbf{s})} \equiv \frac{P(B|A)P(A)}{P(B)} \quad (2)$$

where $H(\mathbf{x})$ is the Hamiltonian of the field, and Z is the partition function. The Hamiltonian is a functional of the field, and is often expressed as a sum of terms, each of which is a function of the field at a certain point in space.

$$H(\mathbf{x}) = \sum_i H_i(\mathbf{x}) \equiv -\log P(B|A)P(A) \quad (3)$$

The partition function is the normalization constant, and is given by the integral of the exponential of the Hamiltonian over all possible field configurations.

$$Z(\mathbf{s}) = \int_{-\infty}^{\infty} d\mathbf{x} e^{-H(\mathbf{x}) + \mathbf{s}^T \mathbf{x}} \quad (4)$$

where \mathbf{s} is a book-keeping vector that keeps track of the moments of the field. The term $\mathbf{s}^T \mathbf{x}$ acts as a perturbation in the Hamiltonian. Given the similarity between the partition function and the moment-generating function, it is possible to calculate the moments of the field by taking derivatives of the partition function with respect to the book-keeping vector.

$$\langle x_1 \dots x_n \rangle = \mu_n = \frac{1}{Z(\mathbf{s})} \frac{\partial^n Z(\mathbf{s})}{\partial s_1 \dots \partial s_n} \bigg|_{\mathbf{s}=0} = \frac{\partial^n \log Z(\mathbf{s})}{\partial s_1 \dots \partial s_n} \bigg|_{\mathbf{s}=0} \quad (5)$$

For most practical cases, the partition function $Z(\mathbf{s})$ cannot be computed in a closed-form analytical expression, making it highly difficult to extract the moments of the distribution using Eq(4). However, the diagrammatic expansion technique provides us a method to extract information about the moments from the partition function up to a certain order of accuracy. To do so, we begin by expanding the Hamiltonian in a Taylor series expansion,

$$H(\mathbf{x}) = H_0 - \mathbf{j}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{D}^{-1} \mathbf{x} + \sum_{n=3}^{\infty} \frac{1}{n!} \Lambda_{x_1 \dots x_n}^{(n)} \mathbf{x}^n \quad (6)$$

where

$$\mathbf{j}(x_1) = -\frac{\partial H}{\partial x_1} \Big|_{\mathbf{x}=0} \quad (7)$$

$$\mathbf{D}^{-1}(x_1, x_2) = \frac{\partial^2 H}{\partial x_1 \partial x_2} \Big|_{\mathbf{x}=0} \quad (8)$$

$$\Lambda_{x_1 \dots x_n}^{(n)} = \frac{\partial^n H}{\partial x_1 \dots \partial x_n} \Big|_{\mathbf{x}=0} \quad (9)$$

In this expression, $\mathbf{j}(x_1)$ represents the source of information where the field's influence is most pronounced. The term $\mathbf{D}^{-1}(x_1, x_2)$ acts as the propagator, functioning similarly to a covariance, facilitating the influence of the random field at location x_1 on the field at location x_2 . The higher-order interactions are encapsulated by the tensors $\Lambda_{x_1 \dots x_n}^{(n)}$, referred to as the n th-order interaction coefficients. These coefficients quantify the contribution's magnitude of the n th power of the random field within the Hamiltonian framework. Repeated indices are integrated over (Einstein summation convention), and the n th-order interaction coefficients are symmetric under permutation of their indices.

$$\Lambda_{x_1 \dots x_n}^{(n)} x_1 \dots x_n = - \int_{-\infty}^{\infty} dx_1 \dots dx_n \Lambda_{x_1 \dots x_n}^{(n)} x_1 \dots x_n \quad (10)$$

In a discrete setting, \mathbf{j} is a vector, \mathbf{D}^{-1} is a matrix, and $\Lambda^{(n)}$ is a tensor. For discrete cases we summed over repeated indices as per the Einstein summation convention. For example,

$$M_{ij} A_i B_j = \sum_{i,j} M_{ij} A_i B_j$$

To solve this numerically, we will be dealing with the discrete cases.

