

# Gaussian Process Models

Lecture 14

Dr. Colin Rundel

# Multivariate Normal

# Multivariate Normal Distribution

For an n-dimension multivariate normal distribution with covariance  $\Sigma$  (positive semidefinite) can be written as

$$\begin{matrix} \mathbf{y} \\ n \times 1 \end{matrix} \sim N\left(\begin{matrix} \boldsymbol{\mu} \\ n \times 1 \end{matrix}, \begin{matrix} \boldsymbol{\Sigma} \\ n \times n \end{matrix}\right)$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, \begin{pmatrix} \rho_{11}\sigma_1\sigma_1 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \cdots & \rho_{nn}\sigma_n\sigma_n \end{pmatrix}\right)$$

# Density

For the  $n$  dimensional multivariate normal given on the last slide, its density is given by

$$(2\pi)^{-n/2} \det(\Sigma)^{-1/2} \exp\left(-\frac{1}{2} \begin{matrix} \mathbf{y} - \boldsymbol{\mu} \\[1ex] 1 \times n \end{matrix}' \begin{matrix} \Sigma^{-1} \\[1ex] n \times n \end{matrix} \begin{matrix} \mathbf{y} - \boldsymbol{\mu} \\[1ex] n \times 1 \end{matrix}\right)$$

and its log density is given by

$$-\frac{n}{2} \log 2\pi - \frac{1}{2} \log \det(\Sigma) - \frac{1}{2} \begin{matrix} \mathbf{y} - \boldsymbol{\mu} \\[1ex] 1 \times n \end{matrix}' \begin{matrix} \Sigma^{-1} \\[1ex] n \times n \end{matrix} \begin{matrix} \mathbf{y} - \boldsymbol{\mu} \\[1ex] n \times 1 \end{matrix}$$

# Sampling

To generate draws from an  $n$ -dimensional multivariate normal with mean  $\mu$  and covariance matrix  $\Sigma$ ,

- Find a matrix  $A$  such that  $\Sigma = A A^t$ , most often we use  $A = \text{Chol}(\Sigma)$  where  $A$  is a lower triangular matrix.
- Draw  $n$  iid unit normals  $\begin{pmatrix} & (0, 1) \end{pmatrix}$  as  $z$
- Obtain multivariate normal draws using

$$\mathbf{y} = \boldsymbol{\mu} + A \mathbf{z}$$

# Bivariate Examples

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

# Marginal distributions

- For an n-dimensional multivariate normal with mean  $\mu$  and covariance matrix  $\Sigma$ , any marginal or conditional distribution of the y's will also be (multivariate) normal.

For a univariate marginal distribution,

$$y_i = (\mu_i, \Sigma_{ii})$$

For a bivariate marginal distribution,

$$y_{ij} = \left( \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix}, \begin{pmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ji} & \Sigma_{jj} \end{pmatrix} \right)$$

For a  $k$ -dimensional marginal distribution,

$$\mathbf{y}_{i,\dots,k} = \left( \begin{pmatrix} \boldsymbol{\mu}_i \\ \vdots \\ \boldsymbol{\mu}_k \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{ii} & \cdots & \boldsymbol{\Sigma}_{ik} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{ki} & \cdots & \boldsymbol{\Sigma}_{kk} \end{pmatrix} \right)$$

# Conditional Distributions

If we partition the  $n$ -dimensions into two pieces such that  $\mathbf{y} = (y_1, y_2)^t$  then

$$\underset{n \times 1}{\mathbf{y}} \sim \left( \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

$$\underset{k \times 1}{y_1} \sim \left( \underset{k \times 1}{\boldsymbol{\mu}_1}, \underset{k \times k}{\boldsymbol{\Sigma}_{11}} \right)$$

$$\underset{n-k \times 1}{y_2} \sim \left( \underset{n-k \times 1}{\boldsymbol{\mu}_2}, \underset{n-k \times n-k}{\boldsymbol{\Sigma}_{22}} \right)$$

then the conditional distributions are given by

$$y_1 \mid y_2 = \mathbf{a} \sim (\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{a} - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21})$$

$$y_2 \mid y_1 = \mathbf{b} \sim (\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{b} - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{21})$$

# Gaussian Processes

From Shumway,

A process,  $\mathbf{y} = \{y(t) : t \in T\}$ , is said to be a Gaussian process if all possible finite dimensional vectors  $\mathbf{y} = (y_{t_1}, y_{t_2}, \dots, y_{t_n})^t$ , for every collection of time points  $t_1, t_2, \dots, t_n$ , and every positive integer  $n$ , have a multivariate normal distribution.

So far we have only looked at examples of time series where  $T$  is discrete (and evenly spaces & contiguous), it turns out things get a lot more interesting when we explore the case where  $T$  is defined on a *continuous* space (e.g.  $\mathbb{R}$  or some subset of  $\mathbb{R}$ ).

# Gaussian Process Regression

# Parameterizing a Gaussian Process

Imagine we have a Gaussian process defined such that

$$\mathbf{y} = \{y(t) : t \in [0, 1]\},$$

- We now have an uncountably infinite set of possible  $t$ 's and  $y(t)$ s.
- We will only have a (small) finite number of observations  $y(t_1), \dots, y(t_n)$  with which to say something useful about this infinite dimensional process.
- The unconstrained covariance matrix for the observed data can have up to  $n(n + 1)/2$  unique values\*
- Necessary to make some simplifying assumptions:
  - Stationarity
  - Simple parameterization of  $\Sigma$

# Covariance Functions

More on these next week, but for now some common examples

Exponential covariance:

$$\Sigma(y_t, y_{t'}) = \sigma^2 \exp(-|t - t'|)$$

Squared exponential covariance (Gaussian):

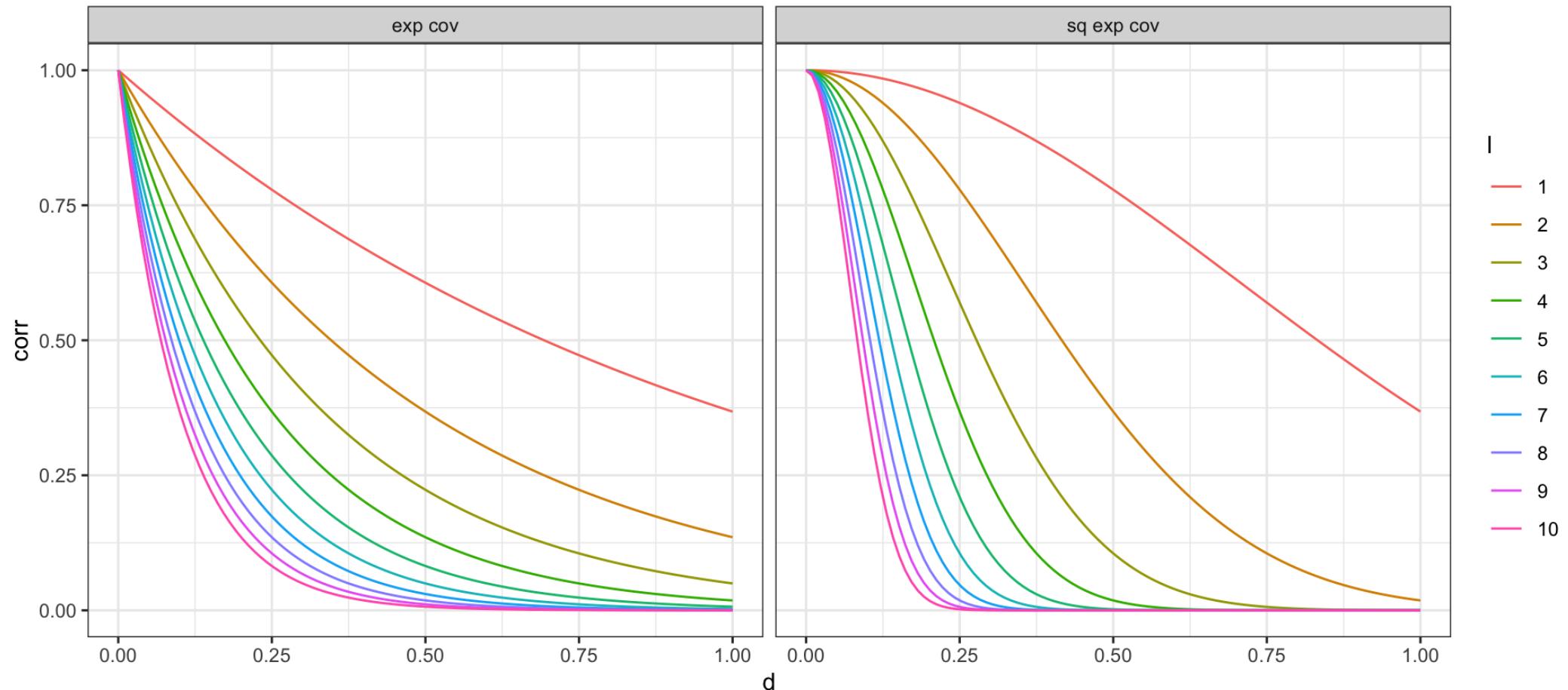
$$\Sigma(y_t, y_{t'}) = \sigma^2 \exp(-(|t - t'|)^2)$$

