Chapter 2
Statistical Models

2.1 Parameter Estimation in Statistical Physics

Linear regression is the statistical procedure most known by physicists. The Figure 2.1 shows a linear fit on semi-logarithmic paper to the data of the Problem 1.1. Less known is that linear regression is an example of what is called *a model belonging to the exponential family* A definition of this notion follows later on). The parameters of such models can be estimated by measuring quantities, called estimators. In statistical physics these estimators are called *extensive variables* because usually their average values are proportional to the size of the system. The averages of extensive variables are called *extensive parameters*. In contrast, the parameters of the model are called *intensive parameters*.

A well-known model of statistical physics is the *Ising model* on a chain or on a (finite part of a) square lattice. This model also belongs to the exponential family, by construction. It has 2 parameters, the inverse temperature β and the external magnetic field h. Its Hamiltonian, in the case of the chain, is given by

$$H(\sigma) = -J\sum_{n=1}^{N-1}\sigma_n\sigma_{n+1} - h\sum_{n=1}^{N}\sigma_n. \tag{2.1}$$

In this expression the σ_n are stochastic variables that can take on the values ± 1. This Hamiltonian is used to write down a probability distribution

$$p_{\beta,h}(\sigma) = \frac{1}{Z_N(\beta,h)}\exp(-\beta H(\sigma)). \tag{2.2}$$

The normalisation $Z_N(\beta,h)$ is called the *partition sum* and is given by

$$Z_N(\beta,h) = \sum_{\sigma}\exp(-\beta H(\sigma)). \tag{2.3}$$

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_2,

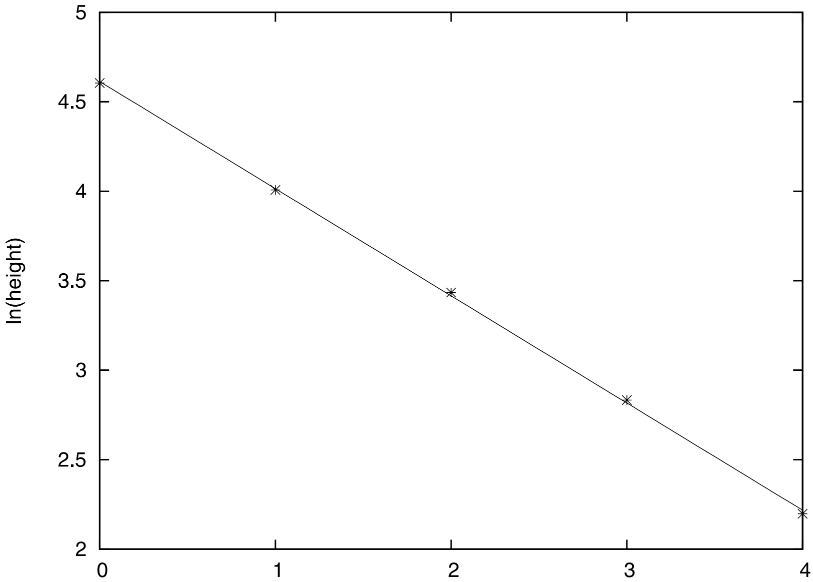

Fig. 2.1 Linear fit to the logarithm of the experimental data of the Problem 1.1

In this expression, the sum $\sum_\sigma$ extends over all possible values of each of the spin variables σ_n.

It is our intention to put the linear regression model and the Ising model on the same footing. To do so requires some work. For the linear regression model we have to introduce a probability distribution $p(y)$, which depends on 3 parameters, the slope a and intercept b of the fitted line, and the root mean square error σ of the fit. We also have to introduce a Hamiltonian H, like that of the Ising model. More precisely, we will introduce 3 extensive variables H_k, one for each of the intensive parameters. Indeed, the Hamiltonian of the Ising model, with its 2 parameters, is the sum of two pieces, one related to the interaction energy, the other due to the external field (here, the size N is not considered as a parameter, although that is a possibility).

Next, the two above mentioned models are considered in the context of statistical parameter estimation. The extensive variables are used as estimators, whose average value can be used to calculate the parameters of the model. In case of linear regression this is clear: the empirical values of the three extensive variables will be used to obtain the fitting parameters a, b, and σ. In case of the Ising model the interaction energy $H_1 = -J \sum_n \sigma_n \sigma_{n+1}$ and the total magnetisation $H_2 = \sum_n \sigma_n$ can be used to estimate the inverse temperature β and the strength of the external field h. The latter looks a little bit strange because experimental measurement of the interaction energy is usually more difficult than measuring temperature. We will come back to this point later on.

