Partition Functions: Variational Bounds, Saddle-Point, and Perturbative Expansions

Danilo J. Rezende

December 24, 2015

1 Introduction

In statistical physics and probabilistic modeling, the *partition function* plays a central role in normalizing unnormalized densities and computing moments. Throughout these notes, we consider a density of the form

$$p(x) \propto \exp\bigl(-S(x)\bigr), \quad x \in \mathbb{R}^N,$$

where $S(x)$ is often referred to as an *action* or *energy*. For a real vector $\lambda \in \mathbb{R}^N$, define the partition function

$$Z(\lambda) \;=\; \int dx \, \exp\bigl(-S(x) \;+\; \lambda^T x\bigr).$$

In many contexts, one also writes $Z(\lambda) = Z\,\mathbb{E}_p[e^{\lambda^T x}]$, where

$$Z \;=\; \int dx \, e^{-S(x)}$$

is simply the partition function at $\lambda = 0$. We will discuss two main uses of $Z(\lambda)$ (and related quantities) in this document:

(i) *Variational bounds:* We show how classical Jensen-Feynman-type bounds, as well as higher-order extensions, can be used to approximate $Z(\lambda)$.

(ii) *Approximation schemes:* We describe standard approximation methods—namely, the saddle-point (or semi-classical) approximation and perturbative expansions—that enable tractable computations of $Z(\lambda)$ and its moments.

2 Partition Function and Moments

A key reason for studying $Z(\lambda)$ is that any moment $\mathbb{E}_p[x_{a_1} \cdots x_{a_n}]$ can be computed from its derivatives. Indeed,

$$\mathbb{E}_p[x_{a_1} \cdots x_{a_n}] \;=\; \frac{1}{Z} \, \frac{\partial^n}{\partial \lambda_{a_1} \cdots \partial \lambda_{a_n}} Z(\lambda)\Big|_{\lambda=0} \;=\; \frac{\partial^n}{\partial \lambda_{a_1} \cdots \partial \lambda_{a_n}} \ln \mathbb{E}_p\bigl[e^{\lambda^T x}\bigr]\Big|_{\lambda=0}.$$

As the above ratio suggests, one often focuses on $\frac{Z(\lambda)}{Z(0)}$, which can be more numerically stable and also remains finite even if $Z(0)$ itself grows large.

3 Variational Bounds on the Partition Function

Rather than bounding a density $p(x)$ itself, one can work directly with variational bounds on $Z(\lambda)$. We begin by discussing the fundamental Jensen-Feynman bound and then outline a higher-order variant.

3.1 Jensen-Feynman Bound

A fundamental variational strategy employs Jensen's inequality for exponentials:

$$Z(\lambda) \;=\; \mathbb{E}_q\Big[\exp\big(\lambda^T x - S(x) - \ln q(x)\big)\Big] \;\;\geq\;\; \exp\Big(\mathbb{E}_q\big[\lambda^T x - S(x) - \ln q(x)\big]\Big).$$

Thus,

$$\ln Z(\lambda) \;\geq\; -F(\lambda),$$

where the *Helmholtz Free Energy* is defined as

$$F(\lambda) \;=\; -\,\mathbb{E}_q\big[\lambda^T x \;-\; S(x) \;-\; \ln q(x)\big].$$

This is known as the *Jensen-Feynman* or *Gibbs-Bogoliubov* bound [1, 7, 2]. The variational distribution q^* achieving the bound satisfies

$$q^*(x,\lambda) \;\propto\; \exp\big(\lambda^T x - S(x)\big).$$

3.2 Higher-Order Jensen-Feynman Bound

For more accurate approximations, one can extend the first-order exponential bound to a higher-order series without breaking the variational guarantee [7]. Define

$$\delta(x) \;=\; \lambda^T x \;-\; S(x) \;-\; \ln q(x) \;+\; F(\lambda).$$

Then

$$Z(\lambda) \;=\; e^{-F(\lambda)}\,\mathbb{E}_q\big[e^{\delta(x)}\big].$$

Consider the truncated exponential series remainder function

$$h_n(x) \;=\; \exp(x) \;-\; \sum_{k=0}^{2n-1} \frac{x^k}{k!},$$

which is convex. By Jensen's inequality,

$$\mathbb{E}_q\big[h_n(\delta(x))\big] \;\geq\; h_n\big(\mathbb{E}_q[\delta(x)]\big).$$

Hence,

$$\ln Z(\lambda) \;\geq\; -F(\lambda) \;+\; \ln\left\{1 \;+\; \sum_{k=2}^{2n-1} \frac{\mathbb{E}_q[\delta(x)^k]}{k!}\right\}.$$

The resulting bound can be sharper than the simpler first-order bound, at the expense of additional terms inside the expectation.

4 Saddle-Point Approximation

We now turn to approximation methods for $Z(\lambda)$. One of the most direct is the *saddle-point approximation*, also called the *semi-classical* approximation. A quadratic expansion around a local maximum of the exponent reduces the integral to a Gaussian form that can be evaluated analytically.

