Posterior for f and posterior predictive distribution

Haziq Jamil

01/04/2017

The regression model

For $i = 1, \dots, n$,

$$\begin{aligned} y_i &= f(x_i) + \epsilon_i \\ \epsilon_i &\overset{\text{iid}}{\sim} \mathrm{N}(0, \psi^{-1}) \end{aligned}$$

An I-prior on f

$$\big(f(x_1), \dots, f(x_n)\big) \sim \mathrm{N}\big(\mathbf{f}_0, \mathcal{I}(f)\big)$$

where $\mathcal{I}$ is the Fisher information for f, and has (i, j) entries

$$\begin{aligned} \mathcal{I}\left(f(x_i), f(x_j)\right) &= \psi \sum_{k,l=1}^{n} h_\lambda(x_i, x_k) h_\lambda(x_j, x_l) \\ &= \psi \mathbf{H}_\lambda(i, \cdot)\mathbf{H}_\lambda(\cdot, j) = \psi \mathbf{H}_\lambda^2(i, j) \end{aligned}$$

and $\mathbf{f}_0$ is a (chosen) prior mean. For now, set $\mathbf{f}_0 = \big(f_0(x_1), \dots, f_0(x_n)\big) = \alpha \mathbf{1}_n$. The regression model is equivalent to

$$\begin{aligned} y_i &= \alpha + \sum_{k=1}^{n} h_\lambda(x_i, x_k) w_k + \epsilon_i \\ w_k &\overset{\text{iid}}{\sim} \mathrm{N}(0, \psi) \\ \epsilon_i &\overset{\text{iid}}{\sim} \mathrm{N}(0, \psi^{-1}) \end{aligned}$$

and this is the model estimated by the `iprior` package.

Posterior distribution for f

Denote $f = f(x)$, the point evaluation of f at some $x \in \mathcal{X}$. We are interested in the posterior of f

$$p(f|\mathbf{y}) = \frac{p(\mathbf{y}|f)p(f)}{\int p(\mathbf{y}|f)p(f)\,\mathrm{d}f}$$

Since everything is Gaussian, this is easy to derive (without the integrals). We have firstly the marginals $f \sim \mathrm{N}\left(f_0, \psi \sum_{k=1}^{n} h_\lambda(x, x_k)\right)$, and $\mathbf{y} \sim \mathrm{N}(\mathbf{f}_0, \mathbf{V}_y)$, where $\mathbf{V}_y = \psi \mathbf{H}_\lambda^2 + \psi^{-1}\mathbf{I}_n$. The covariance between f and $\mathbf{y}$ is the $n \times 1$ row-vector

$$
\begin{aligned}
\operatorname{Cov}[f, \mathbf{y}] &= \operatorname{Cov}[f, \big(f(x_1), \dots, f(x_n)\big)] \\
&= (\operatorname{Cov}[f, f(x_1)], \dots, \operatorname{Cov}[f, f(x_n)])
\end{aligned}
$$

where each $ Cov[f, f(x_i)]$ is equal to $\psi \sum_{k=1}^n h_\lambda(x, x_k) h_\lambda(x_i, x_k)$. The joint distribution is then

$$
\begin{pmatrix} f \\ \mathbf{y} \end{pmatrix} = \left(\begin{pmatrix} f_0 \\ \mathbf{f}_0 \end{pmatrix}, \begin{pmatrix} \psi \mathbf{h}_\lambda^* \mathbf{h}_\lambda^{*\top} & \psi \mathbf{H}_\lambda \mathbf{h}_\lambda^{*\top} \\ \psi \mathbf{h}_\lambda^* \mathbf{H}_\lambda & \mathbf{V}_y \end{pmatrix} \right)
$$

where we have collected the values $\mathbf{h}_\lambda(x, x_k)$, $k = 1, \dots, n$ in the $n \times 1$ row-vector $\mathbf{h}^*$. Thus the posterior distribution $\mathbf{f}|\mathbf{y}$ is normal with mean and variance

$$
\mathrm{E}[\mathbf{f}|\mathbf{y}] = f_0 + \psi \mathbf{h}_\lambda^* \mathbf{H}_\lambda \mathbf{V}_y^{-1} (\mathbf{y} - \mathbf{f}_0)
$$

$$
\begin{aligned}
\operatorname{Var}[\mathbf{f}|\mathbf{y}] &= \psi \mathbf{h}_\lambda^* \mathbf{h}_\lambda^{*\top} - \psi^2 \mathbf{h}_\lambda^* \mathbf{H}_\lambda \mathbf{V}_y^{-1} \mathbf{H}_\lambda \mathbf{h}_\lambda^{*\top} \\
&= \mathbf{h}_\lambda^* (\psi \mathbf{I}_n - \psi^2 \mathbf{H}_\lambda \mathbf{V}_y^{-1} \mathbf{H}_\lambda) \mathbf{h}_\lambda^{*\top} \\
&= \mathbf{h}_\lambda^* \mathbf{V}_y^{-1} \mathbf{h}_\lambda^{*\top}
\end{aligned}
$$

