Noninformative Priors

Aleksandar Dogandžić

2017-03-11

Contents

READING: [Gelman et al. 2014, §2.8].

Preliminaries

THE highest posterior density (HPD) approach is *not* invariant to parameter transformations.

✻ EXAMPLE. Suppose that $0 < \theta < 1$ is a parameter of interest, but that we are equally interested in

$$\gamma = \frac{1}{1-\theta}.$$

If the posterior probability density function (pdf) $f_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x})$ is, say,

$$\begin{aligned} f_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x}) &= \begin{cases} 2\theta, & 0 < \theta < 1 \\ 0, & \text{otherwise} \end{cases} \\ &= 2\theta \, \mathbb{1}_{(0,1)}(\theta) \end{aligned}$$

the corresponding cumulative distribution function (cdf) is

$$F_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x}) = \begin{cases} 0, & \theta < 0 \\ \theta^2, & 0 < \theta < 1 \\ 1, & \theta > 1 \end{cases}$$

and a 95 % HPD credible set for θ is $(\sqrt{0.05}, 1)$. Now, despite the fact that

$$\gamma = \frac{1}{1-\theta}$$

is a monotonic function of θ, the interval

$$\Big(\frac{1}{1-\sqrt{0.05}}, +\infty\Big) \tag{1}$$

is not an HPD credible set for γ. Clearly, (1) is a 95 % credible set for γ, but it is not HPD. To see this, we find the cdf $F_{\Gamma|\boldsymbol{X}}(\gamma \mid \boldsymbol{x})$: for $t \geq 1$,

$$\begin{aligned} F_{\Gamma|\boldsymbol{X}}(t \mid \boldsymbol{x}) &= \Pr_{\Gamma|\boldsymbol{X}}\left\{\Gamma \leq t \mid \boldsymbol{X} = \boldsymbol{x}\right\} \\ &= \Pr_{\Theta|\boldsymbol{X}}\left(\frac{1}{1-\Theta} \leq t \,\middle|\, \boldsymbol{X} = \boldsymbol{x}\right) \\ &= \Pr_{\Theta|\boldsymbol{X}}\left(\Theta \leq 1 - \frac{1}{t} \,\middle|\, \boldsymbol{x} = \boldsymbol{x}\right) \\ &= \left(1 - \frac{1}{t}\right)^2. \end{aligned}$$

Therefore, $\{\Gamma \mid \boldsymbol{x} = \boldsymbol{x}\}$ has the pdf

$$f_{\Gamma|\boldsymbol{X}}(\gamma \mid \boldsymbol{x}) = \begin{cases} 2(1 - 1/\gamma)/\gamma^2, & \gamma \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

with $f_{\Gamma|\boldsymbol{X}}(1 \mid \boldsymbol{x}) = 0$ and, consequently, HPD intervals for $\{\Gamma \mid \boldsymbol{X} = \boldsymbol{x}\}$ must be two-sided, which is in contrast with (1).

☞ IF we switch perspective and think of prior distributions rather than posteriors, this kind of thinking raises concerns about the whole idea of a "flat prior". What is "flat" for one parameterization will not be "flat" for another.

Suppose

$$f_\Theta(\theta) = \mathrm{U}(\theta \mid 0, 1).$$

Then,

$$\gamma = \frac{1}{1-\theta}$$

has pdf

$$f_\Gamma(\gamma) = \frac{1}{\gamma^2}\,\mathbb{1}_{[1,+\infty)}(\gamma).$$

Notions of shape of priors and posteriors are *not independent* of the choice of parameterization.

Reparameterizing the variability parameter of a Gaussian distribution

ALTHOUGH it may seem that picking a noninformative prior distribution is easy, e.g., just use a uniform pdf or probability mass function (pmf), it is not straightforward.

✻ EXAMPLE. Estimating the standard deviation and variance of a Gaussian distribution with known mean.

Consider now N conditionally independent, identically distributed (i.i.d.) observations $(X[n])_{n=0}^{N-1}$ given the standard deviation σ, where

$$\{X[n] \mid \sigma\} \sim \mathcal{N}(0, \sigma^2) \tag{2a}$$

$$f_\sigma(\sigma) \propto \mathbb{1}_{[0,\infty)}(\sigma) \tag{2b}$$

i.e., we assume a uniform prior (from zero to infinity) for the standard deviation σ.

QUESTION: What is the equivalent prior for the variance σ^2?

We now apply the change-of-variables formula to our problem: $h(\sigma^2) = \sqrt{\sigma^2}, h'(\sigma) = 0.5(\sigma^2)^{-0.5}$, yielding

$$f_{\sigma^2}(\sigma^2) \propto \frac{1}{2\sqrt{\sigma^2}} \mathbb{1}_{[0,\infty)}(\sigma^2) \tag{3}$$

Recall transformation of random variables:

$$f_Y(y) = f_X(h(y))|h'(y)|$$

where $x = h(y)$; see handout `revprob`.

which is *not uniform*. Therefore, (2b) implies (3), which means that our prior belief is that the variance σ^2 is small.

Similarly, for the uniform prior on the variance σ^2:

$$f_{\sigma^2}(\sigma^2) \propto \mathbb{1}_{[0,\infty)}(\sigma^2)$$

the equivalent prior on the standard deviation σ is

$$f_\sigma(\sigma) \propto 2\sigma\, \mathbb{1}_{[0,\infty)}(\sigma)$$

implying that we believe that the standard deviation σ is large.

A way to visualize the observed phenomenon observed is to look at what happens to intervals of equal measure.