We break the Hamiltonian into two parts,

$$H_G(\mathbf{x}) = H_0 - \mathbf{j}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{D}^{-1} \mathbf{x}$$

which is the Gaussian part of the Hamiltonian, and

$$H_I(\mathbf{x}) = \sum_{n=3}^{\infty} \frac{1}{n!} \Lambda_{x_1 \dots x_n}^{(n)} \mathbf{x}^n$$

which is the interaction part of the Hamiltonian. Following this, we can write the partition function as a sum of the Gaussian and interaction partition functions,

$$Z(\mathbf{s}) = \int_{-\infty}^{\infty} d\mathbf{x} e^{-H(\mathbf{x}) + \mathbf{s}^T \mathbf{x}} \quad (11)$$

$$= \int_{-\infty}^{\infty} d\mathbf{x} e^{-(H_0 - \mathbf{j}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{D}^{-1} \mathbf{x} + \sum_{n=3}^{\infty} \frac{1}{n!} \Lambda_{x_1 \dots x_n}^{(n)} \mathbf{x}^n) + \mathbf{s}^T \mathbf{x}} \quad (12)$$

we note that since \mathbf{s} enters with a linear power of \mathbf{x} in the Gaussian exponential, and hence derivatives with respect to \mathbf{s} will bring down powers of \mathbf{x} .

$$Z(\mathbf{s}) = \exp \left[- \sum_{n=3}^{\infty} \frac{1}{n!} \Lambda_{x_1 \dots x_n}^{(n)} \frac{\partial^n}{\partial x_1 \dots \partial x_n} \right] \int_{-\infty}^{\infty} d\mathbf{x} e^{-H_G(\mathbf{x}) + \mathbf{s}^T \mathbf{x}} \quad (13)$$

$$= \exp \left[-H_I \left(\frac{\partial^n}{\partial x_1 \dots \partial x_n} \right) \right] Z_G(\mathbf{s}) \quad (14)$$

1.3 Feynman Diagrams

Borrowing from Quantum Field Theory, we can represent the partition function as a sum of diagrams, each of which corresponds to a term in the expansion of the partition function.

Given that the Gaussian partition function takes the form

$$\int_{-\infty}^{\infty} dx e^{-ax^2 + bx + c} = \sqrt{\frac{\pi}{a}} e^{\frac{b^2}{4a} + c} \quad (15)$$

it can show that the Gaussian partition function has the following exact form,

$$Z_G(\mathbf{s}) = \sqrt{|2\pi \mathbf{D}|} \exp \left[\frac{1}{2} (\mathbf{s} + \mathbf{j})^T \mathbf{D} (\mathbf{s} + \mathbf{j}) - H_0 \right] \quad (16)$$

To get the diagrammatic expansion of the partition function, we take the natural logarithm of the partition function

$$\log Z_G(\mathbf{s}) = \frac{1}{2} (\mathbf{s}^T \mathbf{D} \mathbf{s} + \mathbf{j}^T \mathbf{D} \mathbf{s} + \mathbf{s}^T \mathbf{D} \mathbf{j} + \mathbf{j}^T \mathbf{D} \mathbf{j}) - H_0 + \log \sqrt{|2\pi \mathbf{D}|} \quad (17)$$

We represent each information source \mathbf{j} as \mathbf{x} and each external source \mathbf{s} as \bullet and each propagator \mathbf{D} as --- and each interaction $\Lambda^{(n)}$ as a vertex with n lines coming out of it. Thus, the diagrammatic expansion of the Gaussian partition function is given by

$$\log Z_G(\mathbf{s}) = \frac{1}{2} (\bullet \text{---} \bullet + \bullet \text{---} \mathbf{x} + \mathbf{x} \text{---} \bullet + \mathbf{x} \text{---} \mathbf{x}) - H_0 + \log \sqrt{|2\pi \mathbf{D}|} \quad (18)$$

