Chapter 6
Prior Elicitation

As discussed previously, the prior belief depends on a person's knowledge and experience. In order to obtain a good prior, one sometimes needs to perform some expert elicitation. Nevertheless, there is no universal rule and one has to be careful in constructing a prior. In fact, prior construction is a subject of current research, and it is problem-dependent.

6.1 Smooth priors

In this section, we believe that the unknown function $f(t)$ is smooth, which can be translated into, among other possibilities, the following simplest requirement on the pointwise values $f(s_i)$, and hence m_i,

$$m_i = \frac{1}{2}\left(m_{i-1} + m_{i+1}\right), \tag{6.1}$$

that is, the value of $f(s)$ at a point is more or less the same of its neighbor. But, this is by no means the correct behavior of the unknown function $f(s)$. We therefore admit some uncertainty in our belief (6.1) by adding an *innovative* term W_j such that

$$m_i = \frac{1}{2}\left(m_{i-1} + m_{i+1}\right) + W_j,$$

where $W \sim \mathcal{N}\left(0, \gamma^2 I\right)$. The standard deviation γ determines how much the reconstructed function $f(t)$ departs from the smoothness model (6.1). In terms of matrices, we obtain

$$Lm = W,$$

where L is given by

$$L = \frac{1}{2}\begin{bmatrix} -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times(n+1)},$$

which is the second order finite difference matrix approximating the Laplacian Δf. Indeed,

$$\Delta f(s_j) \approx n^2 (Lm)_j. \tag{6.2}$$

Suppose $\det(L) \neq 0$, the prior distribution is therefore given by (using the technique in Section 5.1)

$$\pi_{\text{pre}} \propto \exp\left(-\frac{1}{\gamma^2}\|Lm\|^2\right). \tag{6.3}$$

But L has rank of $n-1$, and hence π_{pre} is a degenerate Gaussian density in $\mathbb{R}^{n+1}$. In fact, (6.3) is not valid since $\det(L) = 0$. The reason is that we have not specified the smoothness of $f(s)$ at the boundary points. In other words, we have not specified any boundary conditions for the Laplacian $\Delta f(s)$. This is a crucial point in prior elicitation via differential operators. One needs to make sure that the operator is positive definite by incorporating some well-posed boundary conditions. Throughout the lecture notes, unless otherwise stated, $\|\cdot\|$ denotes the usual Euclidean norm.[1]

Let us first consider the case with zero Dirichlet boundary condition, that is, we believe that $f(s)$ is smooth and (close to) zero at the boundaries, then

$$\begin{aligned} m_0 &= \frac{1}{2}(m_{-1} + m_1) + W_0 = \frac{1}{2}m_1 + W_0, \quad W_0 \sim \mathcal{N}\left(0, \gamma^2\right) \\ m_n &= \frac{1}{2}(m_{n-1} + m_{n+1}) + W_n = \frac{1}{2}m_{n-1} + W_n, \quad W_n \sim \mathcal{N}\left(0, \gamma^2\right). \end{aligned}$$

Note that we have extended $f(s)$ by zero outside the domain $[0,1]$ since we "know" that it is smooth. Consequently, we have $L_D m = W$ with

$$L_D = \frac{1}{2}\begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}, \tag{6.4}$$

which is the second order finite difference matrix corresponding to zero Dirichlet boundary conditions. The prior density in this case reads

$$\pi^D_{\text{prior}}(m) \propto \exp\left(-\frac{1}{2\gamma^2}\|L_D m\|^2\right). \tag{6.5}$$

[1] The ℓ^2-norm if you wish

It is instructive to draw some random realizations from π_{prior}^D (we are ahead of ourselves here since sampling will be discussed in Chapter 8), and we show five of them in Figure 6.1 together with the prior standard deviation curve. As can be seen, all the draws are almost zero at the boundary and the prior variance (uncertainty) is close to zero as well. This is not surprising since our prior belief says so. How do we compute the standard deviation curve? Well, it is straightforward. We first compute the pointwise variance as

Why are they not exactly zero?

$$\mathbb{V}ar\left[m_j\right] \stackrel{\text{def}}{=} \mathbb{E}\left[m_j^2\right] = e_j^T \left(\int_{\mathbb{R}^{n+1}} mm^T \pi_{\text{prior}}^D \, dm\right) e_j \stackrel{\text{def}}{=} \gamma^2 e_j^T \left(L_D^T L_D\right)^{-1} e_j,$$

where e_j is the jth canonical basis vector in $\mathbb{R}^{n+1}$, and we have used the fact that the prior is Gaussian in the last equality. So we in fact plot the square root of the diagonal of $\gamma^2 \left(L_D^T L_D\right)^{-1}$, the covariance matrix, as the standard deviation curve. One can see that the uncertainty is largest in the middle of the domain since it is farthest from the constrained boundary. The points closer to the boundaries have smaller variance, that is, they are more correlated to the "known" boundary data, and hence less uncertain.

Do we really have the complete continuous curve?

