Parameter priors for Ising models

research notes

Y. Roudi
<yasser.roudi@ntnu.no>

P.G.L. Porta Mana
<piero.mana@ntnu.no>

25 June 2018; updated 26 June 2018

Study of uniform priors in parameter space and in constraint space for Ising models

'Flat priors do not exist'
(anonymous)

1 A two-unit model with sufficient statistics

Consider a population of two binary units $\boldsymbol{s} \coloneqq (s_1, s_2)$ with values in $\{0, 1\}$. One observation of this population can thus give four results: $\boldsymbol{s} \in \{00, 01, 10, 11\}$.

Assume that we have N observations $(\boldsymbol{s}^{(1)}, \dots, \boldsymbol{s}^{(N)})$ of this or other populations prepared in similar conditions, so that knowledge of these observations is relevant for our forecast of a new observation $\boldsymbol{s}$, again in similar conditions. Also assume that only the number, the mean, and the second moments of these past observations are relevant to forecast the new one; that is,

$$N, \qquad \tfrac{1}{N}\bigl(\boldsymbol{s}^{(1)} + \dots + \boldsymbol{s}^{(N)}\bigr) \eqqcolon \boldsymbol{a}, \qquad \tfrac{1}{N}\bigl(s_1^{(1)} s_2^{(1)} + \dots + s_1^{(N)} s_2^{(N)}\bigr) \eqqcolon c \quad (1)$$

are sufficient statistics; note that the second sum contains the first as its diagonal. These assumptions are collectively denoted I. Then the Koopman-Pitman theorem says that our probabilistic forecasts must assume this general form:

$$\mathrm{p}(\boldsymbol{s}^{(1)}, \dots, \boldsymbol{s}^{(n)} \mid I) = \int \Biggl[\prod_{i=1}^{n} g(\boldsymbol{s}^{(i)})\, \frac{\exp\bigl(\mu_1 s_1^{(i)} + \mu_2 s_2^{(i)} + \lambda s_1^{(i)} s_2^{(i)}\bigr)}{Z(\mu_1, \mu_2, \lambda)}\Biggr] \mathrm{p}(\mu_1, \mu_2, \lambda \mid I)\, \mathrm{d}\mu_1\, \mathrm{d}\mu_2\, \mathrm{d}\lambda,$$

$$\text{with} \quad Z(\mu_1, \mu_2, \lambda) \coloneqq 1 + \exp(\mu_1) + \exp(\mu_2) + \exp(\mu_1 + \mu_2 + \lambda). \quad (2)$$

Let's denote $\boldsymbol{\theta} \coloneqq (\mu_1, \mu_2, \lambda) \in \mathbf{R}^3$.

The distribution $g(\boldsymbol{s})$ and the density $\mathrm{p}(\mu_1, \mu_2, \lambda \,|\, I)$ in the formula above are not determined by the theorem: they need to be determined by additional assumptions. The distribution g is often determined by symmetry or combinatorial properties of the system. In the present study we assume it to be unity: $g(\boldsymbol{s}) = 1$. The density $\mathrm{p}(\boldsymbol{\theta} \,|\, I)$ is called *prior parameter density*.

Geometrically, the formula above says that our probability distribution is the convex combination of probability distributions from a three-dimensional family parameterized by $\boldsymbol{\theta}$, with weights $\mathrm{p}(\boldsymbol{\theta} \,|\, I)\,\mathrm{d}\boldsymbol{\theta}$. The parameters $\boldsymbol{\theta}$ are just coordinates of this three-dimensional manifold, and we can choose other coordinates $\boldsymbol{t}$, the weights then being given by $\mathrm{p}(\boldsymbol{t} \,|\, I)\,\mathrm{d}\boldsymbol{t}$, with the densities for $\boldsymbol{\theta}$ and for $\boldsymbol{t}$ related by a Jacobian determinant:

$$\mathrm{p}(\boldsymbol{\theta} \,|\, I) = \mathrm{p}[\boldsymbol{t}(\boldsymbol{\theta}) \,|\, I] \, \det\!\left(\frac{\partial \boldsymbol{t}}{\partial \boldsymbol{\theta}}\right). \tag{3}$$

Using Bayes's theorem with the probabilities (2) we find our forecast for a new observation $\boldsymbol{s}$ conditional on observations $\big(\boldsymbol{s}^{(1)}, \dots, \boldsymbol{s}^{(N)}\big)$:

$$\begin{aligned}\mathrm{p}(\boldsymbol{s} \,|\, \boldsymbol{s}^{(1)}, \dots, \boldsymbol{s}^{(N)}, I) ={}& \\ &\int \frac{\exp(\mu_1 s_1 + \mu_2 s_2 + \lambda s_1 s_2)}{Z(\boldsymbol{\theta})}\, \mathrm{p}(\boldsymbol{\theta} \,|\, \boldsymbol{s}^{(1)}, \dots, \boldsymbol{s}^{(N)} I)\,\mathrm{d}\boldsymbol{\theta}\end{aligned} \tag{4a}$$

with

$$\begin{aligned}\mathrm{p}(\boldsymbol{\theta} \,|\, \boldsymbol{s}^{(1)}, \dots, \boldsymbol{s}^{(N)} I) \propto{}& \left[\prod_{i=1}^{N} \frac{\exp\big(\mu_1 s_1^{(i)} + \mu_2 s_2^{(i)} + \lambda s_1^{(i)} s_2^{(i)}\big)}{Z(\boldsymbol{\theta})}\right] \mathrm{p}(\boldsymbol{\theta} \,|\, I) = \\ & \exp\{N\,[\mu_1 a_1 + \mu_2 a_2 + \lambda c - \ln Z(\boldsymbol{\theta})]\}\, \mathrm{p}(\boldsymbol{\theta} \,|\, I).\end{aligned} \tag{4b}$$

The density $\mathrm{p}(\boldsymbol{\theta} \,|\, \boldsymbol{s}^{(1)}, \dots, \boldsymbol{s}^{(N)} I)$ is called *posterior parameter density*.