2.2 Definition of a Statistical Model

A model, in the present context, is a probability distribution p_θ, depending on a finite number of parameters $\theta^1, \cdots, \theta^m$, together with a set of extensive variables $H_1, \cdots, H_m$, which can be used to estimate the value of the model parameters. The expectation value is denoted $\langle\cdot\rangle_\theta$ and is defined by

$$\langle A\rangle_\theta = \int \mathrm{d}x\, p_\theta(x)A(x). \tag{2.4}$$

In this expression, $A(x)$ is an arbitrary quantity whose value depends on the events x. In the mathematical literature it is called a *stochastic variable*.

We always assume the existence of a function $\Phi(\theta)$ such that

$$\langle H_k\rangle_\theta = -\frac{\partial\Phi}{\partial\theta^k}, \qquad k = 1\cdots m. \tag{2.5}$$

Such a function $\Phi(\theta)$, called a *potential*, exists provided that

$$\frac{\partial}{\partial\theta^l}\langle H_k\rangle_\theta = \frac{\partial}{\partial\theta^k}\langle H_l\rangle_\theta \qquad \text{for all } k, l. \tag{2.6}$$

Its physical meaning is that of a *Massieu function* (this is a kind of *free energy*, well-known in thermodynamics [2]). The precise definition of the Massieu function follows in the next Section.

The estimators H_k are said to be *unbiased* if $\langle H_k\rangle_\theta = \theta^k$. For example, the kinetic energy of a particle in a classical[1] gas, multiplied with an appropriate constant, is an *unbiased estimator* of temperature T. However, most extensive variables of statistical physics are biased estimators.

A quantum model exists of a *density operator*[2] ρ_θ, depending on a finite number of parameters $\theta^1, \cdots, \theta^m$, together with a set of self-adjoint[3] operators $H_1, \cdots, H_m$, which can be used to estimate the value of the model parameters. These operators are the extensive variables of the quantum model. The expectation value of an arbitrary operator A is denoted $\langle\cdot\rangle_\theta$ and is defined by

$$\langle A\rangle_\theta = \operatorname{Tr}\rho_\theta A. \tag{2.7}$$

Characteristic for quantum models is that the operators $H_1, \cdots, H_m$, used to estimate the parameters θ_k, do not necessarily commute between themselves. See the example of Box 2.1.

---

[1] this means, neglecting quantum effects

[2] see Chapter 1

[3] this is, Hermitean (neglecting some mathematical details)

The simplest quantum-mechanical example concerns a magnetic spin described in terms of *Pauli matrices* $\sigma_k, k = 1, 2, 3$. These are defined by

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.8}$$

They satisfy the relations $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \mathbf{I}$, $\sigma_1\sigma_2 = -\sigma_2\sigma_1 = i\sigma_3$, and cyclic permutations of the latter.

An arbitrary density operator in the Hilbert space $\mathbf{C}^2$ can be written as

$$\rho_{\mathbf{r}} = \frac{1}{2}\left(\mathbf{I} - r^k\sigma_k\right), \tag{2.9}$$

with parameters r_1, r_2, r_3 satisfying $|\mathbf{r}|^2 = \sum_\alpha r_\alpha^2 \leq 1$. See the Figure 2.2. The Pauli matrices are the extensive variables of this model, which is known as the *Bloch representation* of the Pauli spin: $H_k = \sigma_k, k = 1, 2, 3$. A short calculation gives

$$\langle\sigma_k\rangle_{\mathbf{r}} = \frac{1}{2}\,\mathrm{Tr}\,\left(\mathbf{I} - r_l\sigma^l\right)\sigma_k = -r_k. \tag{2.10}$$

The potential $\Phi(\mathbf{r})$ is given by

$$\Phi(\mathbf{r}) = \frac{1}{2}|\mathbf{r}|^2 + \text{ constant}. \tag{2.11}$$

Box 2.1 Quantum spin example

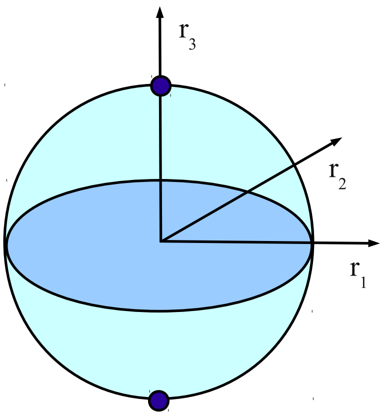

Fig. 2.2 The Bloch sphere. The top point corresponds with the pure state vector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, the bottom point with $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