Posterior predictive distribution for new observations y^*

Consider the joint distribution $(\mathbf{y}, y^*)$. This is multivariate normal with mean $\boldsymbol{\alpha}$ and variance

$$
\begin{pmatrix} \mathbf{V}_y & \mathbf{c}^\top \\ \mathbf{c} & \operatorname{Var}[y^*] \end{pmatrix}
$$

where $\operatorname{Var}[y^*] = \psi \sum_{k=1}^n h_\lambda^2(x^*, x_i) + 1/\psi$, and the covariance between $\mathbf{y}$ and y^* is

$$
\operatorname{Cov}[\mathbf{y}, y^*] = \begin{pmatrix} \operatorname{Cov}[y_1, y^*] & \cdots & \operatorname{Cov}[y_n, y^*] \end{pmatrix} =: \mathbf{c}
$$

and

$$
\begin{aligned}
\operatorname{Cov}[y_i, y^*] &= \operatorname{Cov}\left[\alpha + \sum_{k=1}^n h_\lambda(x_i, x_k) + \epsilon_i, \alpha + \sum_{k=1}^n h_\lambda(x^*, x_k) + \epsilon^* \right] \\
&= \psi \sum_{k=1}^n h_\lambda(x_i, x_k) h_\lambda(x^*, x_k)
\end{aligned}
$$

and in matrix notation, the $n \times 1$ covariance vector $\mathbf{c}^\top$ can be written as

$$
\mathbf{c}^\top = \psi \mathbf{H}_\lambda \mathbf{h}_\lambda^{*\top}
$$

where $\mathbf{h}_\lambda^*$ is the $1 \times n$ row-vector containing $h_\lambda(x^*, x_i)$, $i = 1, \dots, n$.

