A New Class of Objective Priors

Stephen Walker
University of Texas at Austin
Warwick, June 2019
Joint work with Cristiano Villa and Fabrizio Leisen

Introduction

The talk is about a new type of objective prior; the solution to

$$S(\theta, p) = c$$

where S is a score function, p is the prior on Θ space, $\theta \in \Theta$, and c is a constant, which can be 0.

The prior does not depend directly on a posited model; rather only on it through the parameter space Θ.

Most, if not all, objective priors depend in some way on the chosen model $f(x;\theta)$.

Possible concern: if model is misspecified, this error propogates through to the prior.

The prior is

$$\Pi(f \in A) = \int_{f(\cdot|\theta)\in A} p(\theta)\,\mathrm{d}\theta$$

done using $f(\cdot|\theta)$ and $p(\theta)$.

From this perspective there seems no reason to have the $f(\cdot|\theta)$ and the $p(\theta)$ connected.

Many score functions:

Log score;

$$S(\theta, p) = -\log p(\theta).$$

Tsallis score;

$$S(\theta, p) = \int p^2(x)\,dx - 2p(\theta);$$

which is a special case of the Bregman score.

$$S(\theta, p) = \tfrac{1}{2} \int \int g(x, y)\,p(x)\,p(y)\,dx\,dy - \int g(x, \theta)\,p(x)\,dx.,$$

where g is positive and non–negative definite.

A class of score function introduced in Parry et al. (2012);

$$S(\theta, p) = \frac{\partial}{\partial p} L(p, p') - \frac{\mathrm{d}}{\mathrm{d}\theta} \frac{\partial}{\partial p'} L(p, p')$$

where

$$L(p, p')$$

is (for now) a concave function.

For example, if

$$L(p, p') = p - p \log p$$

then

$$S(\theta, p) = -\log p.$$

Setting this to a constant only gives a uniform prior.

The equation

$$\frac{\partial}{\partial p} L(p, p') = \frac{\mathrm{d}}{\mathrm{d}\theta} \frac{\partial}{\partial p'} L(p, p')$$

is the Euler–Lagrange equation for extremal solutions to the problem of maximizing (minimizing)

$$I(p) = \int L(p, p')\,\mathrm{d}\theta \quad \left(I(p) = -\int L(p, p')\, d\theta \right).$$

So interpretation provided by choosing $L(p, p')$ to ensure $I(p)$ represents some measure of information.

The two most well known types of information are based on the self–information and the Fisher information;

$$I_S(p) = \int p \log p \quad \text{and} \quad I_F(p) = \tfrac{1}{2} \int \frac{(p')^2}{p},$$

respectively.

The former is based on differential entropy and is connected with Kullback–Leibler divergence; the latter is based the relative Fisher information,

$$F(p,q) = \int p \left(\frac{p'}{p} - \frac{q'}{q} \right)^2 .$$

The corresponding score functions are

$$S(\theta, p) = -\log p(\theta)$$

and

$$S(\theta, p) = -\frac{p''}{p}(\theta) + \tfrac{1}{2}\left(\frac{p'}{p}(\theta)\right)^2,$$

with respective $L(p, p')$ given by

$$L(p, p') = p \log p$$

and

$$L(p, p') = \tfrac{1}{2}\frac{(p')^2}{p}.$$

Hence, for some weighting parameter w, take

$$I(p) = I_S(p) + w\, I_F(p);$$

i.e.

$$L(p, p') = p \log p + w\, \frac{(p')^2}{p}.$$

The aim then is to set

$$\frac{\partial}{\partial p} L(p, p') - \frac{\mathrm{d}}{\mathrm{d}\theta} \frac{\partial}{\partial p'} L(p, p') \equiv 0.$$

Now

$$\frac{\partial L^2}{\partial p^2} = \frac{1}{p} + w\,4\,\frac{(p')^2}{p^3},$$

$$\frac{\partial L^2}{\partial (p')^2} = 2w\frac{1}{p}$$

and

$$\frac{\partial L^2}{\partial p \partial p'} = -2w\,\frac{p'}{p^2}.$$

Therefore the matrix

$$\begin{pmatrix} \frac{\partial L^2}{\partial p^2} & \frac{\partial L^2}{\partial p \partial p'} \\ \\ \frac{\partial L^2}{\partial p \partial p'} & \frac{\partial L^2}{\partial (p')^2} \end{pmatrix}$$

is positive–definite.