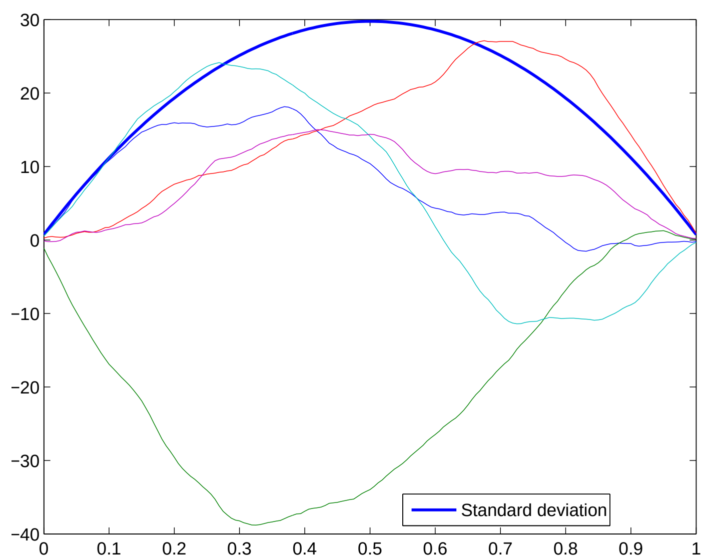

Fig. 6.1 Prior random draws from π_{prior}^D together with the standard deviation curve.

Now, you may ask why $f(s)$ must be zero at the boundary, and you are right! There is no reason to believe that must be the case. However, we don't know the exact values of $f(s)$ at the boundary either, even though we believe that we may have non-zero Dirichlet boundary condition. If this is the case, we have to admit our ignorance and let the data from the likelihood correct us in the posterior. To be consistent with the Bayesian philosophy, if we do not know anything about boundary conditions, let them be, for convenience, Gaussian random variables such as

$$m_0 \sim \mathcal{N}\left(0, \frac{\gamma^2}{\delta_0^2}\right), \quad m_n \sim \mathcal{N}\left(0, \frac{\gamma^2}{\delta_n^2}\right).$$

Hence, the prior can now be written as

$$\pi_{\text{prior}}^R(m) \propto \exp\left(-\frac{1}{2\gamma^2}\|L_R m\|^2\right), \tag{6.6}$$

where

$$L_R = \frac{1}{2}\begin{bmatrix} 2\delta_0 & 0 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & 0 & 2\delta_n \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}.$$

A question that immediately arises is how to determine δ_0 and δ_n. Since the boundary values are now independent random variables, we are less certain about them compared to the previous case. But to which uncertain level we want them to be? Well, let's make every values equally uncertain, meaning we have the same ignorance about the values at these points, that is, we would like to have the same variances everywhere. To approximately accomplish this, we require

$$\mathbb{V}ar[m_0] = \frac{\gamma^2}{\delta_0^2} = \mathbb{V}ar[m_n] = \frac{\gamma^2}{\delta_n^2} = \mathbb{V}ar\left[m_{[n/2]}\right] = \gamma^2 e_{[n/2]}^T \left(L_R^T L_R\right)^{-1} e_{[n/2]},$$

where $[n/2]$ denotes the largest integer smaller than $n/2$. It follows that

$$\delta_0^2 = \delta_n^2 = \frac{1}{e_{[n/2]}^T \left(L_R^T L_R\right)^{-1} e_{[n/2]}}.$$

Is it sensible to do so?

However, this requires to solve a nonlinear equation for $\delta_0 = \delta_n$, since L_R depends on them. To simplify the computation, we use the following approximation

$$\delta_0^2 = \delta_n^2 = \frac{1}{e_{[n/2]}^T \left(L_D^T L_D\right)^{-1} e_{[n/2]}}.$$

Again, we draw five random realizations from π_{prior}^R and put them together with the standard deviation curve in Figure 6.2. As can be observed, the uncertainty is more or less the same at every point and prior realizations are no longer constrained to have zero boundary conditions.

Exercise 6.1. Consider the following general scheme

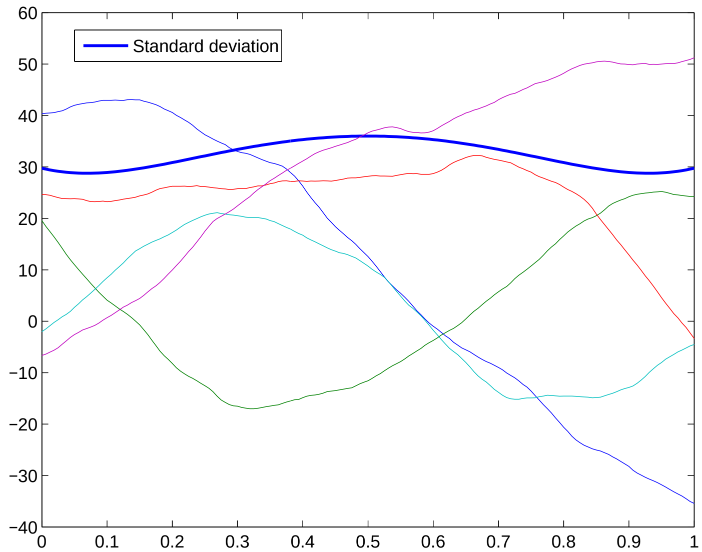

Fig. 6.2 Prior random draws from π^R_{prior} together with the standard deviation curve.