2.3 The Exponential Family

Here we show that the linear regression model belongs to the (curved) exponential family. Let

$$\begin{array}{ll} \theta^1 = \frac{a}{\sigma^2} & H_1(y) = -\sum_n x_n y_n \\ \theta^2 = \frac{b}{\sigma^2} & H_2(y) = -\sum_n y_n \\ \theta^3 = \frac{1}{2\sigma^2} & H_3(y) = \sum_n y_n^2 \end{array}$$

Then one calculates

$$\begin{aligned} \int \mathrm{d}y_1 \cdots \int \mathrm{d}y_N \, e^{-\theta^k H_k(y)} &= \prod_{n=1}^{N} \int \mathrm{d}y_n \, e^{-\frac{1}{2\sigma^2}\left(y_n^2 - 2y_n(ax_n+b)\right)} \\ &= (2\pi\sigma^2)^{N/2} \prod_{n=1}^{N} \exp\left(\frac{1}{2\sigma^2}(ax_n+b)^2\right). \end{aligned}$$

Hence a properly normalised probability distribution $p_\theta(y)$ is obtained when the normalisation is fixed by

$$\begin{aligned} \Phi(\theta) &= \frac{N}{2} \ln 2\pi\sigma^2 + \frac{1}{2\sigma^2} \sum_{n=1}^{N} (ax_n + b)^2 \\ &= -\frac{N}{2} \ln \frac{\theta^3}{\pi} + \frac{1}{4\theta^3} \sum_{n=1}^{N} (\theta^1 x_n + \theta^2)^2. \end{aligned}$$

Three identities are now obtained by taking derivatives

$$\begin{aligned} \langle H_1 \rangle_\theta &= -\frac{\partial \Phi}{\partial \theta^1} = -a \sum_{n=1}^{N} x_n^2 - b \sum_{n=1}^{N} x_n \\ \langle H_2 \rangle_\theta &= -\frac{\partial \Phi}{\partial \theta^2} = -a \sum_{n=1}^{N} x_n - bN \\ \langle H_3 \rangle_\theta &= -\frac{\partial \Phi}{\partial \theta^3} = N\sigma^2 + \sum_{n=1}^{N} (ax_n + b)^2. \end{aligned}$$

These imply the famous fitting formulae

$$\begin{aligned} a &= \frac{1}{N} \frac{\langle x \rangle \langle H_2 \rangle - \langle H_1 \rangle}{\langle x^2 \rangle - \langle x \rangle^2} = \frac{1}{N} \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2} \\ b &= -\frac{1}{N} \langle H_2 \rangle - a \langle x \rangle = \langle y \rangle - a \langle x \rangle. \end{aligned}$$

Box 2.2 Linear regression

The notion of a statistical model has been explained in the previous section. Some statistical models are much easier to analyse than others. This is the case for models belonging to the exponential family. A good understanding of this property, shared by many models, is essential for the present book. It is the corner stone of the first part of the book and is generalised in the second part. The reason why it is so important is that it is the mathematical characterisation of the Boltzmann-Gibbs distribution as it is known in statistical physics – see Chapter 1. The second part of the book deals with generalisations of Boltzmann-Gibbs. These generalised probability distributions are characterised by the property that they belong to a generalised exponential family.

A statistical model is said to belong to the *exponential family* if its probability distribution can be written into the form

$$p_\theta(x) = c(x)\exp(-\Phi(\theta) - \theta^k H_k(x)). \tag{2.12}$$

Note the use of *Einstein's summation convention* (the summation over the index k is implicit). It is essential that the prefactor $c(x)$ and the extensive quantities $H_k(x)$ do not depend on the parameters θ while the normalisation function $\Phi(\theta)$ does not depend on the random variable x. Of course, the prefactor $c(x)$ may not be negative. It plays the role of a *prior probability*, although $\sum_x c(x)$ is not necessarily normalised to one. Therefore, it is a *weight*, rather than a probability distribution. In the physics literature the normalisation $\Phi(\theta)$ is usually written as a prefactor and is then called the *partition sum* $Z(\theta)$. The relation between these functions is $\Phi(\theta) = \ln Z(\theta)$.

It might be necessary to introduce new parameters to bring a statistical model into the canonical form (2.12). Indeed, consider for example the *Poisson distribution*

$$p(n) = \frac{\alpha^n}{n!}e^{-\alpha}, \qquad n = 0, 1, 2, \cdots, \tag{2.13}$$

with parameter $\alpha > 0$. Introduce a new parameter $\theta = -\ln\alpha$. Then the distribution can be written into the form (2.12), with

$$c(n) = \frac{1}{n!} \tag{2.14}$$

$$\Phi(\theta) = \exp(-\theta) \tag{2.15}$$

$$H(n) = n. \tag{2.16}$$

This shows that the Poisson distribution defines a 1-parameter model belonging to the exponential family.