Then the posterior distribution of y^* is normal

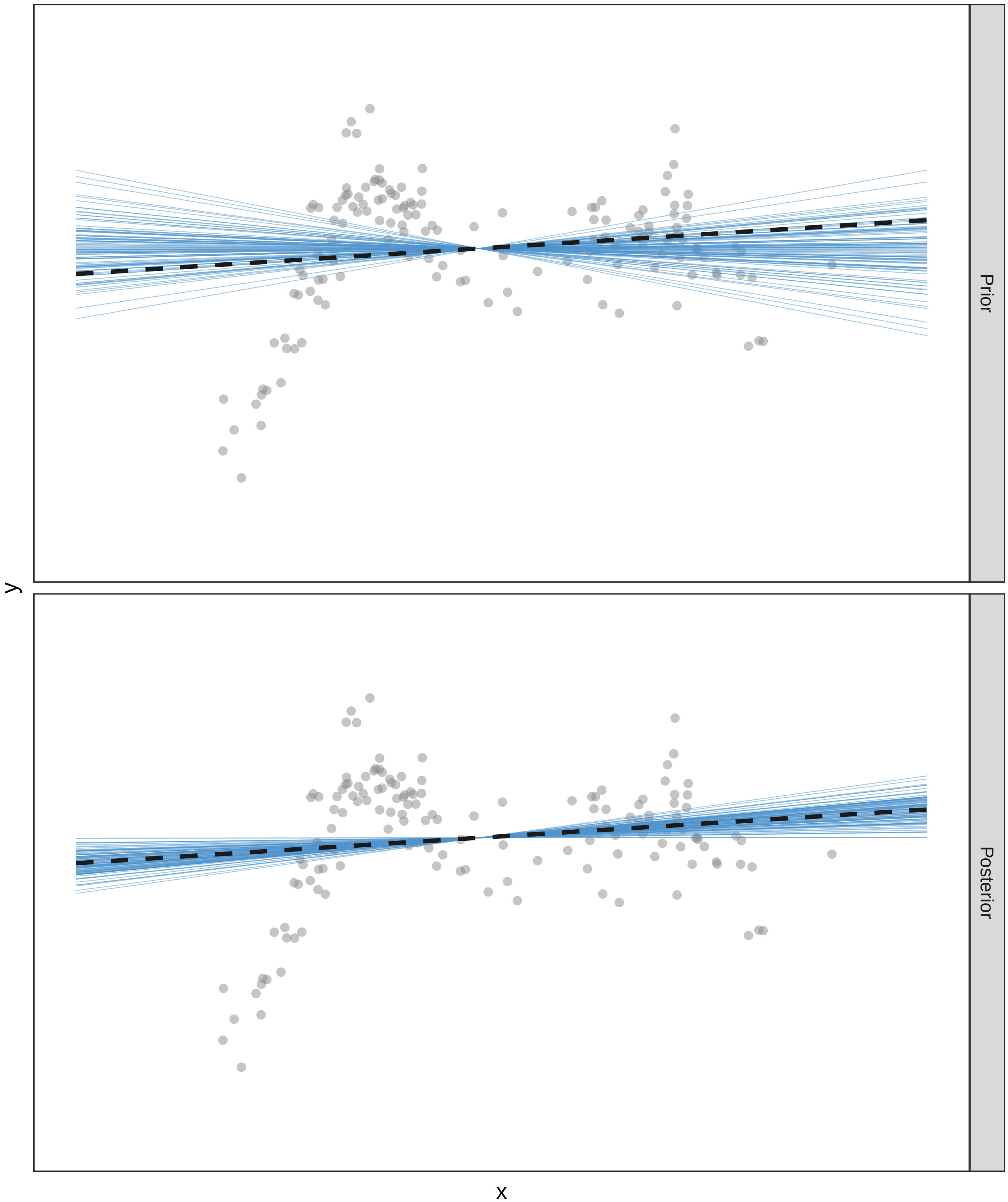
$$
(y^*|\mathbf{y}) \sim \mathrm{N}\left(\alpha + \mathbf{c}\mathbf{V}_y^{-1}(\mathbf{y} - \alpha \mathbf{1}_n), \operatorname{Var}[y^*] - \mathbf{c}\mathbf{V}_y^{-1}\mathbf{c}^\top\right)
$$

Note that the posterior mean is exactly the same as the one derived in the iprior package with the $\tilde{w}$.

