FISHER INFORMATION AND STATISTICAL MODELS

1. Definition

Consider a probability distribution on a measurable space $\mathcal{X}$ modelled by a probability density function $P_\theta(X),\ X \in \mathcal{X}$ dependent on a number of parameters globally identified by $\theta \in \mathcal{M}$, where $\mathcal{M}$ is an arbitrary n-dimensional manifold. We define the Fisher Information tensor of the parametric family P at a point θ as:

$$F_{ij}(\theta) = -\int_{\mathcal{X}} P_\theta(X) \frac{\partial}{\partial \theta_i} \log P_\theta(X) \times \frac{\partial}{\partial \theta_j} \log P_\theta(X) d^n X$$

Note that F is a symmetric and positive semidefinite tensor at any given point, so it defines a Riemannian metric[1] on the manifold of parameters.

1.1. FI and Information.

Defining information as normal, $I_\theta(X) = -\log P_\theta(X)$, we can rewrite the above as

$$F_{ij}(\theta) = \int_{\mathcal{X}} P_\theta(X) \frac{\partial I_\theta(X)}{\partial \theta_i} \frac{\partial I_\theta(X)}{\partial \theta_j} d^n X$$

This can also be written, under simple regularity conditions, as

$$F_{ij}(\theta) = \int_{\mathcal{X}} P_\theta(X) \frac{\partial^2 I_\theta(X)}{\partial \theta_i \partial \theta_j} d^n X.$$

2. FI and physical quantities

Consider a physical system described by the canonical ensemble, i.e. where by definition

$$P_\theta(X) = \frac{e^{-H_\theta(X)}}{Z_\theta}, \quad H_\theta(X) = \sum_{k=1}^{n} \theta_k f_k(X)$$

(I have incorporated the kT factor used in physics into the θs, so we can assume $kT = 1$). The Shannon entropy takes the form

$$S_\theta = \langle H_\theta(X) \rangle + \log Z_\theta$$

Physically, the average of H is the internal energy, and $F = -\log Z$ is the Helmholtz free energy. From the definition of the latter, Shannon entropy equals the canonical, physical entropy.

[1]This should require positive definiteness. I'm not sure how the two things are compatible.

2.1. **Generalised susceptibilities.** Simply applying this particular form of P to the definition of Fisher Information, we find a direct connection between the covariance of the physical quantities f and the FI:

$$F_{ij}(\theta) = \mathrm{Cov}[f_i(X), f_j(X)].$$

This means that the FIM characterises the variances and correlations of the fs intended as stochastic variables which are functions of the state X of the system.

2.2. **Specific heat.** We note that changing the temperature in the traditional canonical ensemble corresponds to scaling the Hamiltonian by a factor $\beta = 1/kT$. In the formulation above, this is equivalent to scaling all the θ_is by β. Given a point on $\mathcal{M}$, the direction $\partial/\partial\beta$ can be expressed as

$$\frac{\partial}{\partial\beta} = \frac{1}{n}\sum_{k=1}^{n}\frac{\partial}{\partial\theta_k}$$

which is just a linear combination. With a change of basis, which includes this vector as a base vector together with other $n-1$ linearly independent ones, we can arrange for the specific heat to be one of the entries of the Fisher Information in matrix form. Analogous considerations can be made for the magnetic field and magnetic susceptibility, and so on.

3. Critical points

There are physical situations, for example due to nonlinear effects in the model defined by a certain Hamiltonian, where a small change of parameters can induce a sharp change in the probability distributions. For example, a volume of water vapour at 100 degrees suddenly becomes a liquid if the temperature is decreased by an amount no matter how small. The distribution of states for the molecules in a gas is qualitatively different from the one which describes the states of the same molecules behaving as a liquid.

3.1. **FI and the KL divergence.** A measure of the "distance" (although it does not satisfy the mathematical definition of a distance) between two probability distributions can be defined by the Kullback-Leibler divergence:

$$D_{KL}(P_{\theta_1}|P_{\theta_0}) = \int_{\mathcal{X}} P_{\theta_1}(X)\log\frac{P_{\theta_1}(X)}{P_{\theta_0}(X)}dX$$

if we now take $\theta_1 = \theta_0 + \delta\theta$, in the limit of small $\delta\theta$, it can be shown that

$$D_{KL}(P_{\theta_0+\delta\theta}|P_{\theta_0}) \approx \frac{1}{2}\sum_{i,j} F_{ij}(\theta_0)\delta\theta_i\delta\theta_j.$$

In other words, the FI is the second order approximation of the KL divergence.

If at a point θ_0 an increase (or decrease) of the k-th parameter is known to immediately lead to a phase transition, such as what happens when θ_0 is a critical point, we expect the KL divergence to be non-zero even for an arbitrarily small value of $\delta\theta$. This implies, by looking at the equation above, a divergence of at least one entry of the Fisher tensor.