Lecture 12

Gaussian Process Models

Colin Rundel

02/27/2017

Multivariate Normal

$$y(t) = \mu(t) + w(t) + v$$

$w(t) \hookrightarrow GP$

Multivariate Normal Distribution

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

$$\underset{n\times 1}{Y} \sim N(\underset{n\times 1}{\boldsymbol{\mu}}, \underset{n\times n}{\boldsymbol{\Sigma}}) \text{ where } \{\boldsymbol{\Sigma}\}_{ij} = \sigma^2_{ij} = \rho_{ij}\,\sigma_i\,\sigma_j$$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, \begin{pmatrix} \rho_{11}\sigma_1\sigma_1 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \cdots & \rho_{nn}\sigma_n\sigma_n \end{pmatrix}\right)$$

Density

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

O(n^3)

$$(2\pi)^{-n/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}(\underset{1\times n}{Y-\boldsymbol{\mu}})' \underset{n\times n}{\boldsymbol{\Sigma}^{-1}} (\underset{n\times 1}{Y-\boldsymbol{\mu}})\right)$$

and its log density is given by

$$-\frac{n}{2}\log 2\pi - \frac{1}{2}\log\det(\boldsymbol{\Sigma}) - -\frac{1}{2}(\underset{1\times n}{Y-\boldsymbol{\mu}})' \underset{n\times n}{\boldsymbol{\Sigma}^{-1}} (\underset{n\times 1}{Y-\boldsymbol{\mu}})$$

n ≈ 2 - 4000

→ n ≈ 5 - 10000

Sampling

To generate draws from an n-dimensional multivate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$,

Sampling

To generate draws from an n-dimensional multivate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$,

- Find a matrix A such that $\boldsymbol{\Sigma} = A\,A^t$, most often we use $A = \text{Chol}(\boldsymbol{\Sigma})$

Sampling

To generate draws from an n-dimensional multivate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$,

- Find a matrix A such that $\boldsymbol{\Sigma} = A\,A^t$, most often we use $A = \text{Chol}(\boldsymbol{\Sigma})$
- Draw n iid unit normals ($\mathcal{N}(0, 1)$) as $\boldsymbol{z}$

Sampling

To generate draws from an n-dimensional multivate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$,

$O(n^3)$

- Find a matrix A such that $\boldsymbol{\Sigma} = A\,A^t$, most often we use $A = \text{Chol}(\boldsymbol{\Sigma})$