Let x^* be a point such that

$$\frac{\partial S(x^*)}{\partial x^*} = 0,$$

and let

$$H(x^*) \;=\; \frac{\partial^2 S(x^*)}{\partial x^* \, \partial x^{*T}}$$

be the Hessian matrix at x^*. Then

$$\begin{aligned}
Z(\lambda) &= \int dx \, \exp\bigl(-S(x) + \lambda^T x\bigr) \\
&\approx \int dx \, \exp\Bigl(-S(x^*) + \lambda^T x - \tfrac{1}{2}(x - x^*)^T H(x^*)(x - x^*)\Bigr) \\
&= \exp\Bigl(-S(x^*) + \lambda^T x^* + \tfrac{1}{2}\,\lambda^T H(x^*)^{-1}\lambda\Bigr) \int dx \, \exp\Bigl(-\tfrac{1}{2}(x - \mu)^T H(x^*)(x - \mu)\Bigr) \\
&= \exp\Bigl(-S(x^*) + \lambda^T x^* + \tfrac{1}{2}\,\lambda^T H(x^*)^{-1}\lambda\Bigr) \, \det\Bigl(\tfrac{H(x^*)}{2\pi}\Bigr)^{-\frac{1}{2}},
\end{aligned}$$

where $\mu = x^* + H(x^*)^{-1}\lambda$. Physically, one interprets x as having a "classical" component x^* (where the gradient vanishes), plus quantum or random fluctuations around it with covariance $\Sigma = H(x^*)^{-1}$.

A virtue of this approximation is that computing n-point functions becomes straightforward when combined with Wick's theorem, which we now recall.

4.1 Wick's Theorem

Wick's theorem expresses n-point functions of Gaussian distributions in terms of combinations of 2-point functions [9, 3, 8, 6]. Consider a zero-mean Gaussian action plus interactions,

$$S(x) = -\tfrac{1}{2}\, x^T \Sigma^{-1} x + V(x),$$

so that $p_0(x) \propto \exp\bigl(-\tfrac{1}{2}\, x^T \Sigma^{-1} x\bigr)$. Then

$$E_{p_0}[x_{a_1} \cdots x_{a_n}] \;=\; \frac{\partial^n}{\partial \lambda_{a_1} \cdots \partial \lambda_{a_n}} \exp\Bigl(\tfrac{1}{2}\, \lambda^T \Sigma \, \lambda\Bigr).$$

In its simplest form, Wick's theorem states:

Theorem 1 (Wick). *For the Gaussian distribution* $p_0(x) \propto \exp\bigl(-\tfrac{1}{2}\, x^T \Sigma^{-1} x\bigr)$,

$$E_{p_0}[x_{a_1} \cdots x_{a_n}] = \begin{cases} \displaystyle\sum_{\{(i_l, j_l)\}} \prod_{l=1}^{k} \Sigma_{i_l, j_l}, & n = 2k, \\ 0, & n = 2k+1, \end{cases}$$

where the notation $\sum_{\{(i_l,j_l)\}}$ denotes summation over all distinct ways of pairing the indices $\{a_1,\ldots,a_n\}$.

5 Perturbative Expansion

Another widely used approach is the *perturbative expansion*, which expands the exponential of the "interaction" term $V(x)$ in a power series. One typically writes

$$S(x) \;=\; S_0(x) \;+\; V(x),$$

where S_0 is a simpler *free-field* action giving rise to a tractable distribution

$$p_0(x) \;\propto\; \exp\bigl(-S_0(x)\bigr),$$

and $V(x)$ captures the remaining (small) coupling among the variables. Then

$$Z(\lambda) \;=\; \int dx\,\exp\bigl(-S_0(x)-V(x)+\lambda^T x\bigr) \;=\; Z_0\,\mathbb{E}_{p_0}\Bigl[e^{-V(x)+\lambda^T x}\Bigr],$$

where $Z_0=\int dx\,e^{-S_0(x)}$. One then expands $\exp\bigl(-V(x)\bigr)$ in a power series:

$$e^{-V(x)} \;=\; \sum_{k=0}^{\infty}\sum_{|a|=k}\frac{a_{k,a}}{k!}\prod_{i=1}^{d}x_i^{a_i},$$

where $\sum_{|a|=k}$ runs over all multi-indices a of length d whose components sum to k. Substituting into $Z(\lambda)$ and reordering yields

$$Z(\lambda) \;=\; Z_0\sum_{k=0}^{\infty}\sum_{|a|=k}\frac{a_{k,a}}{k!}\,E_{p_0}\Bigl[\prod_{i=1}^{d}x_i^{a_i}\;e^{\lambda^T x}\Bigr].$$

One can then derive an expansion for the n-point functions,

$$E_p[x_{a_1}\cdots x_{a_n}] \;=\; \frac{1}{Z}\sum_{k=0}^{\infty}\sum_{|a|=k}\frac{a_{k,a}}{k!}\,E_{p_0}\Bigl[x_{a_1}\cdots x_{a_n}\prod_{i=1}^{d}x_i^{a_i}\Bigr].$$

Crucially, each term involves expectations with respect to the simpler p_0, where variables are independent and thus easily computed.