It is clear that the Ising model belongs to the exponential family. In fact, from the definition of the canonical ensemble follows that all its models belong to the exponential family as well. To see that the linear regression model belongs to the exponential family requires some work. See the Box 2.2. An example of a probability distribution not belonging to the exponential family is the *Cauchy distribution*

$$p(x) = \frac{1}{\pi}\frac{a}{x^2 + a^2}, \tag{2.17}$$

where a is a positive parameter. This function is also called a *Lorentzian*.

A nice property of models belonging to the exponential family is that it is easy to calculate the averages $\langle H_k \rangle_\theta$ in terms of the parameters θ. Indeed, from the normalisation condition $1 = \int \mathrm{d}x\, p_\theta(x)$ follows

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta^k} \int \mathrm{d}x\, p_\theta(x) \\ &= \int \mathrm{d}x\, p_\theta(x) \left(-\frac{\partial \Phi}{\partial \theta^k} - H_k(x) \right). \end{aligned} \tag{2.18}$$

This implies

$$\frac{\partial \Phi}{\partial \theta^k} = -\langle H_k \rangle_\theta. \qquad k = 1, \cdots, m. \tag{2.19}$$

This is a well-known formula of statistical physics: extensive parameters are obtained by taking derivatives of the logarithm of the partition sum with respect to control variables. This expression also shows that $\Phi(\theta)$ is the potential function mentioned earlier in (2.6). In the next Chapter on thermodynamics it will be argued that the function $\Phi(\theta)$, as appearing in (2.12), is Massieu's function.

In the linear regression model the empirical values of the estimators H_k are used as a best guess for the average values $\langle H_k \rangle$. Next, (2.19) is used to obtain estimated values of the model parameters. One can wonder whether this is an optimal procedure. This kind of question is addressed in the *maximum likelihood* method. In this approach one poses the question what is the most likely value of the model parameters, given a sample of the total population.

In statistical physics, it is tradition to proceed in a different manner. Models (like the Ising model) can be so complex that most effort goes into the evaluation of $\Phi(\theta)$ as a function of θ. The result is then used to calculate averages $\langle H_k \rangle_\theta$ as a function of the parameters θ_k. This functional dependence is finally compared with experimental results, often in an indirect manner.

2.4 Curved Exponential Families

As said before, a probability distribution of the form

$$p_\zeta(x) = c(x) \exp(-\Phi(\theta) - \theta^k(\zeta) H_k(x)), \tag{2.20}$$

The probability density function of the normal distribution is

$$f_{a,\sigma}(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(u-a)^2\right). \tag{2.21}$$

It can be written as

$$f(u) = \exp\left(-\Phi(\theta) - \theta_1 H_1(u) - \theta_2 H_2(u)\right), \tag{2.22}$$

with

$$\begin{aligned} H_1(u) &= \frac{1}{2}u^2, \quad H_2(u) = u, \quad \theta_1 = \frac{1}{\sigma^2}, \quad \theta_2 = -\frac{a}{\sigma^2}, \\ \Phi(\theta) &= \frac{1}{2}\frac{\theta_2^2}{\theta_1} - \frac{1}{2}\ln(2\pi\theta_1). \end{aligned} \tag{2.23}$$

The curved coordinates are $\sigma = \zeta_1(\theta) = 1/\sqrt{\theta_1}$ and $a = \zeta_2(\theta) = \theta_2/\theta_1$. One verifies that

$$\begin{aligned} \frac{1}{2}\langle u^2\rangle &= \langle H_1\rangle = -\frac{\partial\Phi}{\partial\theta_1} = \frac{1}{2}a^2 + \frac{1}{2}\sigma^2 \\ \langle u\rangle &= \langle H_2\rangle = -\frac{\partial\Phi}{\partial\theta_2} = a. \end{aligned} \tag{2.24}$$

Box 2.3 The normal distribution

involving functions $\theta^k(\zeta)$, is still considered to belong to the exponential family because reparametrisation is allowed. If the transformation $\theta(\zeta)$ is nonlinear then the model with probability distribution p_ζ is said to belong to the *curved exponential family*. A well-known example of the curved exponential family is the *normal distribution*, also called the *Gauss distribution*. See the Box 2.3. Also the linear regression model is curved. See the Box 2.2.