Getting the moments of the field is now a matter of taking derivatives of the partition function with respect to the book-keeping vector \mathbf{s} and evaluating at $\mathbf{s} = 0$. This is equivalent to counting the number of external sources in the diagram with respect to which we are taking the derivative and removing them from the diagram and evaluating the diagram at $\mathbf{s} = 0$. So the first moment of the field is given by

$$\mu_1 = \left. \frac{\partial \log Z_G(\mathbf{s})}{\partial s_1} \right|_{\mathbf{s}=0} = \mathbf{D} \mathbf{j} = \text{---} \mathbf{x} = \int_{-\infty}^{\infty} dx_1 D(x, x_1) j(x_1) \quad (19)$$

Finding the diagrammatic expansion of the Interactive partition function just requires one to find which nth-order terms contribute (i.e $\Lambda^{(n)} \neq 0$) and then sum over all possible diagrams that can be formed from these terms. Each diagram is multiplied by a symmetry factor, which is the number of ways the diagram can be drawn without changing its appearance.

2 IFT in Practice

2.1 1D case

This section explores the implementation of the diagrammatic expansion technique in the 1-dimensional case. While the problem itself is trivial, it should allow for an easy-to-follow, verifiable example of the technique, which can subsequently be used in higher-dimensional cases. Rather than beginning with a data set, instead consider the following non-Gaussian probability density function:

$$Q(x) = \exp \left[\frac{-(x - x_0)^2}{2\sigma^2} - bx^4 \right] \quad (20)$$

This probability distribution is largely Gaussian, with the $-bx^4$ term slightly perturbing the system. The normalized form of the probability density function is given by:

$$P(x) = \frac{Q(x)}{Z} = \frac{e^{-H(x)}}{\int_{-\infty}^{\infty} e^{-H(x)} dx} \quad (21)$$

Thus we have that the Hamiltonian is given by:

$$H(x) = \frac{(x - x_0)^2}{2\sigma^2} + bx^4 \quad (22)$$

Where the second term is the interaction term due to the perturbation. Taylor expanding the Hamiltonian about $x = 0$ gives:

$$H(x) = \frac{x^2}{2\sigma^2} - \frac{x_0^2}{\sigma^2}x + \frac{1}{\sigma^2}x^2 + 4!bx^4 \quad (23)$$

where

$$\begin{aligned} j &= -\frac{\partial H}{\partial x} \Big|_{x=0} = \frac{x_0^2}{\sigma^2} \\ D^{-1} &= \frac{\partial^2 H}{\partial x^2} \Big|_{x=0} = \frac{1}{\sigma^2} \\ \Lambda^{(4)} &= \frac{\partial^4 H}{\partial x^4} \Big|_{x=0} = 4!b \\ \Lambda^{(n \geq 5)} &= \Lambda^{(3)} = 0 \end{aligned}$$

Since the interaction Hamiltonian is the sum over all possible diagrams formed by the contributing n th-order terms, we have contributions from diagrams $(\Lambda^{(4)})^n$ for $n \geq 1$. So we will have contributions from n th-order terms that are powers of $4n$.

So the log of the partition function would be the sum of all diagrams that can be formed from the n th-order terms and the Gaussian partition function.

To better understand, let us calculate the first moment of the field, μ_1 .

$$\mu_1 = \text{---}\mathbf{x} + \frac{1}{3!} \text{---}\mathbf{x} \begin{array}{c} \mathbf{x} \\ \diagup \quad \diagdown \\ \mathbf{x} \end{array} + \frac{1}{4!} \mathbf{x} \begin{array}{c} \mathbf{x} \\ | \\ \mathbf{x} \end{array} \mathbf{x} + \frac{1}{2} \text{---}\mathbf{x} \begin{array}{c} \text{---}\mathbf{x} \\ \diagup \quad \diagdown \\ \mathbf{x} \end{array} + \frac{1}{8} \text{---}\mathbf{x} \begin{array}{c} \text{---}\mathbf{x} \\ \diagup \quad \diagdown \\ \text{---}\mathbf{x} \end{array} + \frac{1}{8!} \mathbf{x} \begin{array}{c} \mathbf{x} \\ | \\ \mathbf{x} \end{array} \text{---}\mathbf{x} \begin{array}{c} \mathbf{x} \\ | \\ \mathbf{x} \end{array} \mathbf{x} + \dots \quad (24)$$

$$= Dj + \frac{1}{3!} D^3 \Lambda^{(4)} j^3 + \frac{1}{4!} D^4 \Lambda^{(4)} j^4 + \frac{1}{2} D^2 \Lambda^{(4)} j + \frac{1}{8} D^2 \Lambda^{(4)} + \frac{1}{8!} D^6 (\Lambda^{(4)})^2 j^7 + \dots \quad (25)$$