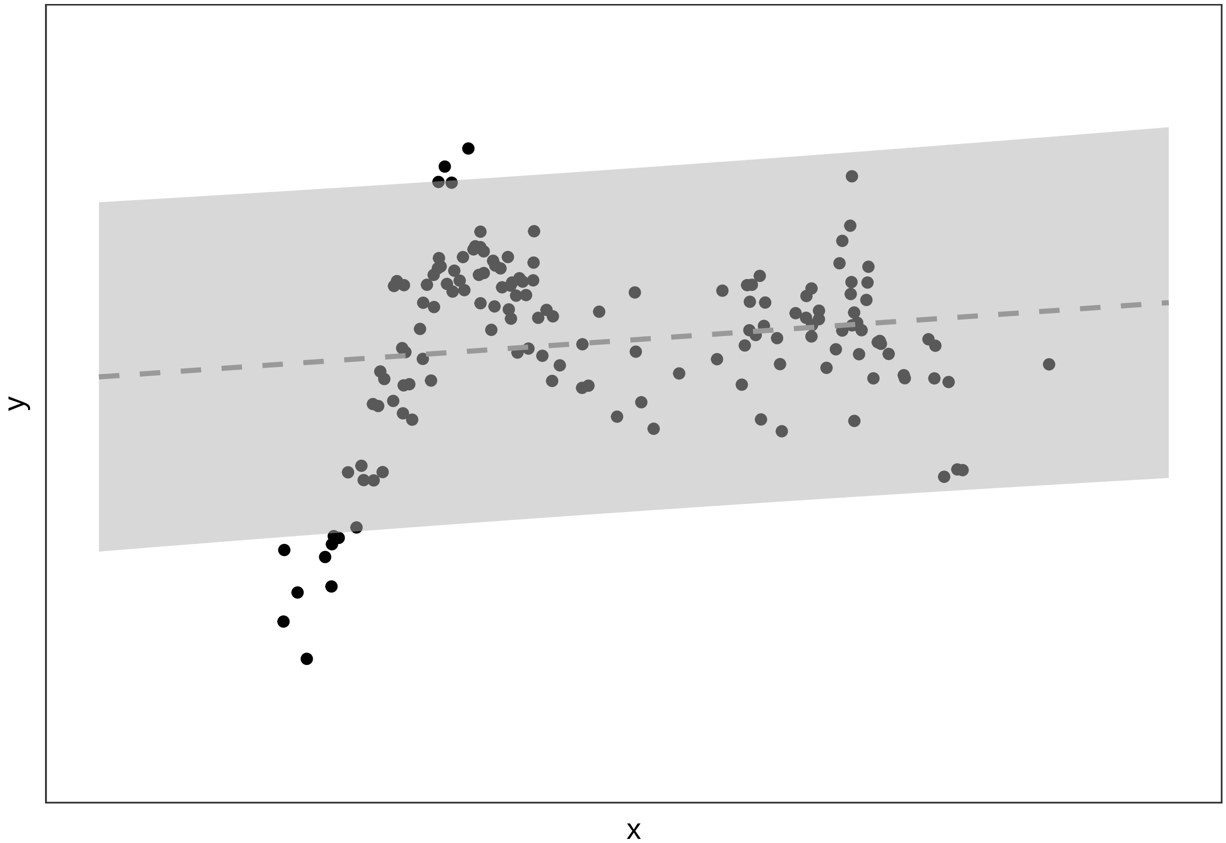
$$
\mathrm{E}[y^*|\mathbf{y}] = \alpha + \mathbf{c}\mathbf{V}_y^{-1}(\mathbf{y} - \alpha \mathbf{1}_n) = \alpha + \mathbf{h}_\lambda^* \overbrace{(\psi \mathbf{H}_\lambda \mathbf{V}_y^{-1} (\mathbf{y} - \alpha \mathbf{1}_n))}^{\tilde{\mathbf{w}} = \mathrm{E}[\mathbf{w}|\mathbf{y}]}
$$