Now, the normalisation condition implies

$$\begin{aligned} 0 &= \frac{\partial}{\partial\zeta^k}\int \mathrm{d}x\, p_\zeta(x) \\ &= \int \mathrm{d}x\, p_\zeta(x)\left(-\frac{\partial\Phi}{\partial\zeta^k} - \frac{\partial\theta^l}{\partial\zeta^k}H_l(x)\right). \end{aligned} \tag{2.25}$$

Hence,

$$\frac{\partial\Phi}{\partial\zeta^k} = -\frac{\partial\theta^l}{\partial\zeta^k}\langle H_l\rangle_\zeta, \tag{2.26}$$

which is not of the form (2.5). If the matrix with components $\dfrac{\partial\theta^l}{\partial\zeta^k}$ is degenerate then this set of equations is underdetermined and it is not possible to obtain the extensive parameters $\langle H_l \rangle_\theta$ as a function of the intensive ζ^k just by solving this set of equations.

2.5 Example: The Ising Model in d=1

It is easy to calculate the partition sum of the Ising chain with $h = 0$. Introduce new stochastic variables $\tau_n = \sigma_n \sigma_{n+1}$. Then the partition sum reads

$$\begin{aligned} Z_N(\beta, h=0) &= 2 \sum_\tau \exp(\beta J \sum_{n=1}^{N-1} \tau_n) \\ &= 2 \prod_{n=1}^{N-1} \sum_{\tau_n = \pm 1} \exp(\beta J \tau_n) \\ &= 2\, (2 \cosh(\beta J))^{N-1} . \end{aligned} \tag{2.27}$$

Box 2.4 Partition sum of the Ising chain with $h = 0$

The probability distribution $p_{\beta,h}$ of the one-dimensional *Ising model* is given by (2.2) in terms of the stochastic variables σ_n, $n = 1..N$ (called spin variables). The actual probability space is the set Γ of all configurations. Each configuration assigns the value ± 1 to each of the spin variables. See Figure 2.3 for an example of a configuration with $N = 6$. The normalisation condition is

$$\sum_{x \in \Gamma} p_{\beta,h}(x) = 1. \tag{2.28}$$

The number of configurations is 2^N and increases exponentially with increasing value of N. Hence, one can expect that individual probabilities $p_{\beta,h}(x)$ are very small numbers.

The Ising model, as defined by (2.2), is called the model with *open boundary conditions*, or still, the Ising chain. Its partition sum $Z_N(\beta, h)$ can be

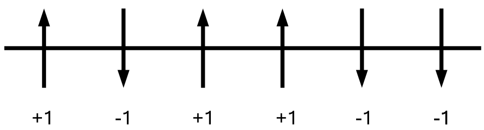

Fig. 2.3 Configuration of spins

Here, we calculate the partition sum of the $d = 1$-Ising model with periodic boundary conditions, using the transfer matrix method.

Write the partition sum as

$$Z_N(\beta, h) = \sum_\sigma \prod_{n=1}^N \exp\left(\beta J \sigma_n \sigma_{n+1} + \frac{1}{2}\beta h(\sigma_n + \sigma_{n+1})\right), \quad (2.29)$$

where σ_{N+1} is identified with σ_1. Next, notice that this is still identical with

$$Z_N(\beta, h) = \operatorname{Tr} T^N \qquad \text{with} \qquad T = \begin{pmatrix} e^{\beta J + \beta h} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta J - \beta h} \end{pmatrix}. \quad (2.30)$$

The two eigenvalues of this matrix are

$$\lambda_\pm = e^{\beta J} \cosh(\beta h) \pm \sqrt{e^{-2\beta J} + e^{2\beta J} \sinh^2(\beta h)}. \quad (2.31)$$

The result is

$$Z_N(\beta, h) = \lambda_+^N + \lambda_-^N. \quad (2.32)$$

In the limit of large N, the so-called *thermodynamic limit,* only the largest eigenvalue is important. The result then simplifies to

$$\ln Z_N(\beta, h) = N \ln \lambda_+ + \cdots. \quad (2.33)$$

Expansion for small values of h then gives

$$\ln Z_N(\beta, h) = N \ln \cosh(\beta J) + N \ln\left(2 + e^{2\beta J}(\beta h)^2 + \cdots\right). \quad (2.34)$$

Box 2.5 Partition sum of the $d = 1$-Ising model with periodic boundary conditions

calculated in closed form when $h = 0$. See the Box 2.4. A slight modification of the model allows to calculate $Z_N(\beta, h)$ in closed form for all values of the parameters. Adding one term to the Hamiltonian, (2.1) becomes

$$H(\sigma) = -J \sum_{n=1}^{N-1} \sigma_n \sigma_{n+1} - J \sigma_N \sigma_1 - h \sum_{n=1}^N \sigma_n. \quad (2.35)$$

The probability distributions, defined by (2.2) with the modified Hamiltonian, are called the Ising model with *periodic boundary conditions,* or, the Ising model on a circle. Its partition sum $Z_N(\beta, h)$ can be calculated in closed form by the *transfer matrix method.* See the Box 2.5.