Since this is a 1-dimensional case, the propagator D is just a scalar, and the interaction term $\Lambda^{(4)}$ is also a scalar.

2.2 Linear Regression

Mathematically, a linear regression model combines a deterministic mean component with an independent error term. This error term, accounting for model uncertainty or measurement errors, is typically modeled as a Gaussian distribution with zero mean and standard deviation σ . The model can be expressed as:

$$y_i = X_i \boldsymbol{\beta} + \sigma \epsilon_i \quad (26)$$

where y is the observed data, \mathbf{X} is the design matrix, $\boldsymbol{\beta}$ is the vector of regression coefficients, and $\boldsymbol{\epsilon}$ is the error term. The goal of linear regression is to estimate the regression coefficients $\boldsymbol{\beta}$ given the observed data y and the design matrix \mathbf{X} .

The natural logarithm of the likelihood function is given by:

$$\log P(\mathbf{y}|\boldsymbol{\beta} = \boldsymbol{\Theta}) = -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \frac{N}{2} \log(2\pi\sigma^2) \quad (27)$$

with the parameter space defined as $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \log \sigma^2\} \equiv \{\boldsymbol{\theta}_1, \theta_2\}$. We let $\sigma^2 = \exp(\theta_2)$.

The Hamiltonian for this model, which combines the prior and likelihood functions, is given by (we have dropped the irrelevant constant terms which are independent of $\boldsymbol{\Theta}$):

$$H(\boldsymbol{\Theta}) = \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_1)^T \exp(-\theta_2) (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_1) + \frac{N}{2} \theta_2 + \frac{1}{2} \boldsymbol{\theta}_1^T \mathbf{S}^{-1} \boldsymbol{\theta}_1 + \frac{1}{2} \frac{\theta_2^2}{\sigma_2^2} \quad (28)$$

Taylor expanding the Hamiltonian about $\boldsymbol{\Theta} = 0$ gives:

$$H(\boldsymbol{\Theta}) = H_0 + \mathbf{j}^T \boldsymbol{\Theta} + \frac{1}{2} \boldsymbol{\Theta}^T \mathbf{D}^{-1} \boldsymbol{\Theta} + \sum_{n=3}^{\infty} \frac{1}{n!} \Lambda^{(n)} \boldsymbol{\Theta}^n \quad (29)$$

where

$$\begin{aligned}
\mathbf{j} &= \begin{pmatrix} \exp(-\theta_2) \mathbf{X}^T \hat{\mathbf{y}} - \mathbf{S}^{-1} \boldsymbol{\theta}_1 \\ \frac{\exp(-\theta_2)}{2} \hat{\mathbf{y}}^T \hat{\mathbf{y}} - \frac{N}{2} - \frac{\theta_2}{\sigma_2^2} \end{pmatrix} \\
\mathbf{D}^{-1} &= \begin{pmatrix} \exp(-\theta_2) \mathbf{X}^T \mathbf{X} + \mathbf{S}^{-1} & \exp(-\theta_2) \mathbf{X}^T \hat{\mathbf{y}} \\ \exp(-\theta_2) \mathbf{X}^T \hat{\mathbf{y}} & \frac{\exp(-\theta_2)}{2} \hat{\mathbf{y}}^T \hat{\mathbf{y}} + \frac{1}{\sigma_2^2} \end{pmatrix} \\
\Lambda_{[0,n]}^{(n)} &= (-1)^n \frac{\exp(-\theta_2)}{2} \hat{\mathbf{y}}^T \hat{\mathbf{y}} \\
\Lambda_{[1,n-1]}^{(n)} &= (-1)^n \exp(-\theta_2) \mathbf{X}^T \hat{\mathbf{y}} \\
\Lambda_{[2,n-2]}^{(n)} &= (-1)^n \exp(-\theta_2) \mathbf{X}^T \mathbf{X}
\end{aligned}$$