Powered exponential covariance ( $p \in (0, 2]$ ):

$$\Sigma(y_t, y_{t'}) = \sigma^2 \exp(-(|t - t'|)^p)$$

# Covariance Function - Correlation Decay

Letting  $\sigma^2 = 1$  and trying different values of the length scale  $l$ ,

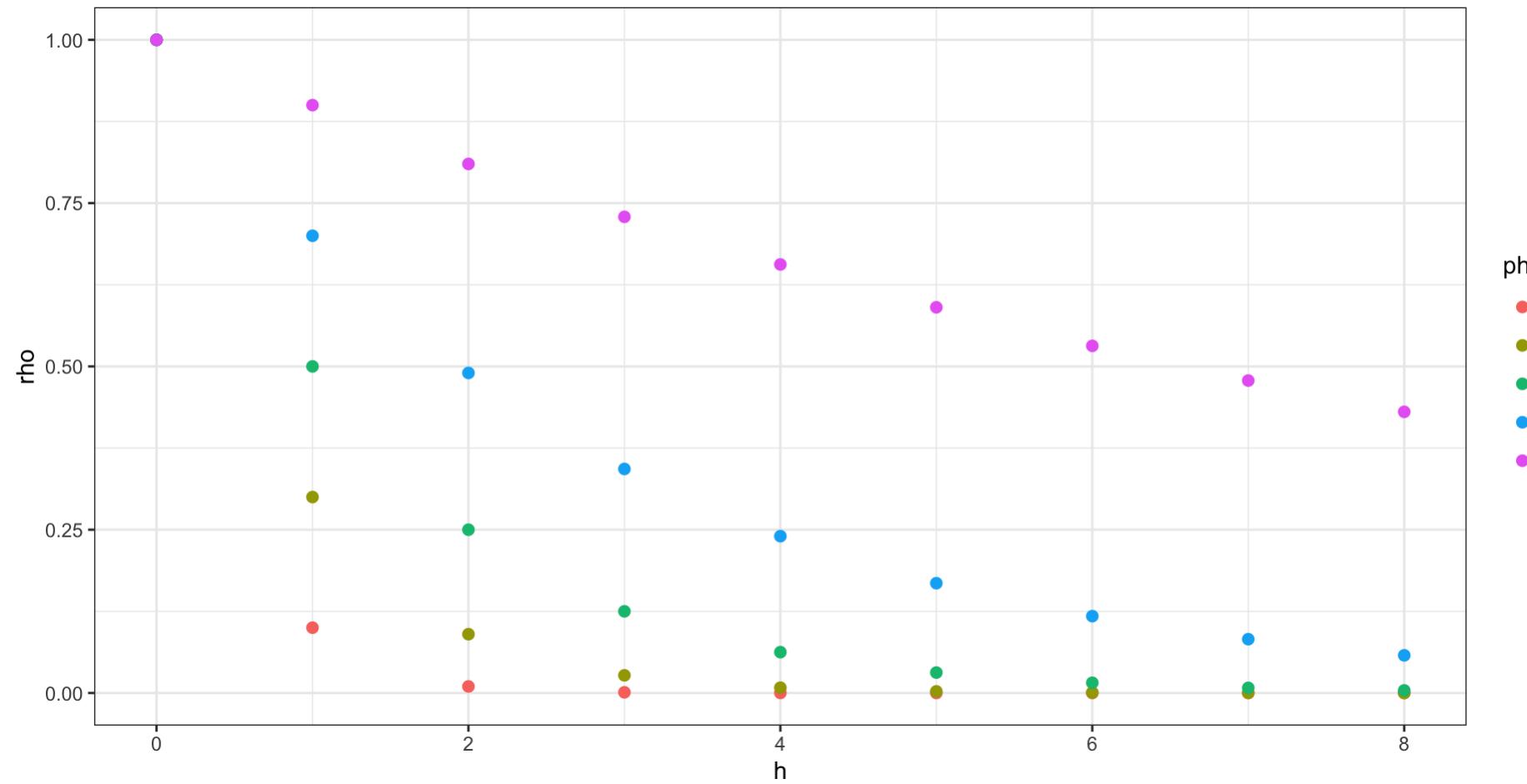


# Correlation Decay - AR(1)

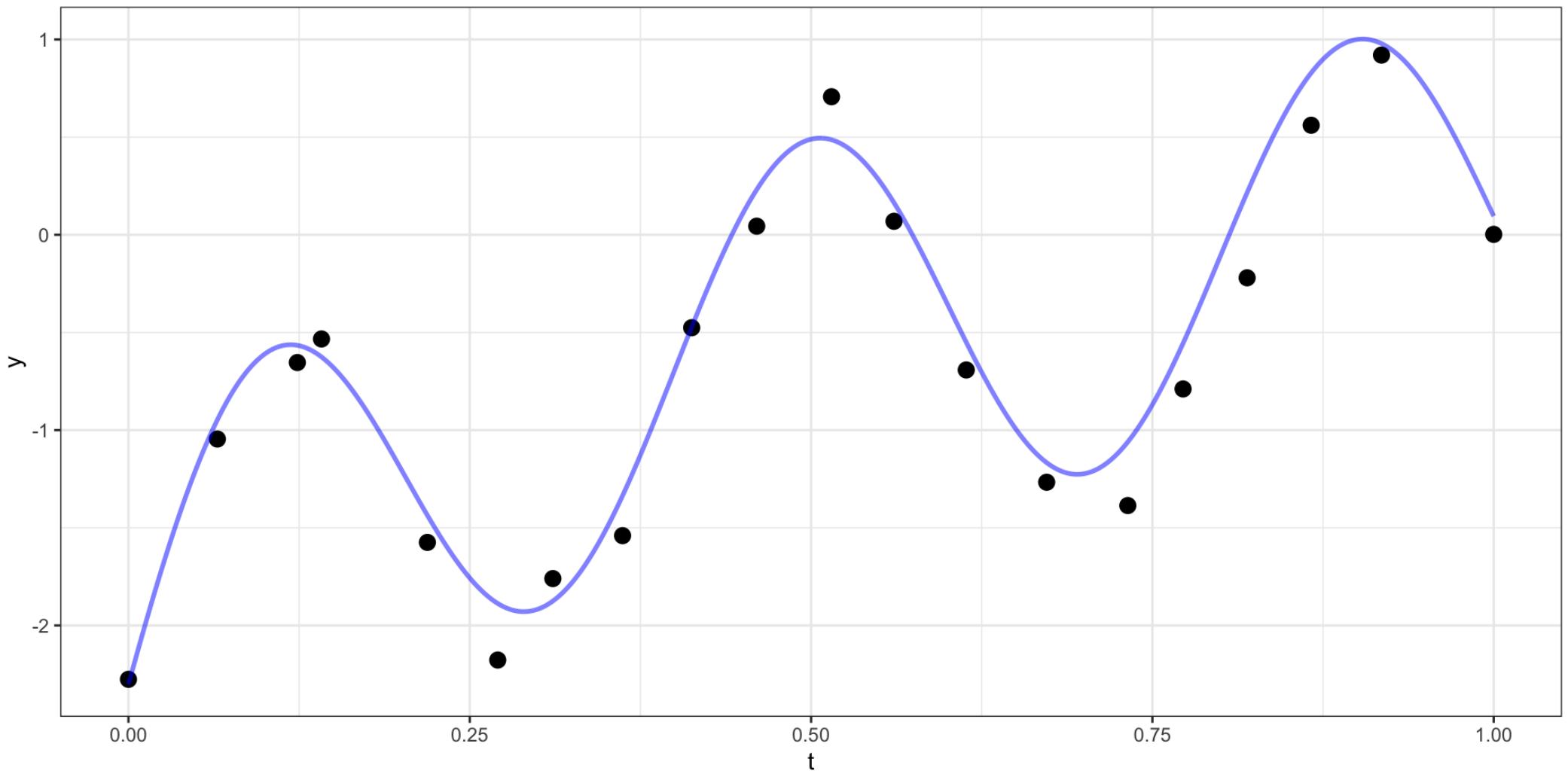
Recall that for a stationary AR(1) process:

$$\gamma(h) = \sigma_w^2 \phi^{|h|} \text{ and } \rho(h) = \phi^{|h|}$$

we can draw a somewhat similar picture about the decay of  $\rho$  as a function of distance.



# Example



# Prediction

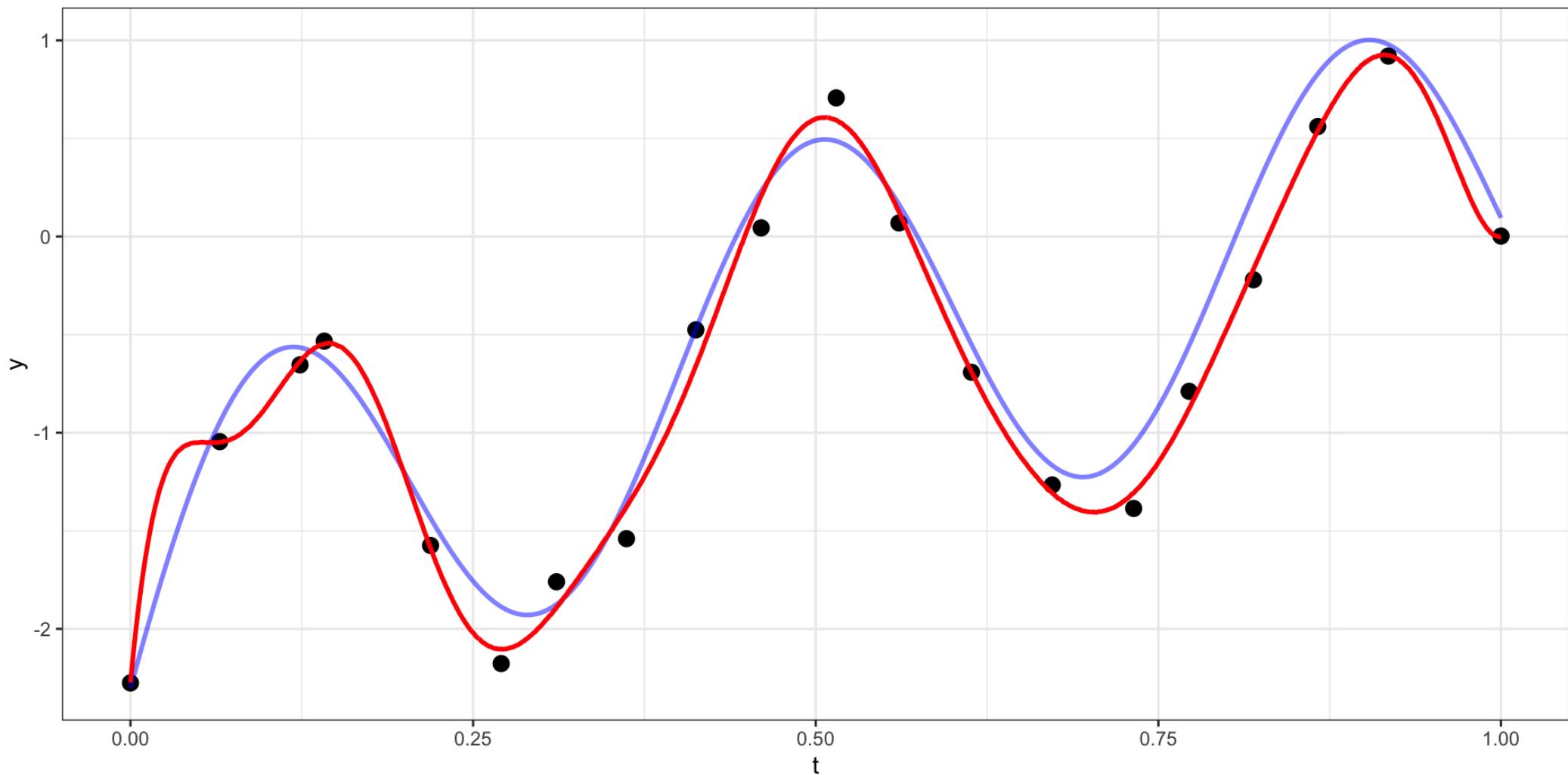
Our example has 20 observations which we would like to use as the basis for predicting  $y(t)$  at other values of  $t$  (say a sequence of values from 0 to 1).