By taking derivatives of $\ln Z_N(\beta, h)$ with respect to β, respectively h, one obtains the averaged quantities $-\langle H \rangle$ and $\beta \langle M \rangle$, with $M(\sigma) = \sum_{n=1}^N \sigma_n$ the total magnetisation. The resulting expressions are rather complicated. A series expansion for small values of $(\beta h)^2$, together with the approximation that the system size N is large, gives

$$\langle H\rangle = -NJ\tanh\beta J - N(\beta^{-1}+J)e^{2\beta J}(\beta h)^2 + \cdots \tag{2.36}$$

$$\langle M\rangle = Ne^{2\beta J}(\beta h) + \cdots . \tag{2.37}$$

Some observations can be made here.

- Both $\langle H\rangle$ and $\langle M\rangle$ are linear in the size of the system N. For this reason, H and M are called extensive variables.
- The average energy $\langle H\rangle$ is a decreasing function of β at constant h. Hence it is an increasing function of the temperature T (the relation between both is $\beta = 1/k_B T$, where k_B is *Boltzmann's constant*; it converts degrees Kelvin into energies, measured in Joule). $\langle H\rangle$ is also called the internal energy. Its derivative with respect to temperature is the heat capacity. A system with negative heat capacity is unstable. Its temperature drops while heating the system. Examples of such behaviour are known in astronomy.
- The average magnetisation $\langle M\rangle$ vanishes when $h = 0$. The derivative of $\langle M\rangle$ with respect to h is called the *static magnetic susceptibility*.
- The Massieu function $\ln Z_N(\beta, h)$ is a real analytic function of $\beta > 0$ and h. The occurrence of a singularity in the function

$$\phi(\beta, h) = \lim_{N\to\infty} \frac{1}{N}\ln Z_N(\beta, h) \tag{2.38}$$

  would be associated with a phase transition. These are discussed in a later chapter. The Ising model on a square lattice exhibits such a phase transition.

2.6 The Density of States

Let be given a continuous distribution $p_\theta(x)$ of the form (2.12), belonging to the exponential family. It quite often happens that one is only interested in calculating average values of quantities which depend only on the average values E_k of the Hamiltonians $H_k(x)$. In such a case it is advantageous to introduce the *density of states*

$$\omega(E) = \int \mathrm{d}x\, c(x) \prod_k \delta(H_k(x) - E_k). \tag{2.39}$$

Indeed, the average of a quantity A, which depends only on the $H_k(x)$, is then obtained by

$$\begin{aligned}\langle A\rangle_\theta &= \int \mathrm{d}x\, p_\theta(x) A(H(x))\\ &= \int \mathrm{d}E\, \omega(E) e^{-\Phi(\theta)-\theta^k E_k} A(E).\end{aligned} \tag{2.40}$$

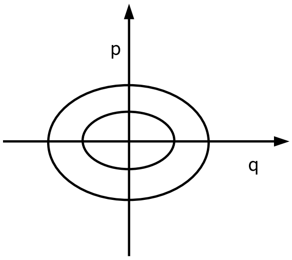

Fig. 2.4 Constant energy lines in the phase space of a harmonic oscillator

Take for example the classical *harmonic oscillator*. Its Hamiltonian is

$$H(x) \equiv H(p,q) = \frac{1}{2m}p^2 + \frac{1}{2}m\omega_0^2 q^2. \tag{2.41}$$

The states of equal energy E in the two-dimensional *phase space* Γ form an ellipse — see the Figure 2.4. Intuitively, one would then guess that the density of states increases with the energy E. However, this is wrong. A careful calculation gives

$$\begin{aligned}\omega(E) &= \frac{1}{2h}\int \mathrm{d}q \int \mathrm{d}p\, \delta(\frac{1}{2m}p^2 + \frac{1}{2}m\omega_0^2 q^2 - E) \\ &= \frac{1}{h\omega_0}\int \mathrm{d}u \int \mathrm{d}v\, \delta(u^2+v^2-E) \\ &= 2\pi \frac{1}{h\omega_0}\int_0^\infty r\mathrm{d}r\, \delta(r^2-E) \\ &= \pi \frac{1}{h\omega_0}\int_0^\infty \mathrm{d}s\, \delta(s-E) \\ &= \frac{\pi}{h\omega_0}.\end{aligned} \tag{2.42}$$

This shows that the density of states of the classical harmonic oscillator is a constant, independent of the energy $E \geq 0$.