Canonical kernel

Prior and posterior draws

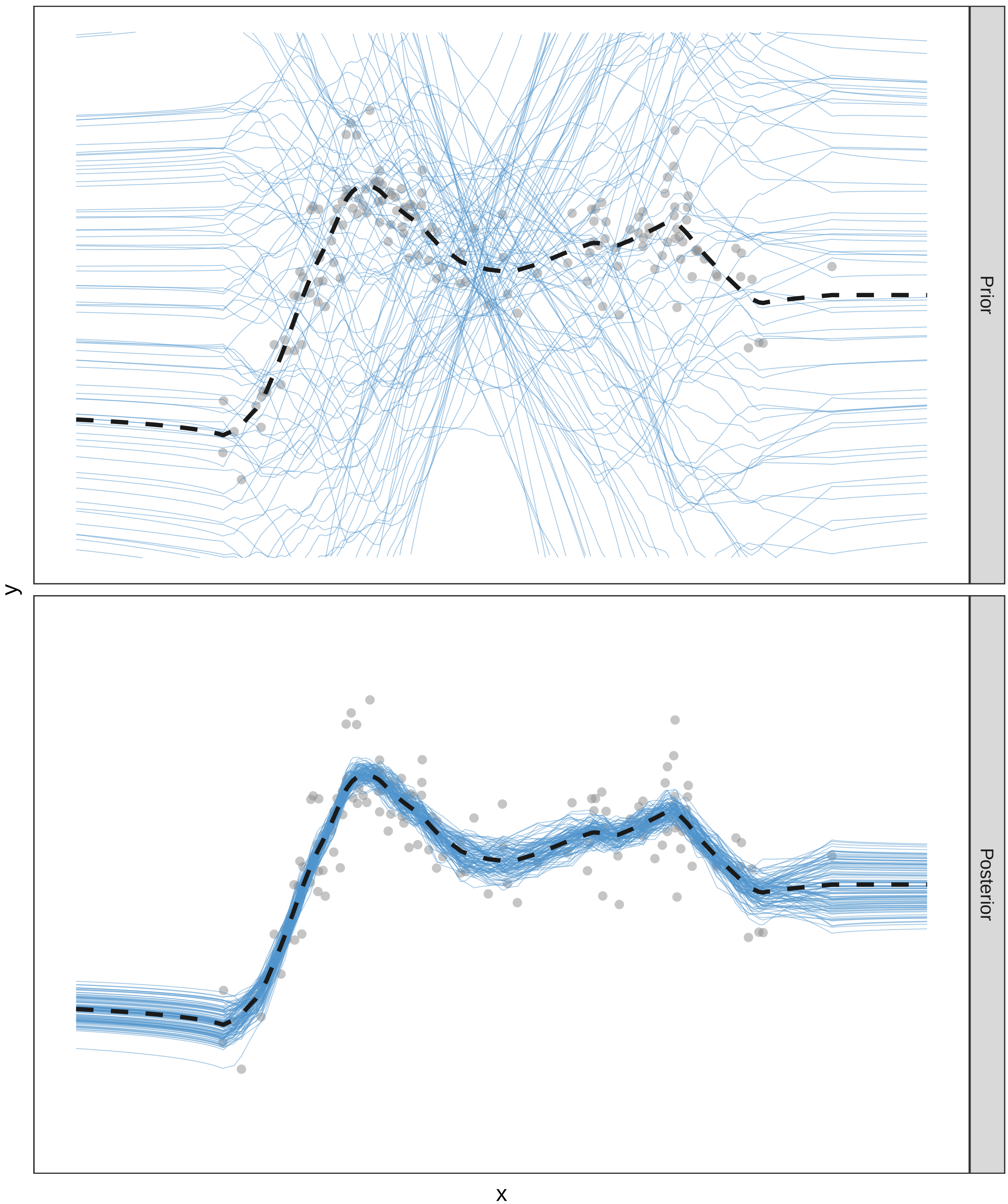


95% posterior predictive bands

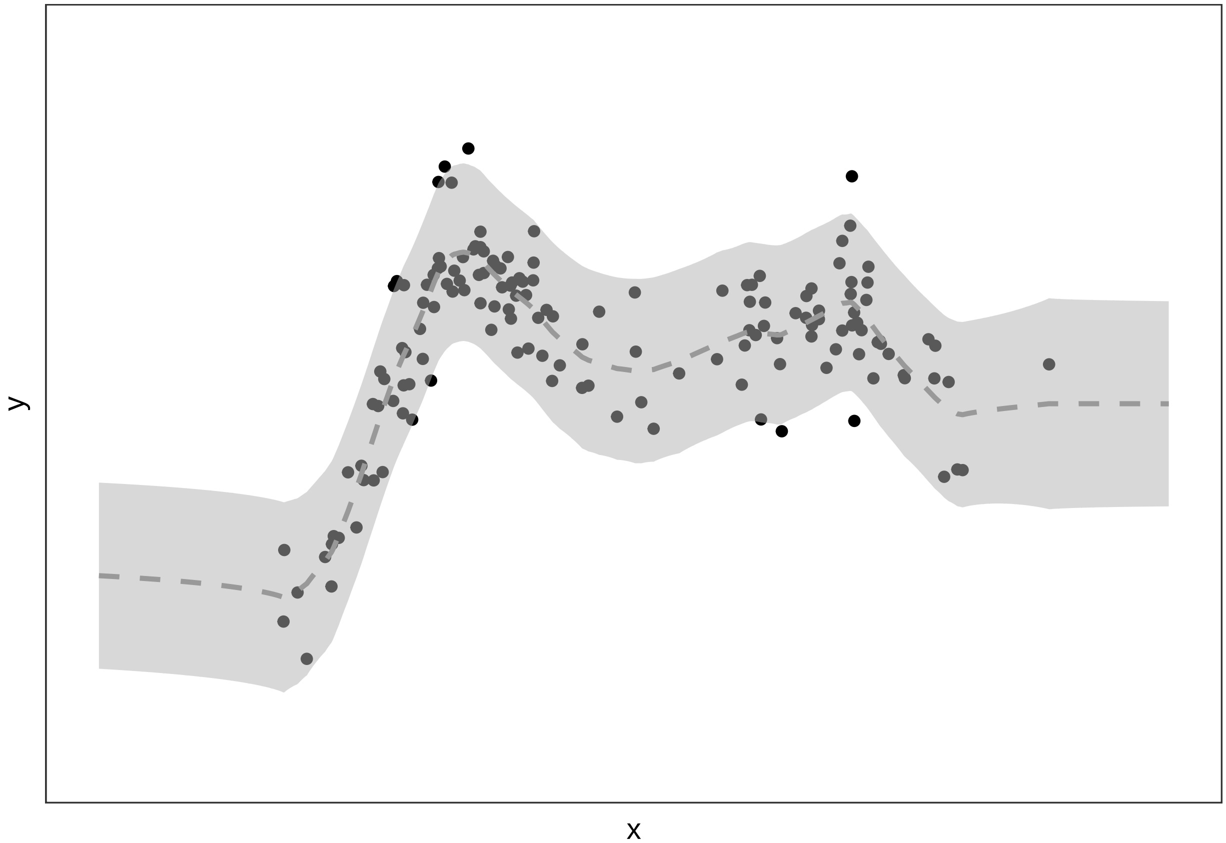


FBM kernel (MLE)