For now lets use a square exponential covariance with  $\sigma^2 = 10$  and  $l = 5$

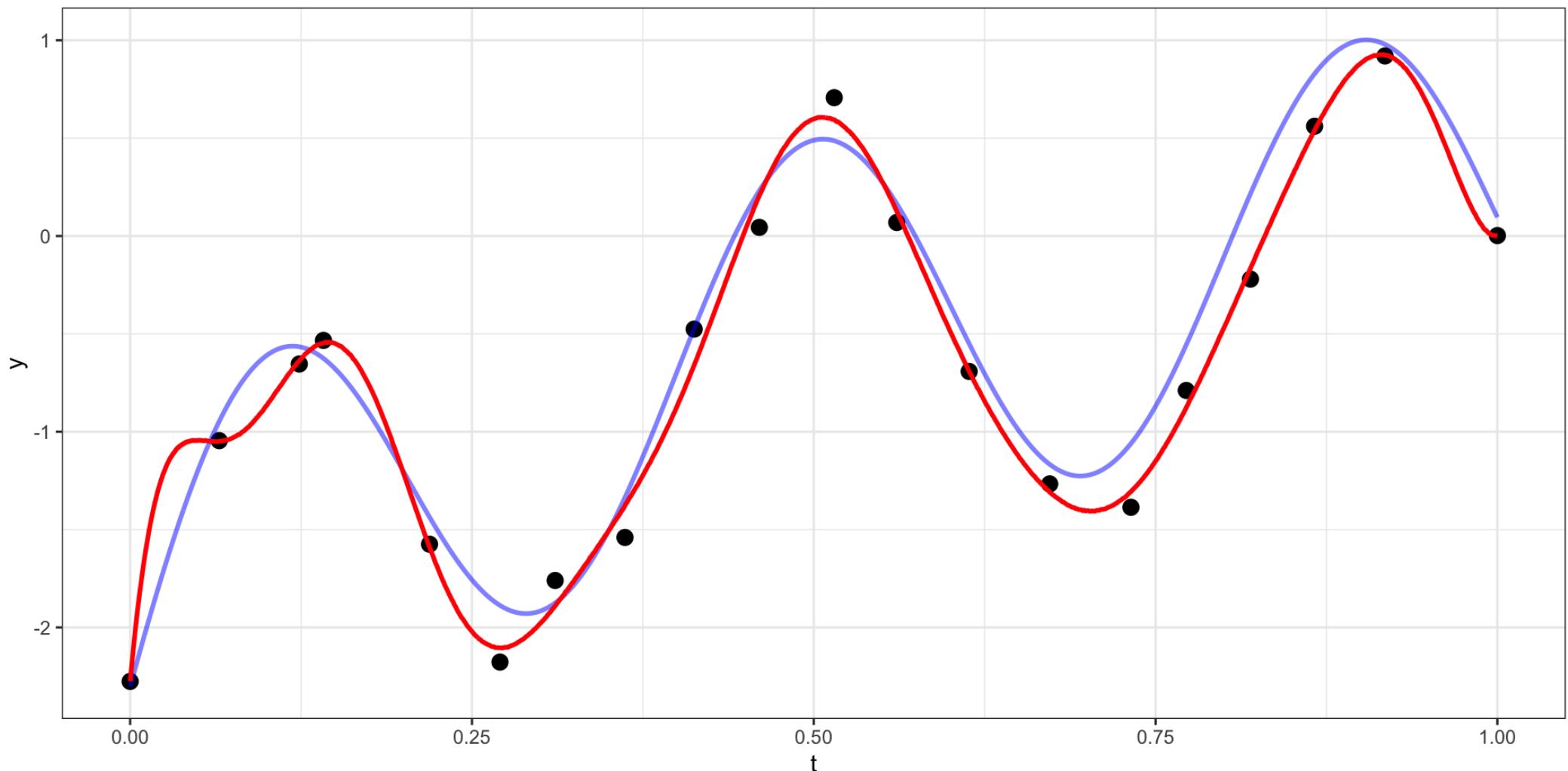
We therefore want to sample from  $y_{\text{pred}} | y_{\text{obs}}$ .

$$y_{\text{pred}} | y_{\text{obs}} = y \sim (\Sigma_{\text{po}} \Sigma_{\text{obs}}^{-1} y, \Sigma_{\text{pred}} - \Sigma_{\text{po}} \Sigma_{\text{pred}}^{-1} \Sigma_{\text{op}})$$