5.1 Concrete Example: Boltzmann Machines

As a concrete example, consider a Boltzmann Machine over binary variables $x\in\{0,1\}^d$ with unnormalized distribution

$$p(x) \;\propto\; \exp\bigl(x^T W x + b^T x\bigr),$$

where W is symmetric ($W^T=W$) and has zero diagonal. Let

$$S(x) = x^T W x + b^T x \;=\; S_0(x) + V(x),$$

with $S_0(x) = b^T x$ and $V(x) = x^T W x$. Then $p_0(x) \propto \exp(b^T x)$ describes d independent Bernoulli variables with $p(x_i = 1) = \text{sigmoid}(b_i) = \mu_i$. One finds

$$Z_0 = \prod_{i=1}^{d} \big(1 + e^{b_i}\big), \qquad E_{p_0}\big[e^{\lambda^T x}\big] = \prod_{i=1}^{d} \big(1 + e^{b_i + \lambda_i}\big).$$

Up to first order in W, the partition function is

$$Z(\lambda) \;\approx\; \prod_{i=1}^{d} \big(1 + e^{b_i + \lambda_i}\big) \Big[1 + \tfrac{1}{2} \sum_{i,j} W_{ij}\, \mu_i\, \mu_j\Big],$$

where $\mu_i = \text{sigmoid}(\lambda_i)$. Differentiating then yields perturbative corrections to the mean,

$$E_p[x_i] \;\approx\; \mu_i + \mu_i\,(1 - \mu_i) \sum_j W_{ji}\, \mu_j,$$

and to the second moments,

$$E_p[x_i\, x_k] \;\approx\; \mu_i\, \mu_k + \mu_i\, \mu_k \Big[(1-\mu_k) \sum_j W_{kj}\, \mu_j + (1-\mu_i) \sum_j W_{ij}\, \mu_j\Big] + \mu_i\,(1-\mu_i)\, \mu_k\,(1-\mu_k)\, W_{ik}.$$

One thus obtains, for instance, the approximate covariance

$$\text{Cov}_p[x_i\,,\, x_k] \;\approx\; \mu_i\,(1 - \mu_i)\, \mu_k\,(1 - \mu_k)\, W_{ik}.$$

Further details and higher-order perturbations can be found in [5, 10], as well as applications to Restricted Boltzmann Machines [4].

6 Concluding Remarks

The partition function $Z(\lambda)$ and its logarithm lie at the heart of both statistical mechanics and probabilistic models. Exact calculations quickly become intractable in high dimensions or strongly coupled systems. Variational methods (including Jensen-Feynman and its higher-order extensions) provide systematic bounds, whereas the saddle-point approximation and perturbative expansions yield tractable analytic forms or series expansions. In combination with gradient-based optimization or sampling, these methods allow one to characterize complex distributions in practice.

References

[1] N. N. Bogoliubov. On representation of green-schwinger's function with the help of functional integrals. *Dokl. Akad. Nauk SSSR*, 99:225, 1954.

[2] R. P. Feynman. *Statistical Mechanics: A Set of Lectures.* Frontiers in Physics. Perseus Books, 1972.

[3] Gerald B Folland. *Quantum Field Theory: A tourist guide for mathematicians*. American Mathematical Soc., 2008.

[4] Marylou Gabrie, Eric W Tramel, and Florent Krzakala. Training restricted boltzmann machine via the thouless-anderson-palmer free energy. In *Advances in Neural Information Processing Systems*, pages 640–648, 2015.

[5] Antoine Georges and Jonathan S Yedidia. How to expand around mean-field theory using high-temperature expansions. *Journal of Physics A: Mathematical and General*, 24(9):2173, 1991.

[6] Krešimir Kumericki. Feynman diagrams for beginners, 2001. Notes for the exercises at the Adriatic School on Particle Physics and Physics Informatics, Split, Croatia.

[7] J.-C. Pain, F. Gilleron, and G. Faussurier. Jensen-feynman approach to the statistics of interacting electrons. *Phys. Rev. E*, 80(2):026703, 2009.

[8] Michael E Peskin and Daniel V Schroeder. *An introduction to quantum field theory*. Westview, 1995.

[9] Michael Polyak. Feynman diagrams for pedestrians and mathematicians. *Graphs and patterns in mathematics and theoretical physics*, 73:15–42, 2005.

[10] Vitor Sessak and Rémi Monasson. Small-correlation expansions for the inverse ising problem. *Journal of Physics A: Mathematical and Theoretical*, 42(5):55001, 2009.