- Draw n iid unit normals ($\mathcal{N}(0,1)$) as $\boldsymbol{z}$

- Construct multivariate normal draws using

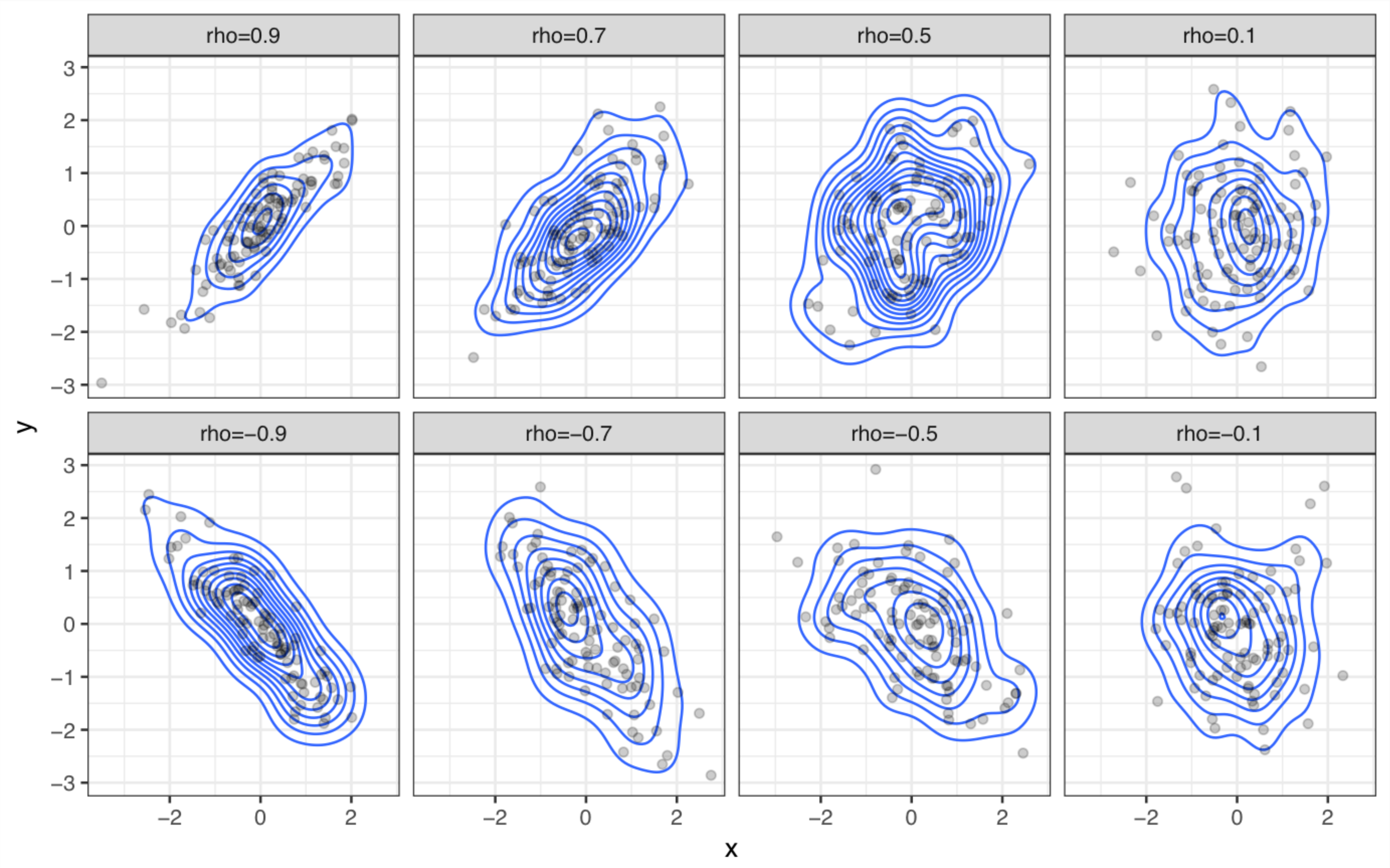
$$Y = \boldsymbol{\mu} + A\,\boldsymbol{z}$$