2.7 The Quantum Exponential Family

A quantum model is said to belong to the *quantum exponential family* if its density operator can be written into the form

$$\rho_\theta = \frac{1}{Z(\theta)}\exp(-\theta^k H_k) = \exp(-\Phi(\theta) - \theta^k H_k), \tag{2.43}$$

with self-adjoint operators H_k and with normalisation

Consider the Bloch representation of the Pauli spin — see the Box 2.1. Let us show that one can write

$$\rho_{\mathbf{r}} = \frac{1}{Z(\mathbf{r})} e^{-\sum_k \theta^k \sigma_k} \tag{2.45}$$

with $\theta^k \equiv \theta^k(\mathbf{r})$ and

$$Z(\mathbf{r}) = \operatorname{Tr} e^{-\theta^k \sigma_k} = 2 \cosh |\theta|. \tag{2.46}$$

In particular, this model belongs to the curved quantum exponential family.

In order to prove (2.45, 2.46), chose a basis in which $\rho_{\mathbf{r}} = \frac{1}{2}\left(\mathbf{I} - r^k \sigma_k\right)$ is diagonal. This is equivalent with assuming $r_1 = r_2 = 0$. In that case it is clear that $\rho_{\mathbf{r}} = \frac{1}{Z} \exp(-\theta_3 \sigma_3)$ with $\tanh \theta_3 = r_3$ and $Z = 2 \cosh \theta_3$. By going back to the original basis $\theta_3 \sigma_3$ transforms into a matrix of the form $\theta^k \sigma_k$. The trace of a matrix does not depend on the choice of basis. Hence,

$$Z = 2 \cosh \theta_3 = 2/\sqrt{1 - \tanh^2 \theta_3} = 2/\sqrt{1 - r_3^2}. \tag{2.47}$$

Under a change of basis the length of the Bloch vector does not change. Hence, (2.46) follows.

Box 2.6 The Bloch sphere belongs to the quantum exponential family

$$Z(\theta) = \operatorname{Tr} \exp(-\theta^k H_k), \qquad \Phi(\theta) = \ln Z(\theta). \tag{2.44}$$

Note that in the example of the *ideal gas* of bosons or of fermions, discussed in Chapter 1, the relevant observables H_N, N, and n_j, two-by-two commute. However, in general, the operators H_k do not mutually commute, except of course in the one-parameter case. As a consequence, several properties, which hold classically or when the H_k mutually commute, cannot be easily generalised. But thanks to the property called ‘cyclic permutation under the trace’ the basic relations (2.19), with $\Phi(\theta) = \ln Z(\theta)$, still hold. Indeed, one has

$$\begin{aligned}
\frac{\partial \Phi}{\partial \theta^k} &= \frac{1}{Z(\theta)} \frac{\partial}{\partial \theta^k} \operatorname{Tr} e^{-\theta^l H_l} \\
&= \frac{1}{Z(\theta)} \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\partial}{\partial \theta^k} \operatorname{Tr} (-\theta^l H_l)^n \\
&= \frac{1}{Z(\theta)} \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{j=0}^{n-1} \operatorname{Tr} (-\theta^l H_l)^j (-H_k) (-\theta^l H_l)^{n-1-j} \\
&= \frac{1}{Z(\theta)} \sum_{n=0}^{\infty} \frac{1}{n!} n \operatorname{Tr} (-\theta^l H_l)^{n-1} (-H_k) \\
&= -\frac{1}{Z(\theta)} \operatorname{Tr} e^{-\theta^l H_l} H_k
\end{aligned}$$

$$= -\langle H_k \rangle. \tag{2.48}$$

A reparametrisation of the parameter space is allowed. In this case the expressions become

$$\rho_\zeta = \frac{1}{Z(\theta)} \exp(-\theta^k H_k), \tag{2.49}$$

with self-adjoint operators H_k, with $\theta^k \equiv \theta^k(\zeta)$, and with normalisation

$$Z(\theta) = \operatorname{Tr} \exp(-\theta^k H_k). \tag{2.50}$$

The model is said to be *curved* if the transformation $\theta(\zeta)$ is non-linear. For an example, see the Box 2.1.

Note that one cannot add a prior weight (the $c(x)$ in (2.12) and (2.20)) in the definition of the quantum exponential family because it would spoil the property of cyclic permutation under the trace, essential in the calculation of (2.48).

Problems

2.1. Correlations in the one-dimensional Ising model
Calculate $\langle \sigma_1 \sigma_n \rangle$ for the one-dimensional Ising model with periodic boundary conditions, in absence of an external field (this means $h = 0$). A quantity like $\langle \sigma_1 \sigma_n \rangle$ is an example of a two-point *correlation function* [4]

2.2. The Gamma distribution
The density function of the *Gamma distribution* is given by

$$p(x) = \frac{x^{k-1} e^{-x/b}}{b^k \Gamma(k)}. \tag{2.51}$$

It coincides with the exponential distribution when $k = 1$. Show that the Gamma distribution belongs to the exponential family with two parameters $\theta_1 = 1 - k$ and $\theta_2 = 1/b$.