The Euler–Lagrange equation become, fixing w at 1,

$$\frac{p''(\theta)}{p(\theta)} - \tfrac{1}{2}\left(\frac{p'(\theta)}{p(\theta)}\right)^2 = 1 + \log p(\theta).$$

If

$$p(\theta) \propto e^{-u(\theta)}$$

then

$$u'(\theta) = \pm\sqrt{ce^{u(\theta)} - 2(1 + u(\theta))}.$$

Here c and e.g. $u(0)$ need to be specified; done so using constraints.

In most cases will use numerical solutions.

This is not too much of a problem; for if we end sampling θ via Markov chain methods, then interest is in

$$\frac{p(\theta')}{p(\theta)} = \exp\{u(\theta) - u(\theta')\}$$

where

$$|\theta' - \theta|$$

is small.

That is

$$u(\theta') - u(\theta) \approx (\theta' - \theta)\, u'(\theta) + \tfrac{1}{2}(\theta' - \theta)^2 u''(\theta) + \cdots$$

and

$$u''(\theta) = \tfrac{1}{2} c e^{u(\theta)} - 1.$$

$\Theta = (0,1)$: Constrain $p(0) = p(1) = 0$.

For this take $c = 2$ as an extremal value; i.e. $e^u > 1 + u$ and $u(\frac{1}{2}) > 0$. As $u(\frac{1}{2}) \uparrow$ so $p(0)$ and $p(1)$ get closer to 0.

$$u' = \begin{cases} \sqrt{2}\sqrt{e^u - 1 - u} & \theta > \frac{1}{2} \\ -\sqrt{2}\sqrt{e^u - 1 - u} & \theta < \frac{1}{2} \end{cases}$$

Here

$$u(\theta) \uparrow \infty \quad \text{as} \quad \theta \to 0, 1$$

and u, and hence p, is symmetric about $\theta = \frac{1}{2}$.

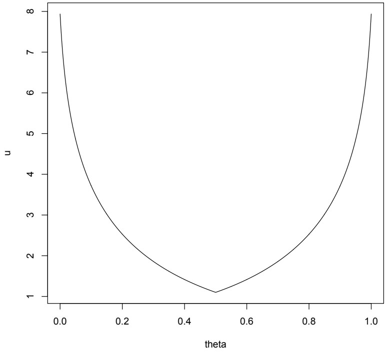

Figure: $u(\frac{1}{2}) = 1.1$

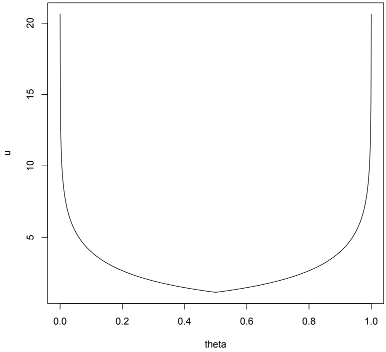

Figure: $u(\frac{1}{2}) = 1.14$

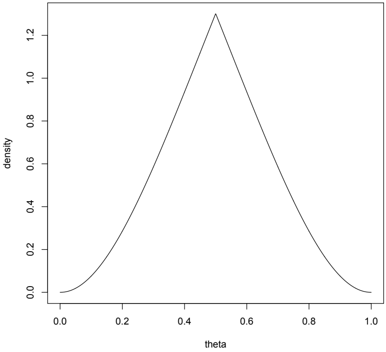

Figure: Normalized p with $u(\frac{1}{2}) = 1.14$

$\Theta = (0, \infty)$: constrain p convex and decreasing.

For $p' < 0$ we need $u' > 0$, so taking $c = 2$ as extremal,

$$u' = \sqrt{2}\sqrt{e^u - 1 - u}.$$

For p to be convex we require $p'' > 0$.

For this we need

$$(u')^2 \geq u''.$$

For this require

$$u(0) > \tfrac{1}{2} \quad \text{and} \quad 1 + 2u(0) = e^{u(0)}$$

which gives

$$u(0) = 1.31.$$

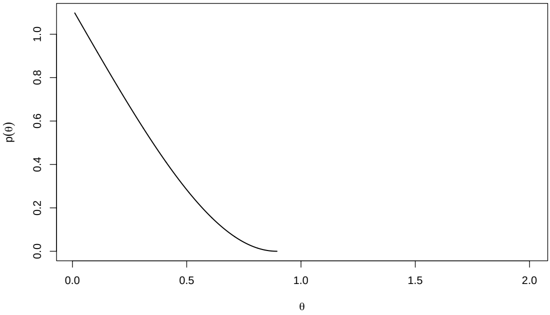

Figure: Plot of the prior on $(0, \infty)$ with $c = 2$ and $u(0) = 1.31$

$\Theta = (-\infty, +\infty)$: Constrain $p'(0) = 0$.