where the adjusted prediction $\hat{\mathbf{y}}$ is defined as $\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_1$. The equations for the interaction coefficients incorporate $\Lambda^{(n)}$, an n-rank tensor that aggregates all nth derivative terms of the Hamiltonian. Since the Hamiltonian possesses continuous second-order derivatives, the sequence of differentiation does not alter the outcome, leading to identical values for multiple tensor elements. This redundancy is captured by the notation $\Lambda_{[a,b]}^{(n)}$, signifying tensor entries that correspond to 'a' number of derivatives with respect to θ_1 and 'b' number of derivatives with respect to θ_2 .

From here we can calculate the moments of the field, and subsequently the regression coefficients, using the diagrammatic expansion technique. The partition function will be just the sum of all diagrams that can be formed from the nth-order terms and the Gaussian partition function.

2.3 The MAP estimate and Closed Diagrams

The Maximum A Posteriori (MAP) estimate is a statistical technique used to estimate the mode of the posterior distribution of a parameter vector given the data. In mathematical terms, the MAP estimate is found by setting the gradient of the Hamiltonian to zero:

$$\nabla H|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0 \quad (30)$$

The MAP estimate is crucial for several reasons:

- **Simplification of Calculations:** By centering our parameter estimates around the MAP, we can significantly simplify the problem of calculating the full posterior distribution. The MAP estimate serves as a central point around which the posterior distribution can be approximated, often reducing the computational complexity.
- **Effective Approximation:** In cases where the posterior distribution is nearly Gaussian, the MAP estimate is a good approximation for the mean of the distribution. This makes it particularly useful in simplifying higher-order calculations.
- **Perturbative Expansion:** The MAP estimate allows us to perform a perturbative expansion around the most probable point. This means we can focus on small devia-

tions (perturbations) from the MAP estimate, which are easier to handle analytically and computationally.

When we subtract the MAP estimate from the parameter vector being inferred, we obtain a shifted parameter vector

$$\mathbf{\Delta} = \mathbf{\Theta} - \mathbf{\Theta}_0 = \begin{pmatrix} \theta_1 - \theta_{1,0} \\ \theta_2 - \theta_{2,0} \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} \quad (31)$$

If we Taylor expand the Hamiltonian about $\mathbf{\Theta} = \mathbf{\Theta}_0$, we get

$$H(\mathbf{\Delta}) = H_0 + \mathbf{j}^T \mathbf{\Delta} + \frac{1}{2} \mathbf{\Delta}^T \mathbf{D}^{-1} \mathbf{\Delta} + \sum_{n=3}^{\infty} \frac{1}{n!} \mathbf{\Lambda}^{(n)} \mathbf{\Delta}^n \quad (32)$$

At the MAP estimate, the gradient of the Hamiltonian is zero. Therefore, when we Taylor expand the Hamiltonian around the MAP estimate, the linear term $\mathbf{j}^T \mathbf{\Delta}$ vanishes ($\mathbf{j} = 0$ at the MAP estimate). With the linear term gone, the only remaining contributions to the partition function come from terms involving $\mathbf{\Delta}$. These terms correspond to loop diagrams in Feynman diagram language. Loop diagrams represent self-interactions or correlations within the system that do not involve external sources.

Odd-order diagrams (third, fifth, etc.) vanish due to symmetry or other properties of the Hamiltonian. This leaves only even-order terms, which are typically easier to manage and interpret. Even n th-order diagrams can be composed of odd order diagrams, but the overall diagram will be even order.