Prior and posterior draws

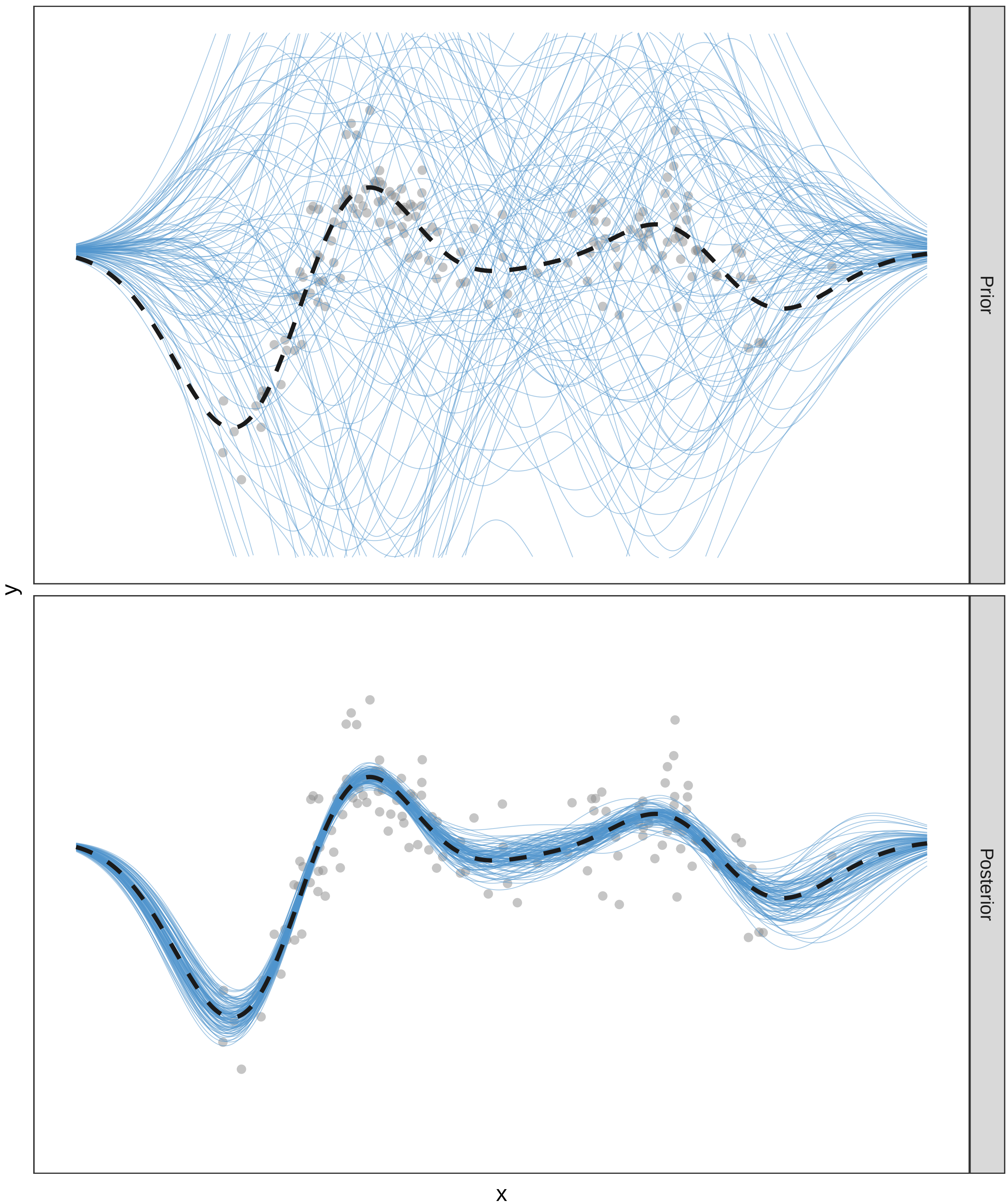


95% posterior predictive bands



SE kernel (MLE)

Prior and posterior draws



95% posterior predictive bands