2.3. Example of the quantum exponential family
Show that the density matrix

$$\rho_\theta = \begin{pmatrix} \theta & 0 \\ 0 & 1-\theta \end{pmatrix}, \qquad 0 < \theta < 1, \tag{2.52}$$

belongs to the curved quantum exponential family.

[4] In fact, $\langle \sigma_1 \sigma_n \rangle - \langle \sigma_1 \rangle \langle \sigma_n \rangle$ is a correlation function. But $\langle \sigma_1 \rangle = \langle \sigma_n \rangle = 0$ holds when $h = 0$.

2.4. Density profile of the earth — See [1].
The mass density of the earth $\rho(r)$ decreases as a function of the distance r to the centre of the earth. Assume a perfect sphere. The radius R, the mass M, and the moment of inertia J are experimentally known. Experimental numbers are $R \simeq 6.36 \times 10^6$ m, $M \simeq 6.0 \times 10^{24}$ kg, and $J \simeq 4.0 \times 10^{37}$ kg m^3. Predict the density at the centre of the earth.

Hints Discretise the density $\rho(r)$ by dividing the sphere into N shells of equal volume $V = 4\pi R^3/3N$. This introduces N variables $\rho_1, \cdots \rho_N$, satisfying

$$H_1(\rho) \equiv \frac{1}{N}\sum_{n=1}^{N} \rho_n = \frac{3M}{4\pi R^3} \tag{2.53}$$

$$H_2(\rho) \equiv \frac{1}{N^{5/3}}\sum_{n=1}^{N} \rho_n (n^{5/3} - (n-1)^{5/3}) = 15J/4\pi R^5. \tag{2.54}$$

Next assume a probability distribution $p_\theta(\rho)$ belonging to the two-parameter exponential family with Hamiltonians $H_1(\rho)$ and $H_2(\rho)$. Fix the parameters θ^1, θ^2 so that $\langle H_1 \rangle = 5568$ kg/m^3 and $\langle H_2 \rangle = 4588$ kg/m^3. Finally, integrate $p_\theta(\rho)$ over all ρ_j but that of the most inner shell to obtain the probability distribution of the latter.

Result The predicted density at the center of the earth is 17140 kg/m^3 (do not take this result too serious!).

2.5. Binomial distribution
Fix an integer $n \geq 2$. The binomial distribution is given by

$$p_a(m) = \binom{n}{m} a^m (1-a)^{n-m}, \quad m = 0, 1, \cdots, n, 0 \leq a \leq 1. \tag{2.55}$$

Show that as a function of the parameter a it belongs to the curved exponential family.

2.6. Weibull distribution
Fix positive parameters k and λ. The *Weibull distribution* is defined on the positive axis by

$$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}. \tag{2.56}$$

For $k = 1$ this is the *exponential distribution*, for $k = 2$ this is the *Raleigh distribution*. Show that as a function of λ it belongs to the curved exponential family.

Notes

The present chapter is inspired by a paper [3] on parameter estimation in the context of generalised thermostatistics. But most of the contents of this Chapter is fairly standard.

The problem of estimating the state of a quantum system has started only recently — see for instance [4] and the references quoted there.

Objectives

- Explain the notion of an estimator.
- Know how to describe an n-parameter model of statistical physics, both classically and quantum mechanically. Give an example of each of these.
- Give the definition of the exponential family, with and without curvature. Give examples.
- Show that the linear regression model belongs to the exponential family.
- Calculate the average value of an estimator of a model belonging to the exponential family, by taking a derivative of the Massieu function.
- Solve the 1-dimensional Ising model both with open and with periodic boundary conditions.
- Give the definition of the quantum exponential family.

References

1. Bevensee, R.: Maximum entropy solutions to scientific problems. Prentice Hall (1993) 35
2. Callen, H.: Thermodynamics and an Introduction to Thermostatistics, 2nd edn. Wiley, New York (1985) v, 23, 37
3. Naudts, J.: Parameter estimation in nonextensive thermostatistics. Physica A **365**, 42–49 (2006) 36, 162
4. Petz, D., Hangos, K., Magyar, A.: Point estimation of states of finite quantum systems. J. Phys. A **40**, 7955–7970 (2007) 36

http://www.springer.com/978-0-85729-354-1

Generalised Thermostatistics
Naudts, J.
2011, XI, 201 p., Hardcover
ISBN: 978-0-85729-354-1