By focusing on loop diagrams, we avoid the complexities associated with external source terms. This makes the computation of the partition function and related quantities more tractable, especially in high-dimensional parameter spaces.

$$\begin{aligned}
\log Z = & -H_0 + \text{circle} - \frac{1}{8} \text{infinity} - \frac{1}{48} \text{triskelion} + \frac{1}{8} \text{figure-eight} + \frac{1}{12} \text{double-circle} \\
& + \frac{1}{384} \text{four-leaf} + \frac{1}{96} \text{three-circles} + \frac{1}{48} \text{double-oval} + \frac{1}{32} \text{figure-eight} \\
& + \frac{1}{12} \text{figure-eight} + \mathcal{O}(\Lambda^{(10)})
\end{aligned} \tag{33}$$

$$\begin{aligned}
= & -H_0 + \log \sqrt{|2\pi \mathbf{D}|} - \frac{1}{8} \Lambda_{ijkl}^{(4)} \mathbf{D}_{ij} \mathbf{D}_{kl} - \frac{1}{48} \Lambda_{ijklmn}^{(6)} \mathbf{D}_{ij} \mathbf{D}_{kl} \mathbf{D}_{mn} \\
& + \frac{1}{8} \Lambda_{ijk}^{(3)} \Lambda_{lmn}^{(3)} \mathbf{D}_{ij} \mathbf{D}_{kl} \mathbf{D}_{mn} - \frac{1}{384} \Lambda_{ijklmnpq}^{(8)} \mathbf{D}_{ij} \mathbf{D}_{kl} \mathbf{D}_{mn} \mathbf{D}_{pq} \\
& + \frac{1}{96} \Lambda_{ijkl}^{(4)} \Lambda_{mnop}^{(4)} \mathbf{D}_{ij} \mathbf{D}_{km} \mathbf{D}_{ln} \mathbf{D}_{op} + \frac{1}{48} \Lambda_{ijkl}^{(4)} \Lambda_{mnop}^{(4)} \mathbf{D}_{im} \mathbf{D}_{jn} \mathbf{D}_{ko} \mathbf{D}_{lp} \\
& + \frac{1}{32} \Lambda_{ijklm}^{(5)} \Lambda_{nop}^{(3)} \mathbf{D}_{ij} \mathbf{D}_{kl} \mathbf{D}_{mn} \mathbf{D}_{op} + \frac{1}{12} \Lambda_{ijklm}^{(5)} \Lambda_{nop}^{(3)} \mathbf{D}_{ij} \mathbf{D}_{kn} \mathbf{D}_{lo} \mathbf{D}_{mp} + \mathcal{O}(\Lambda^{(10)})
\end{aligned} \tag{34}$$

We recall that, repeated indices are summed over under the Einstein summation convention. We can continue to go up till the correction term we want.

To test this numerically we used Python. Using $N = 10$, $\theta_1 = 2$, $\theta_2 = 5$, $\epsilon = \mathcal{N}(0, \sigma)$, and $\mathbf{X} = \mathcal{N}(0, 1)$, we generated the data set.

We first ran a Markov Chain Monte Carlo (MCMC) simulation to estimate the regression coefficients. The MCMC simulation yielded $\beta = 2.340$ and $\log \sigma = 3.121$. We then used the diagrammatic expansion technique to calculate the moments of the field. The technique yielded $\beta = 2.34454322$ and $\log \sigma = 3.02591931$. The two results are in good agreement, with the IFT result being slightly higher than the MCMC result.

Using the diagrammatic expansion technique, we calculated the partition function to be $\log Z = -19.958436807763317$. We then compared this to the numerical result, which was $\log Z = -19.91886605407998$. The two results are in good agreement, with the numerical result being slightly higher than the IFT result.

We compare the results of the MCMC simulation and the IFT technique in the plots above. The numbered runs represent the 4 chains runned in the MCMC simulation and P1 represents the numerical result and P2 represents the IFT result. We see that the two results are in good agreement, with the IFT results being able to capture the underlying trend in the data.