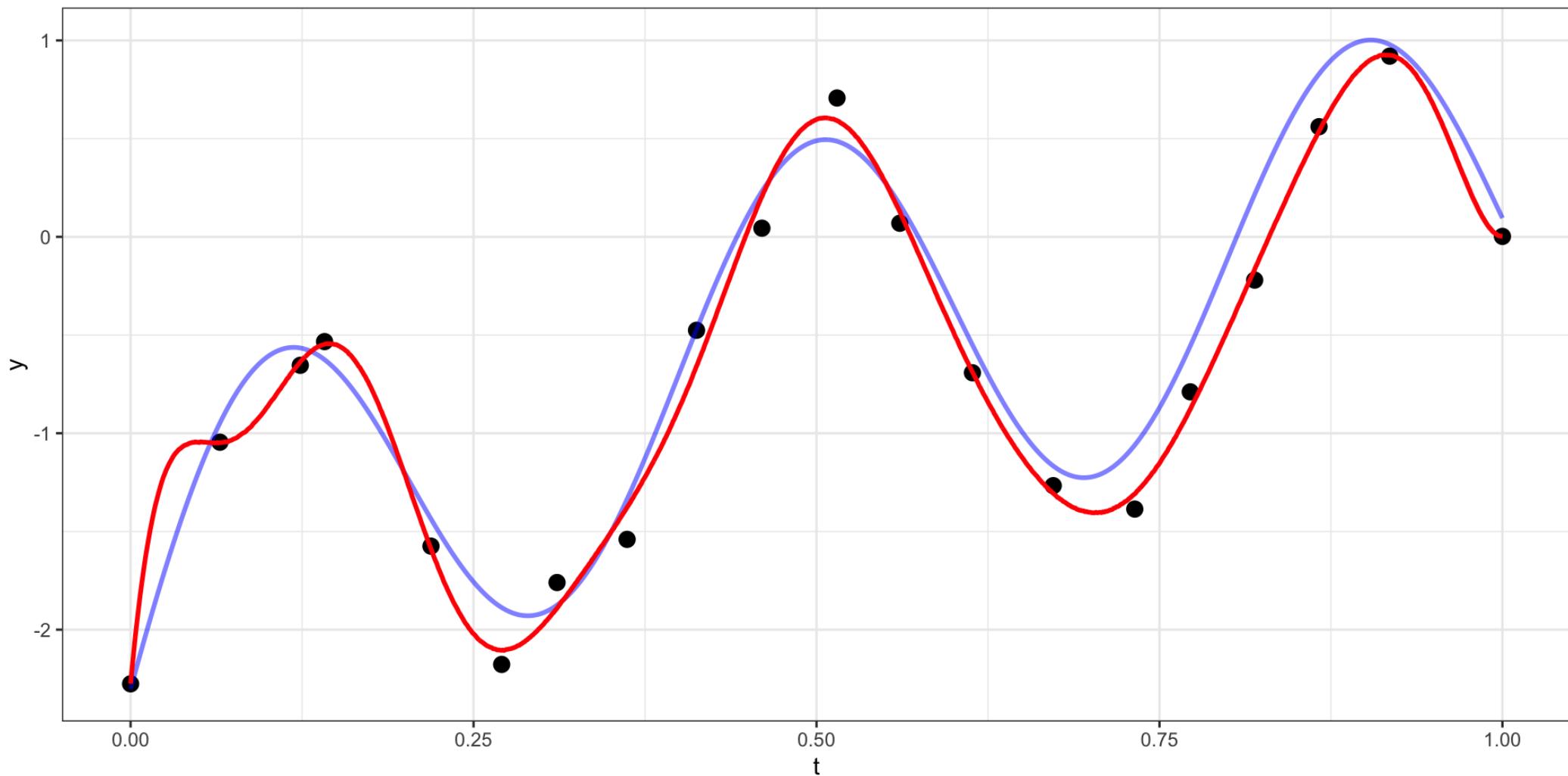
# Draw 1



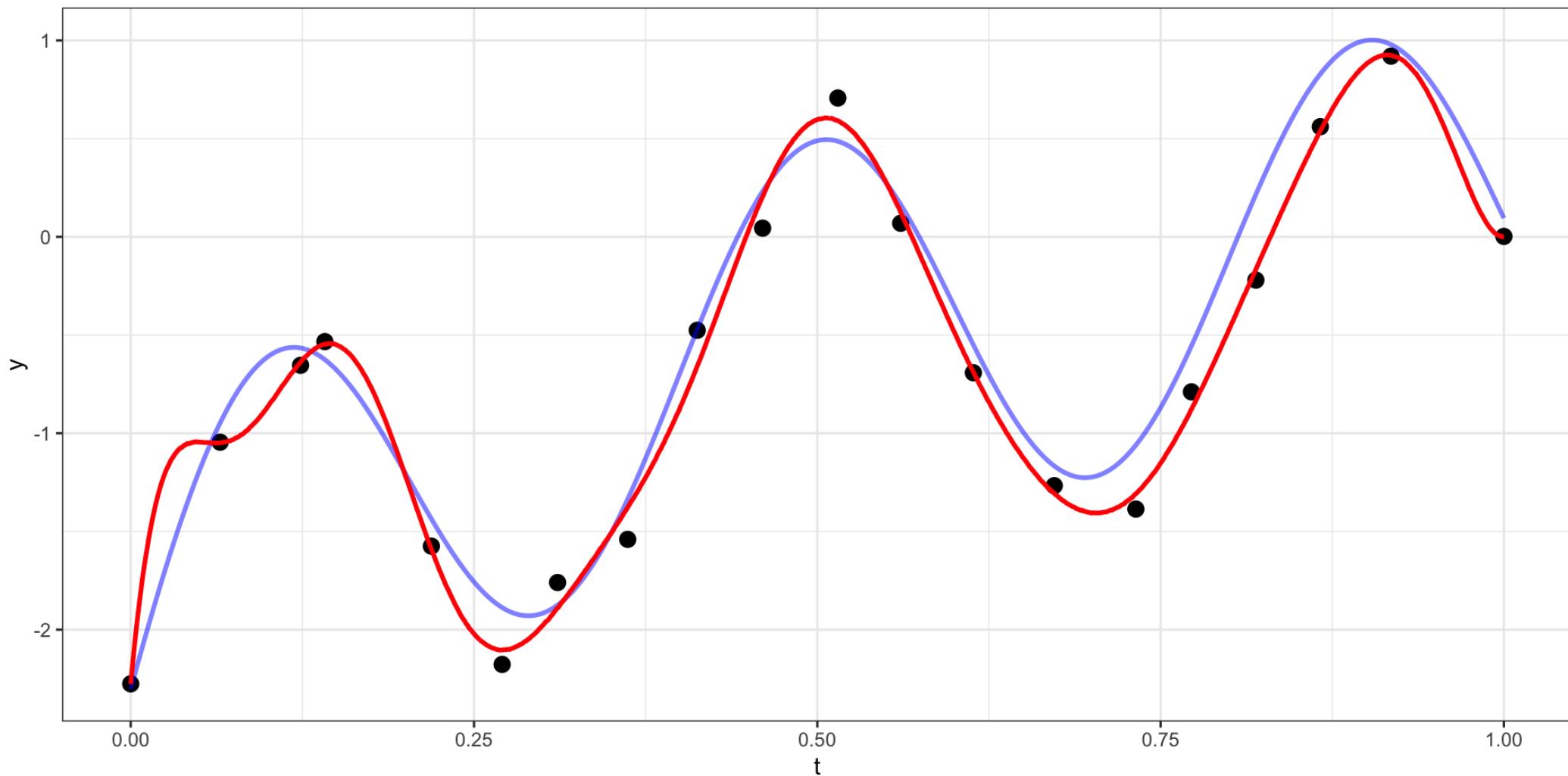
## Draw 2



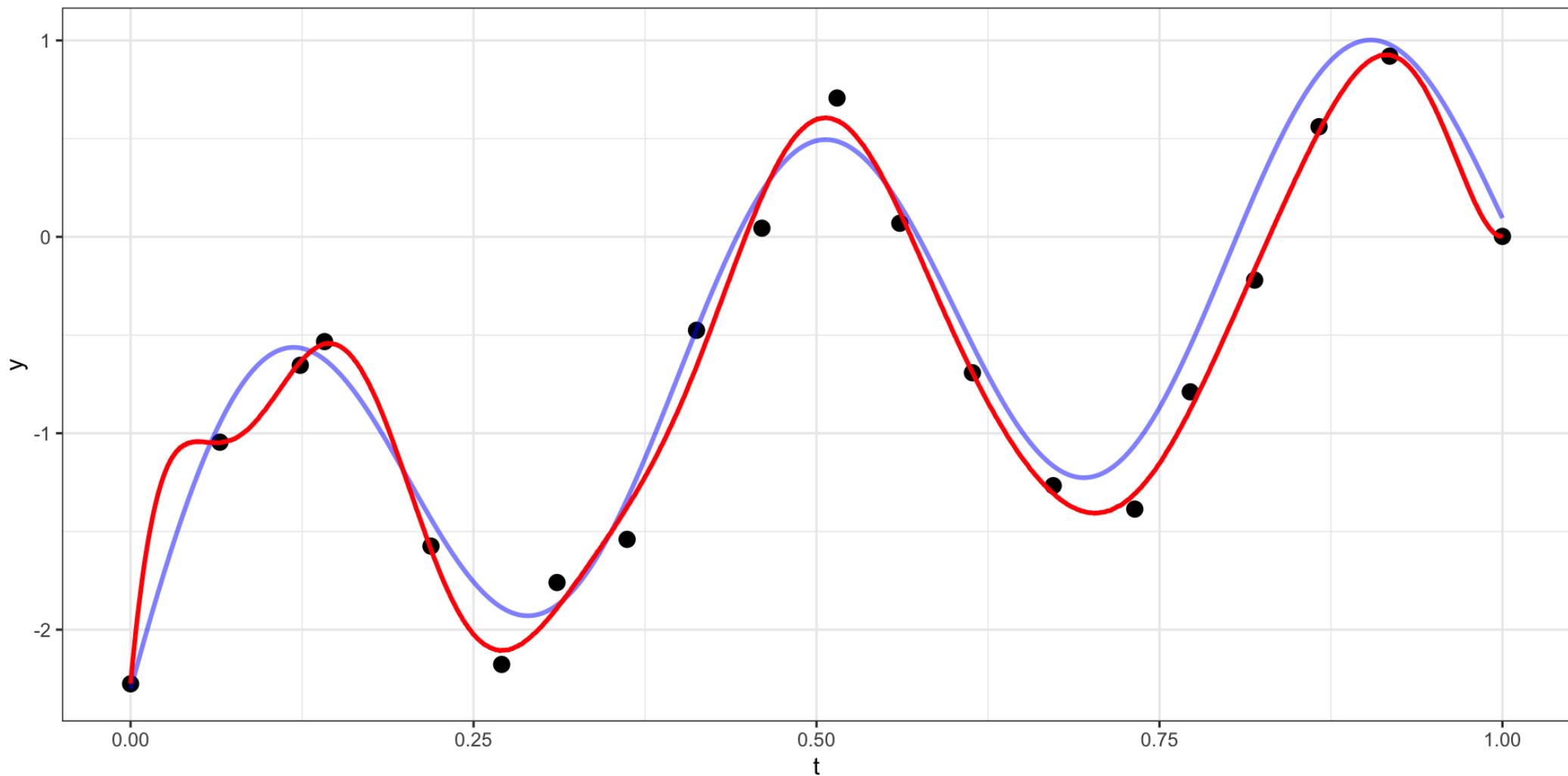
# Draw 3



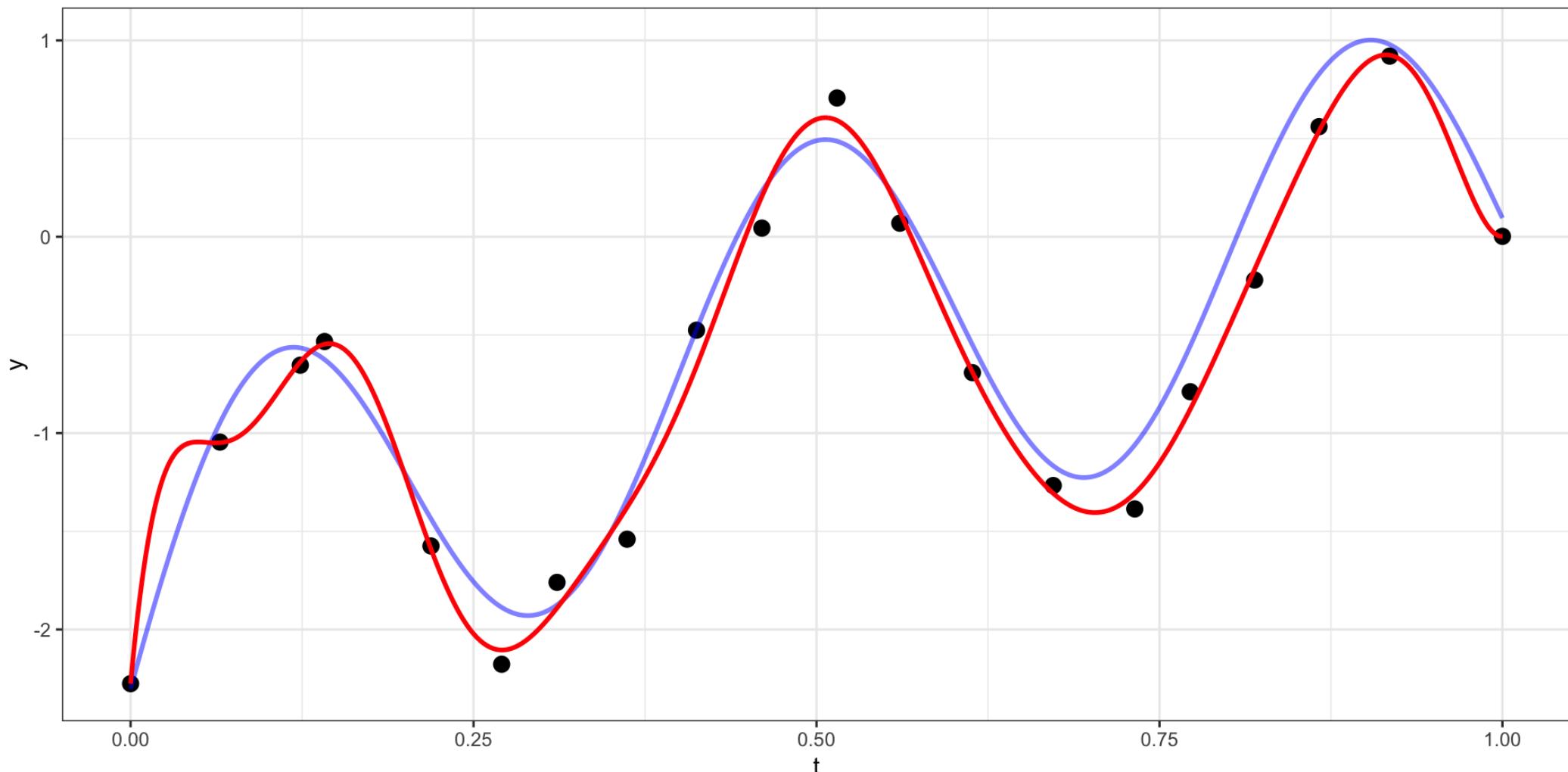
# Draw 4



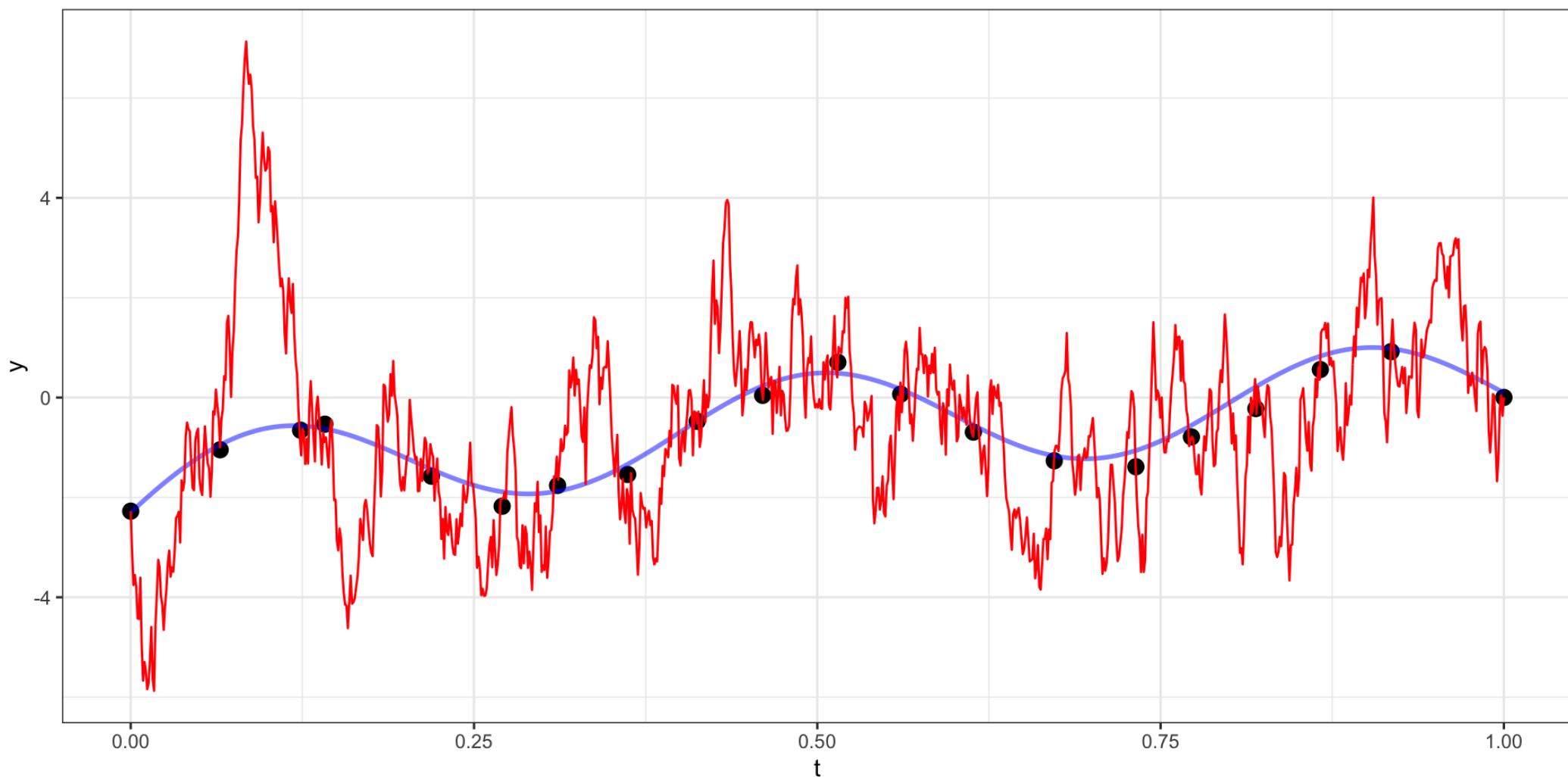
# Draw 5



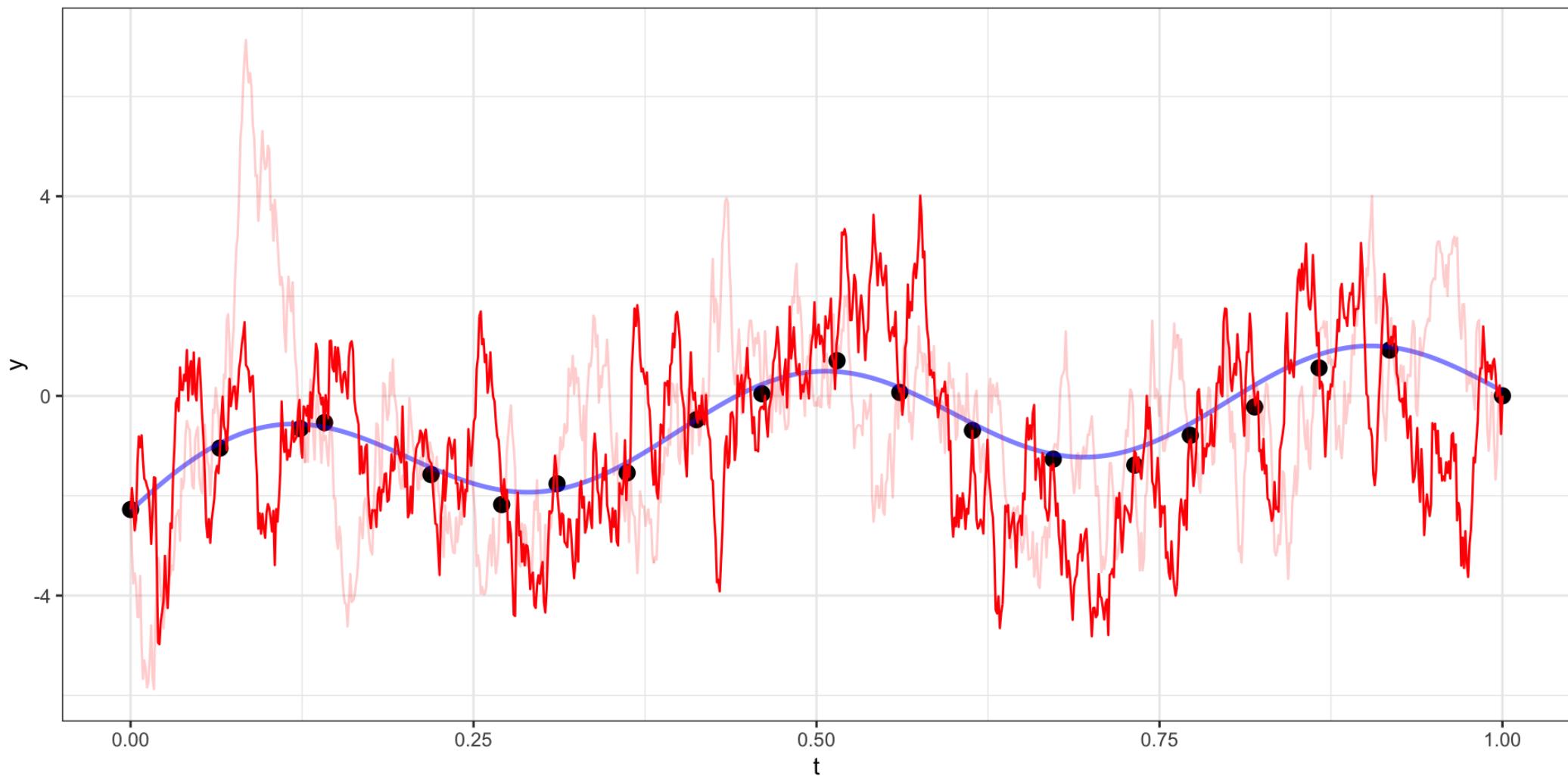
# Many draws later



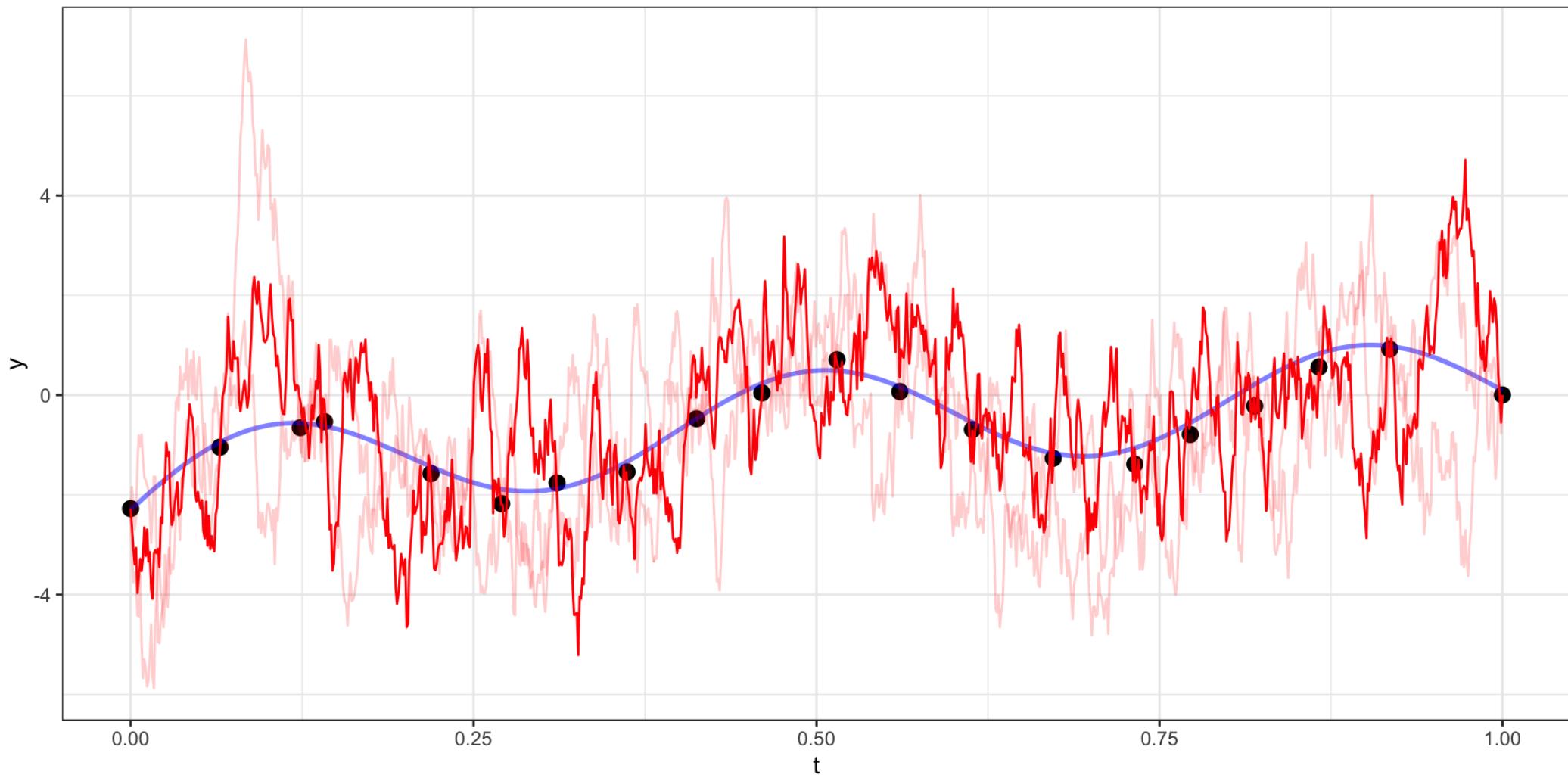