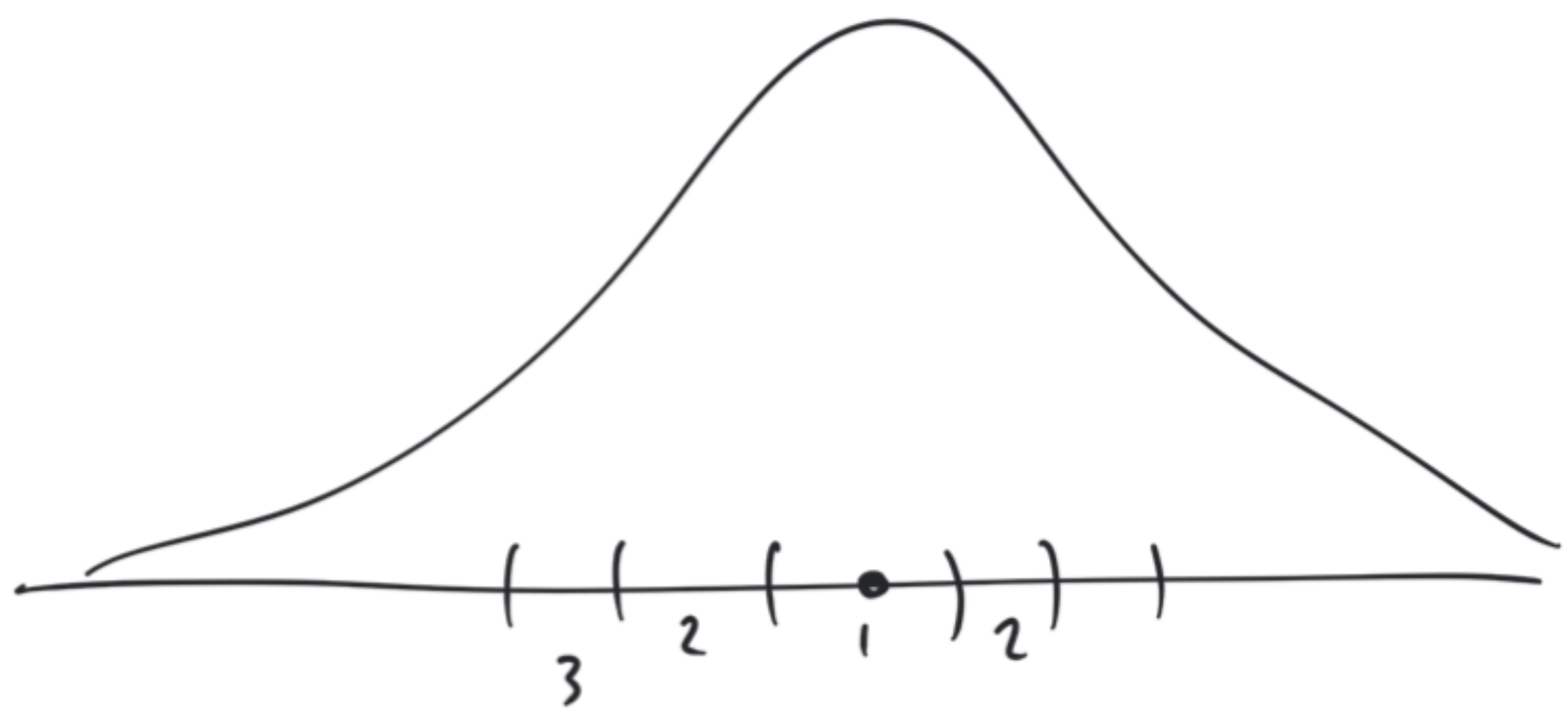
$$E(Y) = E(\mu) + A\,E(z) = \mu$$

$$Var(Y) = Var(\mu) + Var(Az) = A\,Var(z)\,A^t = \Sigma$$

Bivariate Example

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




3
2
1
2

Marginal distributions

Proposition - For an n-dimensional multivate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, any of the possible marginal distributions will also (multivariate) normal.

Marginal distributions

Proposition - For an n-dimensional multivate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, any of the possible marginal distributions will also (multivariate) normal.

For a univariate marginal distribution,

$$y_i = \mathcal{N}(\mu_i,\, \gamma_{ii})$$

Marginal distributions

Proposition - For an n-dimensional multivate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, any of the possible marginal distributions will also (multivariate) normal.

For a univariate marginal distribution,

$$y_i = \mathcal{N}(\mu_i, \gamma_{ii})$$

For a bivariate marginal distribution,

$$\boldsymbol{y}_{ij} = \mathcal{N}\left(\begin{pmatrix}\mu_i \\ \mu_j\end{pmatrix}, \begin{pmatrix}\gamma_{ii} & \gamma_{ij} \\ \gamma_{ji} & \gamma_{jj}\end{pmatrix}\right)$$

Marginal distributions

Proposition - For an n-dimensional multivate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, any of the possible marginal distributions will also (multivariate) normal.

For a univariate marginal distribution,

$$y_i = \mathcal{N}(\mu_i,\, \gamma_{ii})$$

For a bivariate marginal distribution,

$$\boldsymbol{y}_{ij} = \mathcal{N}\left(\begin{pmatrix}\mu_i\\ \mu_j\end{pmatrix},\, \begin{pmatrix}\gamma_{ii} & \gamma_{ij}\\ \gamma_{ji} & \gamma_{jj}\end{pmatrix}\right)$$

For a k-dimensional marginal distribution,

$$\boldsymbol{y}_{i_1,\cdots,i_k} = \mathcal{N}\left(\begin{pmatrix}\mu_{i_1}\\ \vdots\\ \mu_{i_k}\end{pmatrix},\, \begin{pmatrix}\gamma_{i_1 i_1} & \cdots & \gamma_{i_1 i_k}\\ \vdots & \ddots & \vdots\\ \gamma_{i_k i_1} & \cdots & \gamma_{i_k i_k}\end{pmatrix}\right)$$

Conditional Distributions

If we partition the n-dimensions into two pieces such that $Y = (Y_1,\ Y_2)^t$ then