In the case of σ^2 being uniform, an interval $[a, a+0.1]$ must have the same prior measure as the interval $[0.1, 0.2]$. When we transform to σ, the corresponding prior measure must have intervals $[\sqrt{a}, \sqrt{a+0.1}]$ having equal measure. But, the length of $[\sqrt{a}, \sqrt{a+0.1}]$ is a decreasing function of a, which agrees with the increasing density in σ.

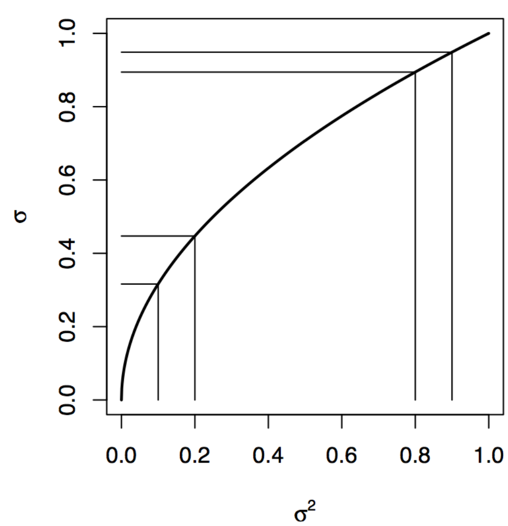

Figure 1: $\sigma = \sqrt{\sigma^2}$ as a function of σ^2.

Therefore, when talking about non-informative priors, we need to think about scale.

Jeffreys' Priors

CAN we pick a prior where the scale of the parameter does not matter?

Jeffreys' general principle states that any rule for determining the prior density $f_\Theta(\theta)$ for parameter θ should yield an equivalent result if applied to the transformed parameter ϕ, where $\theta(\phi)$ is a one-to-one transform. Therefore, applying the prior

$$f_\Phi(\phi) = f_\Theta(\theta(\phi))\,\left|\theta'(\phi)\right|$$

for Φ should give the same answer as dealing directly with the transformed model,

$$f_{\boldsymbol{X},\Phi}(\boldsymbol{x},\phi) = f_\Phi(\phi)\, f_{\boldsymbol{X}\,|\,\Phi}(\boldsymbol{x}\,|\,\phi).$$

Recall the change-of-variables formula for Cramér-Rao bound (CRB) derived in HW:

$$\mathcal{I}_\Phi(\phi) = \mathcal{I}_\Theta(\theta(\phi))\,\left|\theta'(\phi)\right|^2$$

implying

$$\sqrt{\mathcal{I}_\Phi(\phi)} = \sqrt{\mathcal{I}_\Theta(\theta(\phi))}\,\left|\theta'(\phi)\right|.$$

✻ JEFFREYS' prior:

$$f_\Theta(\theta) \propto \sqrt{\mathcal{I}_\Theta(\theta)} \tag{4}$$

where $\mathcal{I}_\Theta(\theta)$ is the Fisher information for θ. Now, the transformed prior for Φ is

$$\begin{aligned} f_\Phi(\phi) &= \underbrace{f_\Theta(\theta(\phi))}_{\propto \sqrt{\mathcal{I}(\theta(\phi))}}\,\left|\theta'(\phi)\right| \\ &\propto \sqrt{\mathcal{I}_\Phi(\phi)} \end{aligned}$$

because $\sqrt{\mathcal{I}_\Phi(\phi)} = \sqrt{\mathcal{I}_\Theta(\theta(\phi))}|\theta'(\phi)|$.

☞ IF we make a one-to-one transform $\phi = \phi(\theta)$, then the Jeffreys' prior for the transformed model is

$$f_\Phi(\phi) \propto \sqrt{\mathcal{I}_\Phi(\phi)}. \tag{5}$$

Estimating the variance of a Gaussian distribution with known mean

THE Fisher information for variance σ^2 of i.i.d. Gaussian measurements is[1]:

$$\mathcal{I}_{\sigma^2}(\sigma^2) = \frac{N}{2(\sigma^2)^2}.$$

[1] see eq. (**??**) in handout `multipar_gauss_crb`

Therefore, the Jeffreys' prior for σ^2 is

$$f_{\sigma^2}(\sigma^2) \propto \frac{1}{\sigma^2}\,\mathbb{1}_{[0,+\infty)}(\sigma^2). \tag{6}$$

Alternative descriptions under different parameterizations for the variance parameter are (for $\sigma > 0$)

uniform on $(-\infty, +\infty)$

$$f_\sigma(\sigma) \propto \frac{1}{\sigma}\mathbb{1}_{[0,+\infty)}(\sigma), \qquad f_{\ln\sigma^2}(\ln\sigma^2) \propto 1. \tag{7}$$

Here, $f_{\ln\sigma^2}(\ln\sigma^2) \propto 1$ means that

$$f_Q(q) \propto 1$$

U$(-\infty,+\infty)$

where $Q = \ln\sigma^2$.

Estimating the mean of a Gaussian distribution with known variance

CONSIDER N i.i.d. observations $(X[n])_{n=0}^{N-1}$ given $\Theta = \theta$, following

$$\{X[n] \mid \Theta = \theta\} \sim \mathcal{N}(\theta, \sigma^2)$$

where σ^2 is a known constant. Here

$$\mathcal{I}(\theta) = \frac{N}{\sigma^2} = \text{const}$$

and, therefore, the clearly improper Jeffreys' prior for θ is

$$f(\theta) \propto 1.$$

U$(-\infty,+\infty)$

References

Gelman, A., J. B. Carlin, H. S. Stern, David B. Dunson, Aki Vehtari, and D. B. Rubin (2014). *Bayesian Data Analysis*. 3rd ed. Boca Raton, FL: Taylor & Francis (cit. on p. 1).