# Exponential Covariance



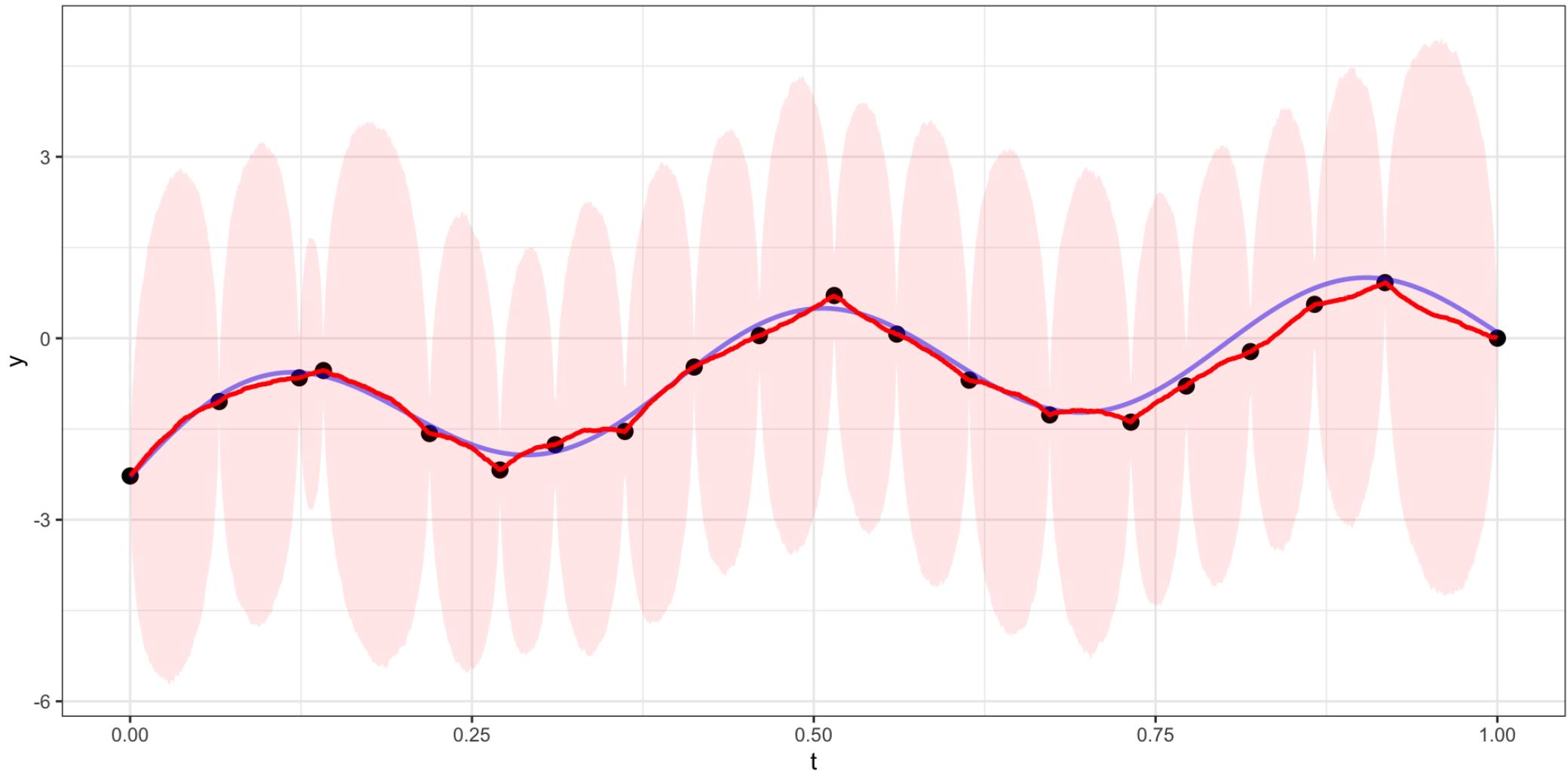
# Exponential Covariance - Draw 2



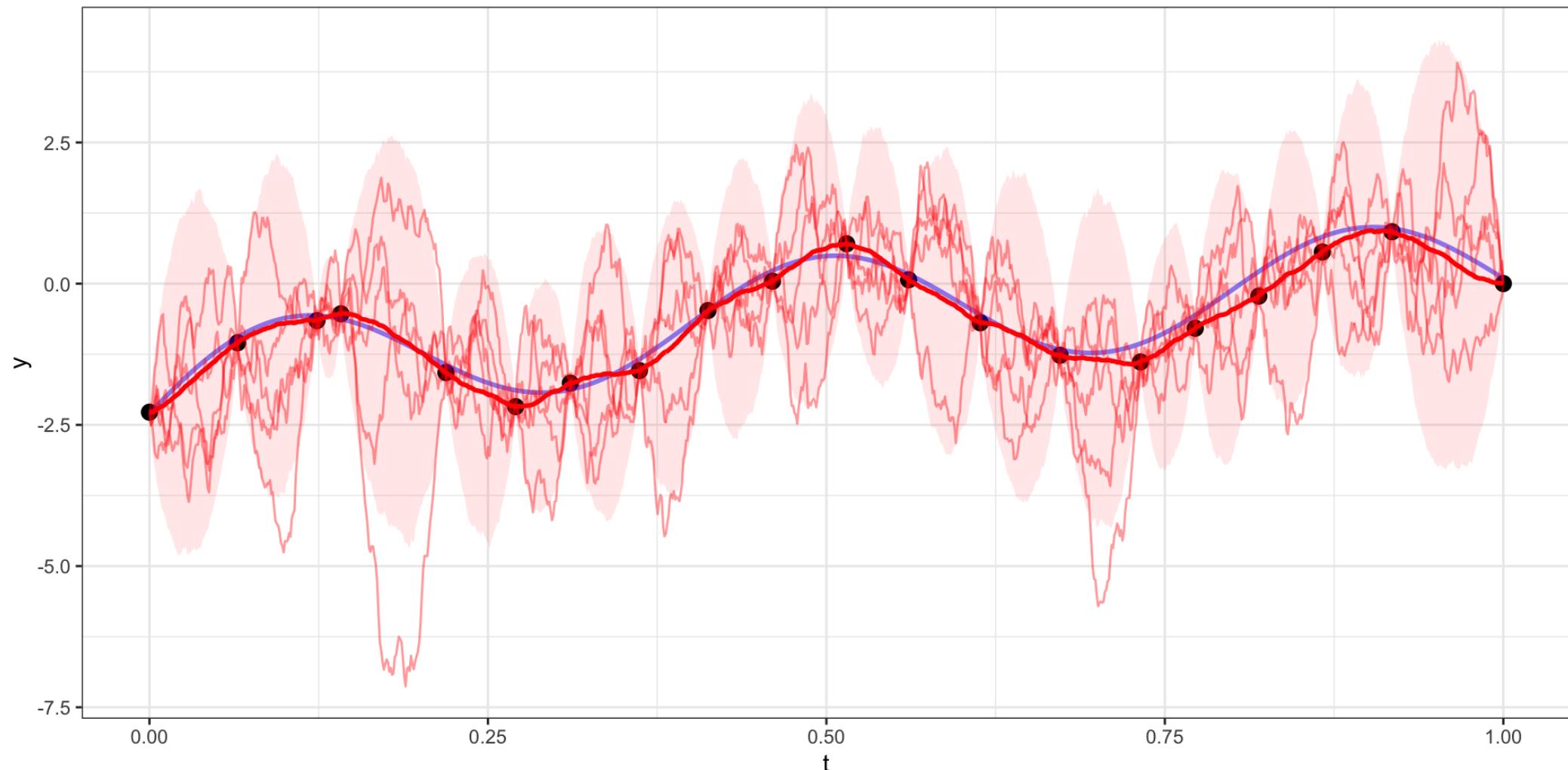
# Exponential Covariance - Draw 3



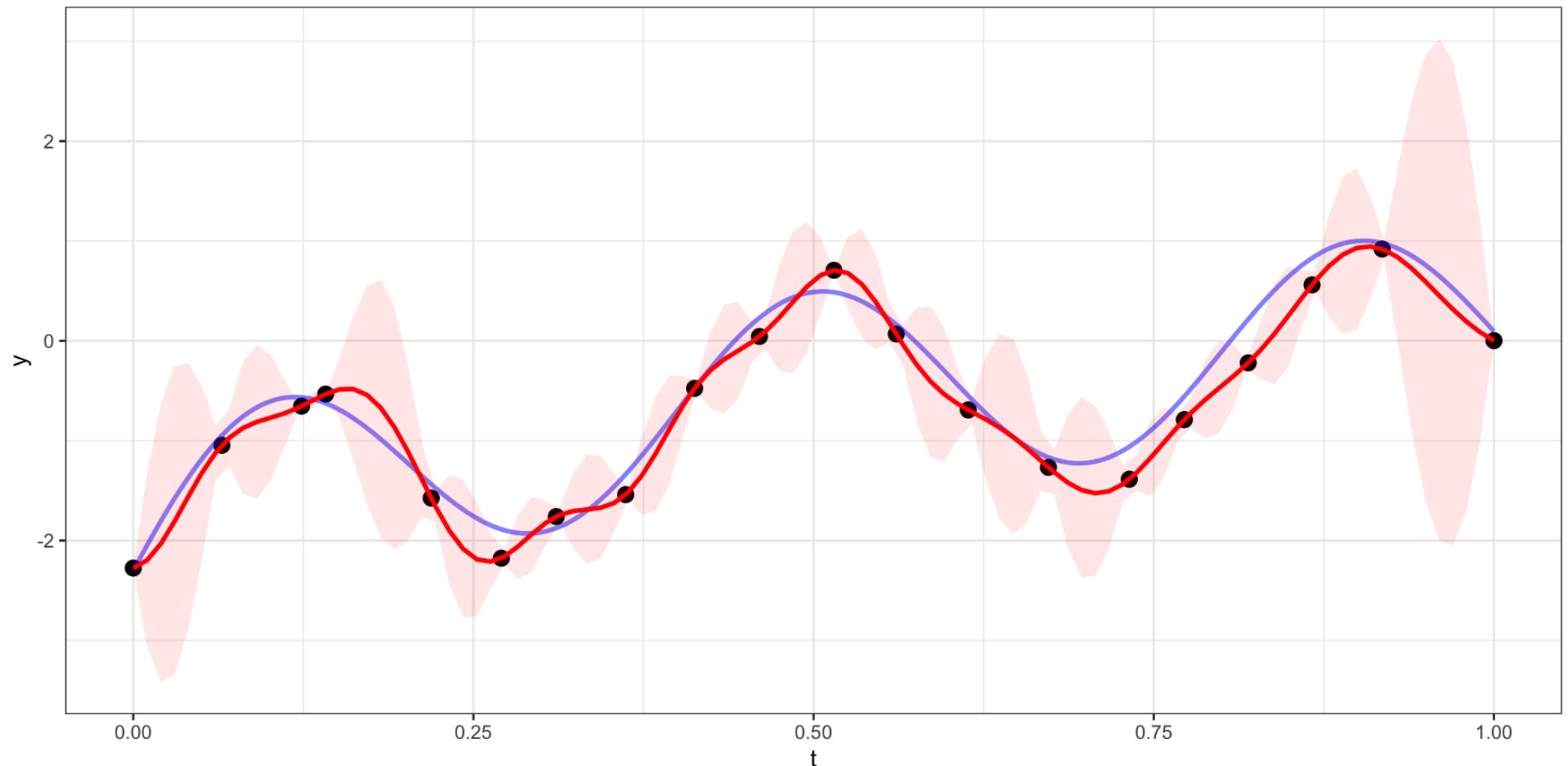
# Exponential Covariance - Posterior



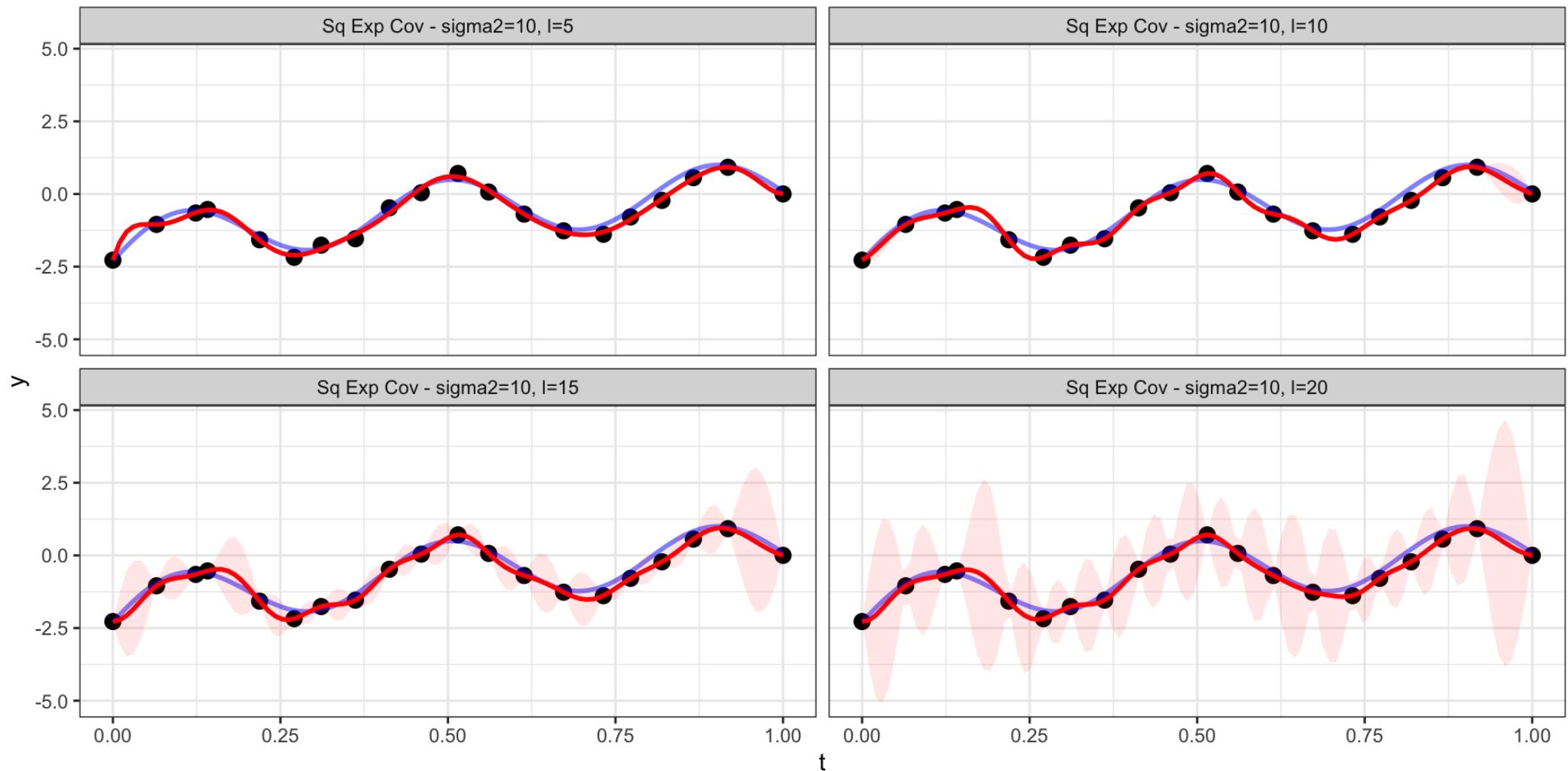
# Powered Exponential Covariance ( $p = 1.5$ )



# Back to the square exponential



# Changing the range (1)



# Effective Range

For the square exponential covariance