The last expression shows that the N observations affect our forecast only through the averages $\boldsymbol{a}$ and c, eq. (1), as we assumed.

The proportionality relation of the last formula reminds us that we must perform an integral over $\boldsymbol{\theta}$ to calculate the posterior parameter density. We must also perform an integral over $\boldsymbol{\theta}$ to calculate the conditional probability for $\boldsymbol{s}$. These integrals are difficult when we consider populations with many units. When the number N of known observations is large, the posterior parameter density is often approximated by a Dirac delta centred on the maximum of the posterior,

$$\theta_{\mathrm{m}} := \arg\sup_{\theta} \{N\,[\mu_1 a_1 + \mu_2 a_2 + \lambda c - \ln Z(\theta)] + \ln \mathrm{p}(\theta \mid I)\}. \tag{5}$$

The probability for s then equals the exponential calculated at θ_{m}. If the prior parameter density $\mathrm{p}(\theta \mid I)$ is constant or very broad, it can be dropped in the calculation of the maximum, as an approximation.

2 Other prior parameter densities

The literature often assumes a prior parameter density $\mathrm{p}(\theta \mid I)$ that is constant in θ. This is an 'improper', non-normalizable prior; we are really considering a sequence of normalizable priors of increasing width – for example, normal distributions with increasing variance – and the resulting limit if it exists.

As noted before, the parameters θ are just coordinates in a manifold of distributions for s. A constant density in these coordinates corresponds to an non-constant density in other coordinates. But are these coordinates 'special' in any way, to consider a constant density in them? Are there other coordinates in which it makes more sense to consider a constant density? How does a different choice affect our inference about s?

To consider other coordinates it is useful to give the predictive formula (2) a particular interpretation.

First of all it must be noted that each choice of parameters θ gives a distribution

$$\mathrm{p}(s \mid \theta, I) = \frac{\exp\big(\mu_1 s_1^{(i)} + \mu_2 s_2^{(i)} + \lambda s_1^{(i)} s_2^{(i)}\big)}{Z(\theta)} \tag{6}$$

with different first and second moments

$$\begin{aligned} \mathrm{E}(s \mid \theta, I) &:= \sum_s s\, \mathrm{p}(s \mid \theta, I), \\ \mathrm{E}(s_1 s_2 \mid \theta, I) &:= \sum_s s_1 s_2\, \mathrm{p}(s \mid \theta, I). \end{aligned} \tag{7}$$

In other words the set $\theta \in \mathbf{R}^3$ is in one-one correspondence with the possible values of the three expectations above. We can therefore introduce coordinates $\boldsymbol{t} \coloneqq (m_1, m_2, l)$ that identify the distributions above via the equations

$$
\begin{aligned}
m_1 = \mathrm{E}(s_1 | \,\boldsymbol{\theta}, I) &\equiv \frac{\exp(\mu_1) + \exp(\mu_1 + \mu_2 + \lambda)}{Z(\mu_1, \mu_2, \lambda)},\\
m_2 = \mathrm{E}(s_2 | \,\boldsymbol{\theta}, I) &\equiv \frac{\exp(\mu_2) + \exp(\mu_1 + \mu_2 + \lambda)}{Z(\mu_1, \mu_2, \lambda)},\\
l = \mathrm{E}(s_1 s_2 | \,\boldsymbol{\theta}, I) &\equiv \frac{\exp(\mu_1 + \mu_2 + \lambda)}{Z(\mu_1, \mu_2, \lambda)},
\end{aligned}
\tag{8}
$$

which are coordinate transformations with the inverse

$$
\begin{gathered}
\mu_1 = \ln \frac{m_1 - l}{1 + l - m_1 - m_2}, \qquad \mu_2 = \ln \frac{m_2 - l}{1 + l - m_1 - m_2},\\
\lambda = \ln \frac{(1 + l - m_1 - m_2)\, l}{(m_1 - l)\,(m_2 - l)}.
\end{gathered}
\tag{9}
$$

In terms of the coordinates $\boldsymbol{t}$ the family of probability distributions for $\boldsymbol{s}$ has the form

$$
\begin{aligned}
&\mathrm{p}(\boldsymbol{s} | \,\boldsymbol{t}, I) = (1 + l - m_1 - m_2) \times\\
&\quad \left(\frac{m_1 - l}{1 + l - m_1 - m_2}\right)^{s_1} \left(\frac{m_2 - l}{1 + l - m_1 - m_2}\right)^{s_2} \left[\frac{(1 + l - m_1 - m_2)\, l}{(m_1 - l)\,(m_2 - l)}\right]^{s_1 s_2}\\
&\quad \equiv l^{s_1 s_2}\,(m_1 - l)^{s_1 - s_1 s_2}\,(m_2 - l)^{s_2 - s_1 s_2}\,(1 + l - m_1 - m_2)^{1 - s_1 - s_2 + s_1 s_2}.
\end{aligned}
\tag{10}
$$