$$m_i = \lambda_i m_{i-1} + (1-\lambda_i) m_{i+1} + e_i, \quad 0 \le \lambda_i \le 1.$$

Convince yourself that by choosing a particular set of λ_i, you can recover all the above prior models. Replace `BayesianPriorElicitation.m` by a generic code with input parameters λ_i. Experience new prior models by using different values of λ_i (those that don't reproduce priors presented in the text). •

Exercise 6.2. Construct a prior with a non-zero Dirichlet boundary condition at $s = 0$ and zero Neumann boundary condition at $s = 1$. Draw a few samples together with the variance curve to see whether your prior model indeed conveys your belief. •

6.2 “Non-smooth” priors

We first consider the case in which we believe that $f(s)$ is still smooth but may have discontinuities at known locations on the mesh. Can we design a prior to convey this belief? A natural approach is to require that m_j is equal to m_{j-1} plus a random jump, i.e.,

$$m_j = m_{j-1} + e_j,$$

where $e_j \sim \mathcal{N}\left(0, \gamma^2\right)$, and for simplicity, let us assume that $m_0 = 0$. The prior density in this case would be

$$\pi_{\text{pren}}(m) \propto \exp\left(-\frac{1}{2\gamma^2}\|L_N m\|^2\right), \tag{6.7}$$

where

$$L_N = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

But, if we think that there is a particular big jump, relative to others, from m_{j-1} to m_j, then the mathematical translation of this belief is $e_j \sim \mathcal{N}\left(0, \frac{\gamma^2}{\theta^2}\right)$ with $\theta < 1$. The corresponding prior in this case reads

$$\pi^{O}_{\text{prior}}(m) \propto \exp\left(-\frac{1}{2\gamma^2}\|JL_N m\|^2\right), \tag{6.8}$$

with

$$J = \text{diag}\left(1,\ldots,1, \underbrace{\theta}_{j\text{th index}}, 1,\ldots,1\right).$$

Let's draw some random realizations from $\pi^{O}_{\text{prior}}(m)$ in Figure 6.3 with $n = 160$, $j = 80$, $\gamma = 1$, and $\theta = 0.01$. As desired, all realizations have a sudden jump at $j = 80$, and the standard deviation of the jump is $1/\theta = 100$. In addition, compared to priors in Figure 6.1 and 6.2, the realizations from $\pi^{O}_{\text{prior}}(m)$ are less smooth, which confirms that our belief is indeed conveyed.

Exercise 6.3. Use `BayesianPriorElicitation.m` to construct examples with 2 or more sudden jumps and plot a few random realizations to see whether your belief is conveyed. •

A more interesting and more practical situation is the one in which we don't know how many jump discontinuities and their locations. A natural prior in this situation is a generalized version of (6.8), e.g.,

$$\pi^{M}_{\text{prior}}(m) \propto C(M)\exp\left(-\frac{1}{2\gamma^2}\|ML_N m\|^2\right), \tag{6.9}$$

with

$$M = \text{diag}(\theta_1,\ldots,\theta_n),$$

What is wrong with this?

where θ_i, $i = 1,\ldots,n$, are unknown. In fact, these are called *hyper-parameters* and one can determine them using information from the likelihood; the readers are referred to [10] for the details.

Exercise 6.4. Modify the scheme in Exercise 6.1 to include priors with sudden jumps. •

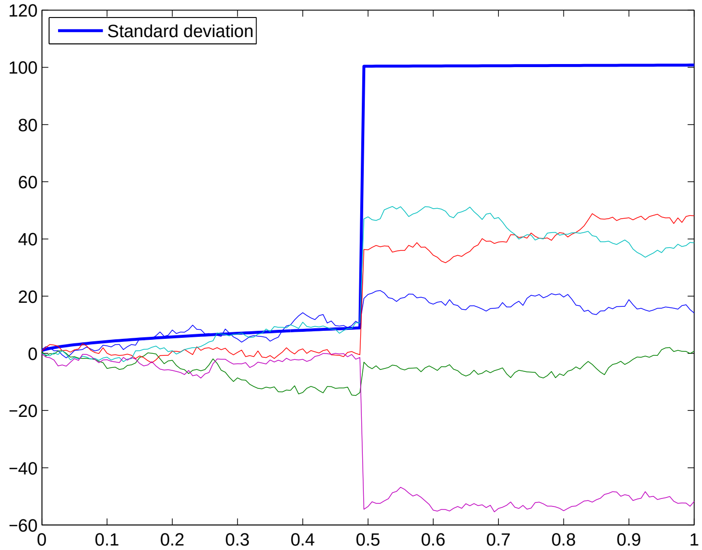

Fig. 6.3 Prior random draws from π^{O}_{prior} together with the standard deviation curve.