$$\text{Cov}(d) = \sigma^2 \exp(-(l \cdot d)^2)$$

$$\text{Corr}(d) = \exp(-(l \cdot d)^2)$$

we would like to know, for a given value of  $l$ , beyond what distance apart must observations be to have a correlation less than 0.05?

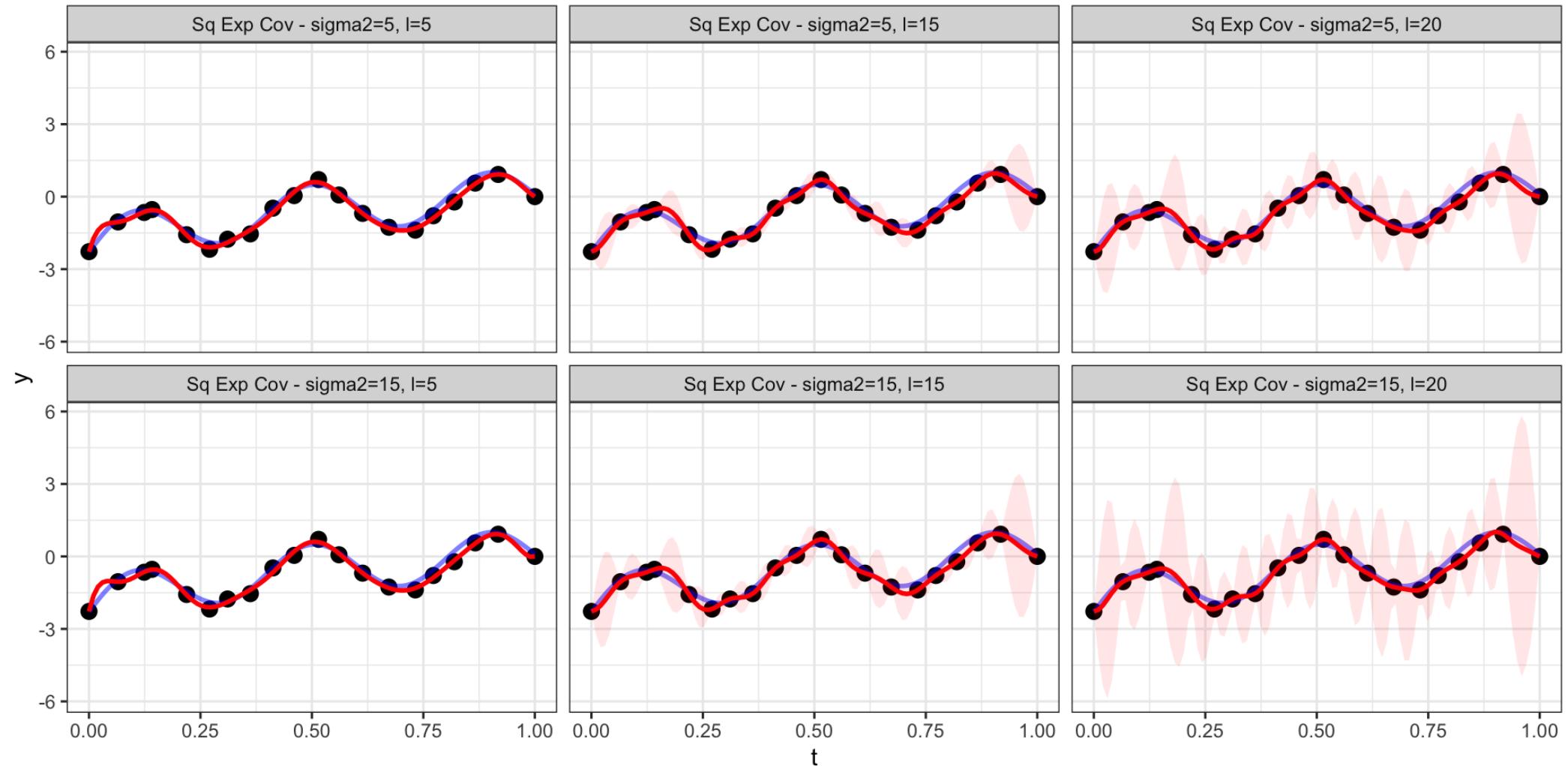
$$\exp(-(l \cdot d)^2) < 0.05$$

$$-(l \cdot d)^2 < \log 0.05$$

$$l \cdot d < \sqrt{3}$$

$$d < \sqrt{3}/l$$

# Changing the scale ( $\sigma^2$ )



# Fitting w/ BRMS

```
1 library(brms)
2 gp = brm(y ~ gp(t), data=d, cores=4, refresh=0)
```

```
1 summary(gp)
```

Family: gaussian  
Links: mu = identity; sigma = identity  
Formula: y ~ gp(t)  
Data: d (Number of observations: 20)  
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
total post-warmup draws = 4000