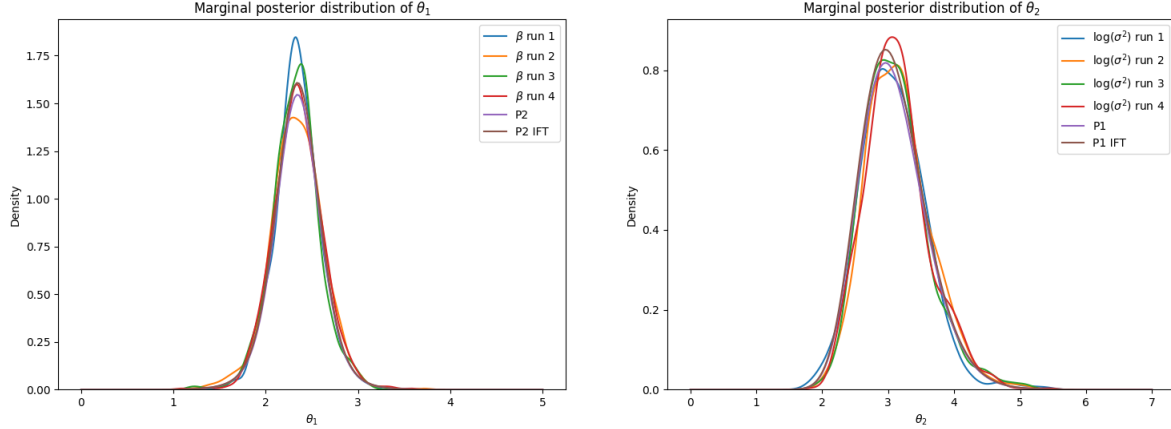


Figure 1: Linear Regression Plots

3 Conclusion

Information Field Theory is a powerful tool for inferring the moments of a random field, given a perturbatively non-Gaussian distribution. The diagrammatic expansion technique allows for the calculation of the moments of the field by summing over all possible diagrams that can be formed from the n th-order terms in the Hamiltonian. The technique is particularly useful in cases where the partition function cannot be computed in a closed-form analytical expression, as it provides a method to extract information about the moments from the partition function up to a certain order of accuracy. The technique can be applied to a wide range of problems, including linear regression, where it can be used to estimate the regression coefficients given the observed data and the design matrix. The closed diagram expansion technique allows for the calculation of the moments of the field by summing over all possible closed diagrams that can be formed from the n th-order terms in the Hamiltonian. The technique is particularly useful in cases where the partition function cannot be computed in a closed-form analytical expression, as it provides a method to extract information about the moments from the partition function up to a certain order of accuracy.

3.1 Future Work

We plan to explore the diagrammatic expansion technique in higher-dimensional cases, where the field is defined on a continuous underlying space. We also plan to investigate the use of basis functions for more efficient calculations and not need to solve complicated integrals. We wish to test the technique on other named distributions to see if it can be applied more broadly.

3.1.1 Basis Functions

The basis functions are a set of functions that span the space of the field.

We want to find functions $\phi_i(x)$ such that

$$\int dx \phi_i(x) \phi_j(x) = \delta_{ij} \quad (35)$$

so that

$$f(x) = \sum_k c_k \phi_k(x) \quad (36)$$

We want to find such functions so that the following holds

$$\int_{-\infty}^{\infty} dx f(x) P(x) = \sum_k c_k \int_{-\infty}^{\infty} dx \phi_k(x) P(x) = \sum_k c_k \int_{-\infty}^{\infty} dx \phi_k(x) \frac{e^{-H(x)}}{Z} \quad (37)$$

If we can find exponential basis functions, then the basis functions could be absorbed into the $P(x)$ term. This exponential basis functions would be of the form e^{kax} , where a is some expression, and would act as the perturbation term in the Hamiltonian.

$$\sum_k c_k \int_{-\infty}^{\infty} dx \frac{e^{kax} e^{-H(x)}}{Z} = \sum_k c_k \int_{-\infty}^{\infty} dx \frac{e^{-H(x)+kax}}{Z} \quad (38)$$

Then the integral just becomes the sum of diagrams dependant on the basis functions times the coefficients of the basis functions.

Some functions would have basis functions of the form x^{kax} , where a is some expression, and so the integral then just becomes the sum of moments of the field times the coefficients of the basis functions.