For this require $u'(0) = 0$.

If now take $c = 2$ then u becomes a constant; we need $u(0) = 0$ etc. So p is flat.

For a proper p, take $u(0) > 0$ and then

$$c = 2(1 + u(0))\, e^{-u(0)}.$$

Here $u(0)$ will determine the variance; we have $u''(0) = u(0)$ and in the normal case

$$u(\theta) = \tfrac{1}{2}\theta^2/\sigma^2$$

yielding $u''(0) = 1/\sigma^2$.

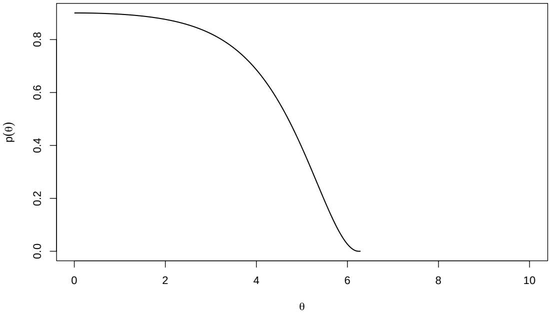

Figure: Plot of the positive half of the prior on $(-\infty, \infty)$, obtained by symmetrising one on $(0, \infty)$ and making differentiable at the origin

$\Theta = (-M, +M)$: Constrain a direct solution.

To pick up a direct solution we need to take $c = 0$; so we need $u < -1$.

The solution to u is then of the form

$$1 + u(\theta) = -\tfrac{1}{2}(\theta - \mu)^2$$

for some μ.

Yielding

$$p(\theta) \propto \exp\left\{\tfrac{1}{2}(\theta - \mu)^2\right\}$$

which is proper.

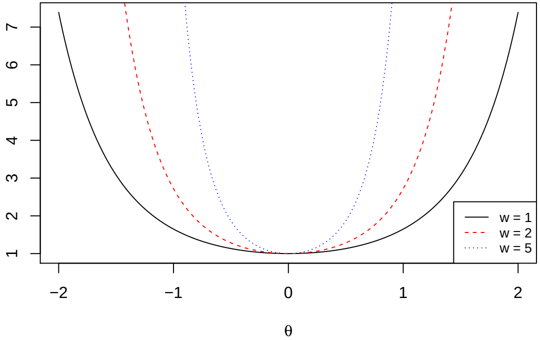

Figure: Prior for the parameter space $(-2, 2)$ with weights $w = 1$ (continuous black line), $w = 2$ (dashed red line) and $w = 5$ (dotted blue line).

•: Extension to multivariate parameter $(\theta_1, \ldots, \theta_d)$;

$$\frac{\partial u}{\partial \theta_j} = \pm\sqrt{c_j e^{u(\theta)} - 2(1 + u(\theta))}.$$

•: General weighting of two components;

$$S(\theta, p) = w\, S_L(\theta, p) + S_H(\theta, p)$$

yields

$$u'(\theta) = \pm\sqrt{c e^{u(\theta)} - 2w(1 + u(\theta))}.$$

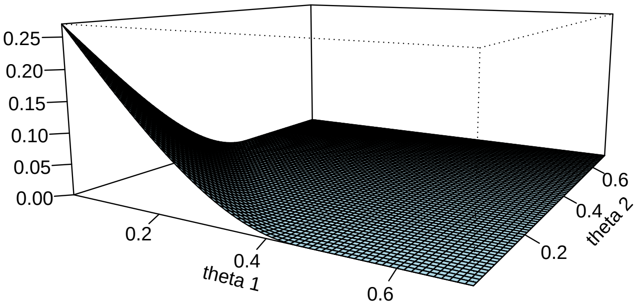

Figure: Surface plot of the prior for the bidimensional $(0, \infty)^2$.

Consider

$$I(p) = \int_\Omega p \log p$$

with $\Omega = (0, 1)$.

Well known that $I(p) \geq 0$ and $I(p) = 0$ only when p is uniform.

Now transform to $\Omega = (0, \infty)$ using $-logx$;

$$p(x) \rightarrow q(x) = p(e^{-x})\, e^{-x}.$$

Specifically,

$$p(x) = ax^{a-1} \rightarrow q(x) = a\, e^{-ax}.$$

So

$$I(p) = \log a - 1 + 1/a \geq 0$$

and is equal 0 when $a = 1$.

And

$$I(q) = \log a - 1.$$

When $a < 1$,

$$I(p_a) > I(p_1)$$

yet

$$I(q_a) < I(q_1).$$