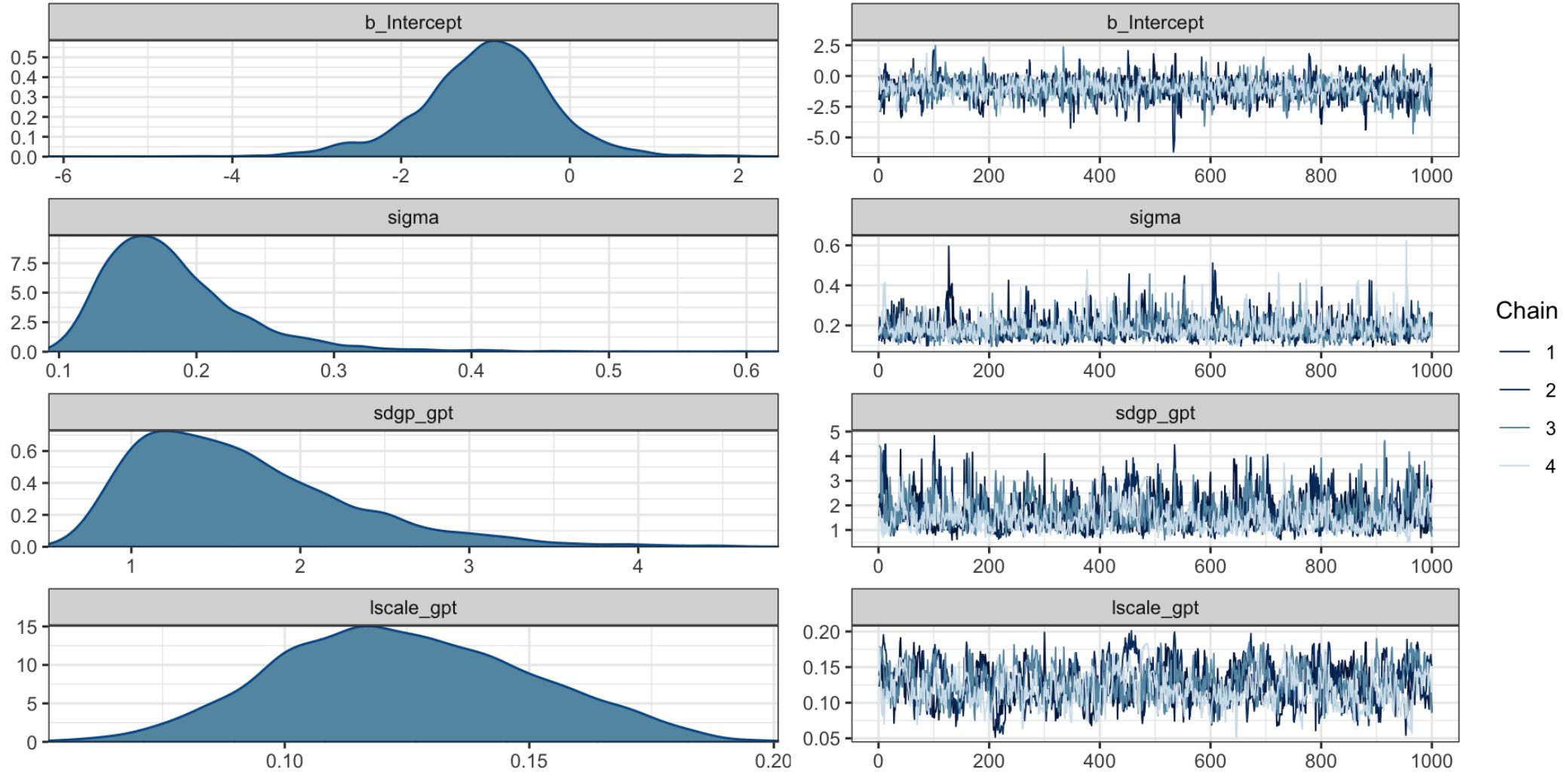
Gaussian Process Terms:

	Estimate	Est.Error	l-95%	CI	u-95%	CI
sdgp(gpt)	1.66	0.64	0.80		3.24	
lscale(gpt)	0.12	0.03	0.08		0.18	
	Rhat	Bulk_ESS	Tail_ESS			
sdgp(gpt)	1.02	393	1156			
lscale(gpt)	1.02	294	700			

Population-Level Effects:

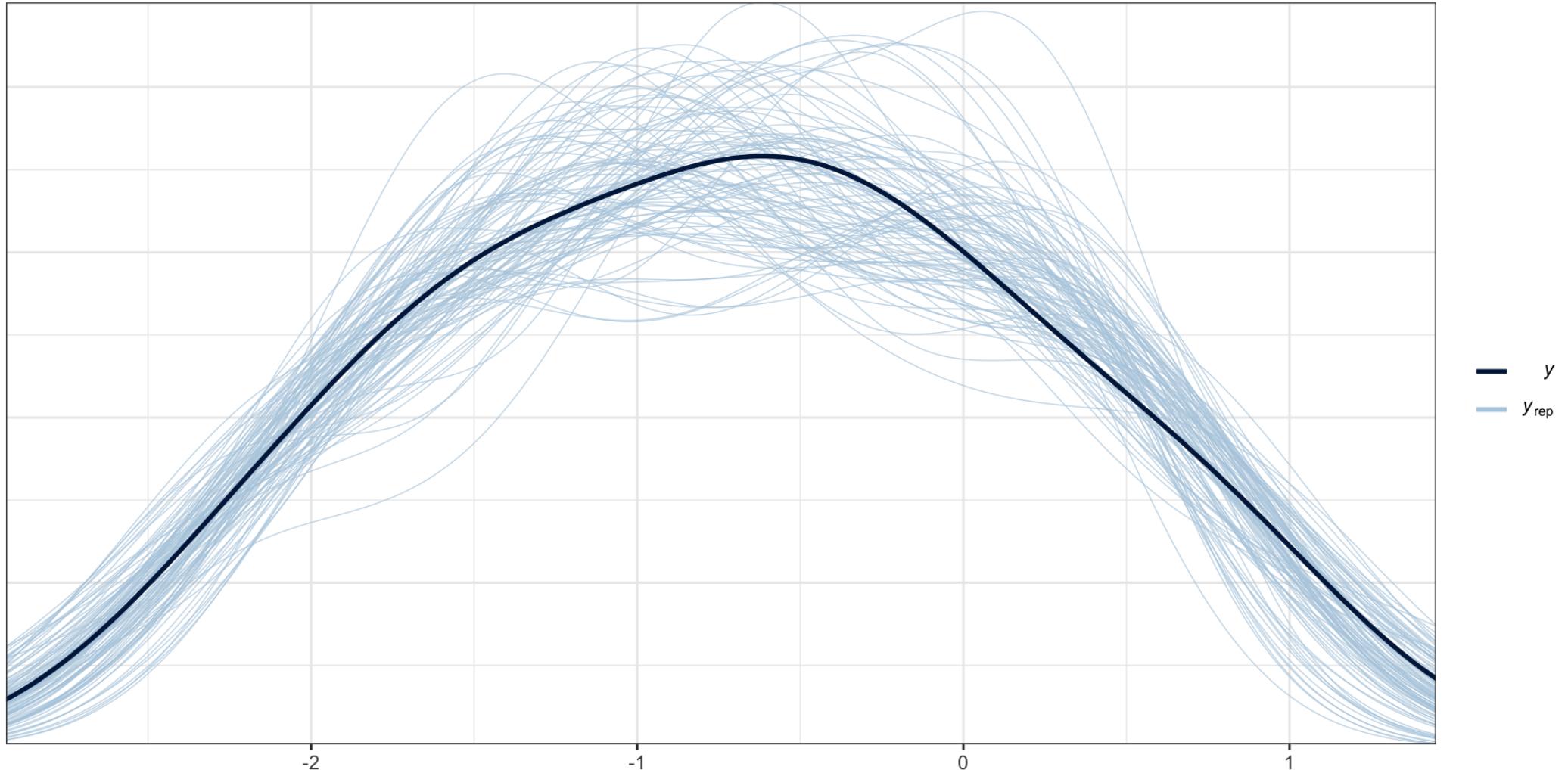
# Trace plots

```
1 plot(gp)
```

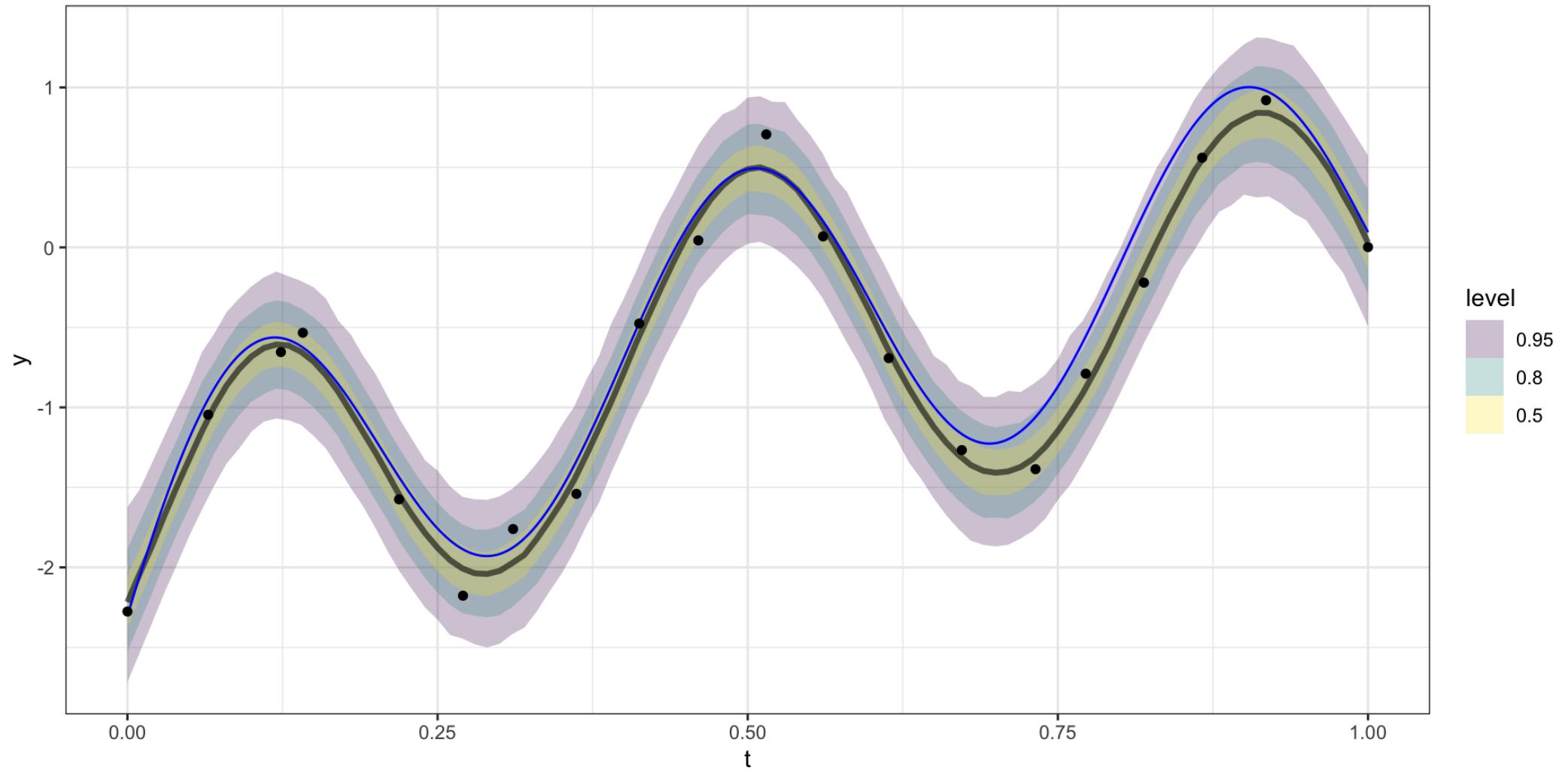


# PP Checks

```
1 pp_check(gp, ndraws = 100)
```



# Model predictions



# Forecasting