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

$$\underset{k\times 1}{Y_1} \sim \mathcal{N}(\underset{k\times 1}{\boldsymbol{\mu}_1},\ \underset{k\times k}{\boldsymbol{\Sigma}_{11}})$$

$$\underset{n-k\times 1}{Y_2} \sim \mathcal{N}(\underset{n-k\times 1}{\boldsymbol{\mu}_2},\ \underset{n-k\times n-k}{\boldsymbol{\Sigma}_{22}})$$

Conditional Distributions

If we partition the n-dimensions into two pieces such that $Y = (Y_1,\ Y_2)^t$ then

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

$$\underset{k\times 1}{Y_1} \sim \mathcal{N}(\underset{k\times 1}{\boldsymbol{\mu}_1},\ \underset{k\times k}{\boldsymbol{\Sigma}_{11}})$$

$$\underset{n-k\times 1}{Y_2} \sim \mathcal{N}(\underset{n-k\times 1}{\boldsymbol{\mu}_2},\ \underset{n-k\times n-k}{\boldsymbol{\Sigma}_{22}})$$

then the conditional distributions are given by

$$Y_1 \mid Y_2 = a\ \sim \mathcal{N}(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\,\boldsymbol{\Sigma}_{22}^{-1}\,(a - \boldsymbol{\mu}_2),\ \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\,\boldsymbol{\Sigma}_{22}^{-1}\,\boldsymbol{\Sigma}_{21})$$

$$Y_2 \mid Y_1 = b\ \sim \mathcal{N}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\,\boldsymbol{\Sigma}_{11}^{-1}\,(b - \boldsymbol{\mu}_1),\ \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\,\boldsymbol{\Sigma}_{11}^{-1}\,\boldsymbol{\Sigma}_{21})$$

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* $\mathbf{y} = (y_{t_1}, y_{t_2}, ..., y_{t_n})^t$*, for every collection of time points* $t_1, t_2, \ldots, t_n$*, and every positive integer* n*, have a multivariate normal distribution.*

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 $\mathbf{y} = (y_{t_1}, y_{t_2}, ..., y_{t_n})^t$, for every collection of time points $t_1, t_2, \ldots, t_n$, and every positive integer n, have a multivariate normal distribution.*

So far we have only looked at examples of time series where T is discete (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 $\mathbb{R}$).

[0, 1]

Gaussian Process Regression

Parameterizing a Gaussian Process

Imagine we have a Gaussian process defined such that
$Y = \{Y_t \ : \ t \in [0, 1]\}$,

Parameterizing a Gaussian Process

Imagine we have a Gaussian process defined such that
$Y = \{Y_t \ : \ t \in [0, 1]\}$,

- We now have an uncountably infinite set of possible Y_ts.

Parameterizing a Gaussian Process

Imagine we have a Gaussian process defined such that
$Y = \{Y_t \ : \ t \in [0,1]\}$,

- We now have an uncountably infinite set of possible Y_ts.
- We will only have a (small) finite number of observations $Y_1, \ldots, Y_n$ with which to say something useful about this infinite dimension process.

Parameterizing a Gaussian Process

Imagine we have a Gaussian process defined such that
$Y = \{Y_t \ : \ t \in [0, 1]\}$,

- We now have an uncountably infinite set of possible Y_ts.
- We will only have a (small) finite number of observations $Y_1, \ldots, Y_n$ with which to say something useful about this infinite dimension process.
- The unconstrained covariance matrix for the observed data can have up to $n(n+1)/2$ unique values ($p >>> n$)

Parameterizing a Gaussian Process

Imagine we have a Gaussian process defined such that $Y = \{Y_t \; : \; t \in [0,1]\}$,

- We now have an uncountably infinite set of possible Y_ts.
- We will only have a (small) finite number of observations $Y_1, \ldots, Y_n$ with which to say something useful about this infinite dimension process.
- The unconstrained covariance matrix for the observed data can have up to $n(n+1)/2$ unique values ($p >>> n$)
- Necessary to make some simplifying assumptions:
    - Stationarity
    - Simple parameterization of Σ

Covariance Functions

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

$L = 0 \quad \Sigma \Rightarrow \sigma^2$

$L \to \infty \quad \Sigma \to 0$

Exponential Covariance:

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

σ^2 → Scale; l → length / range

Squared Exponential Covariance:

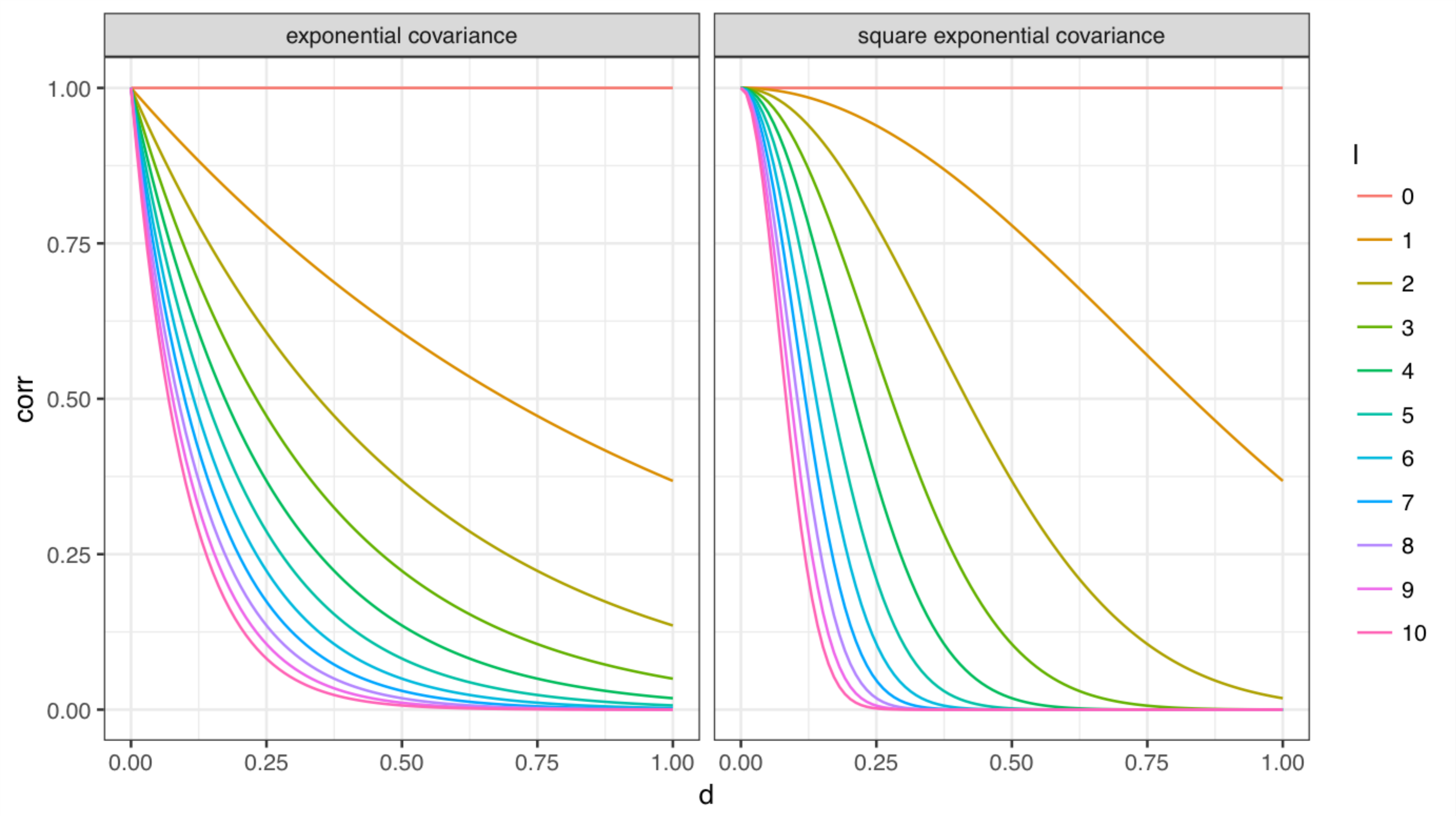
$$\Sigma(y_t, y_{t'}) = \sigma^2 \exp\left(-\, (|t - t'| \; l\,)^2 \right)$$

$\exp(\ldots)$: Corr

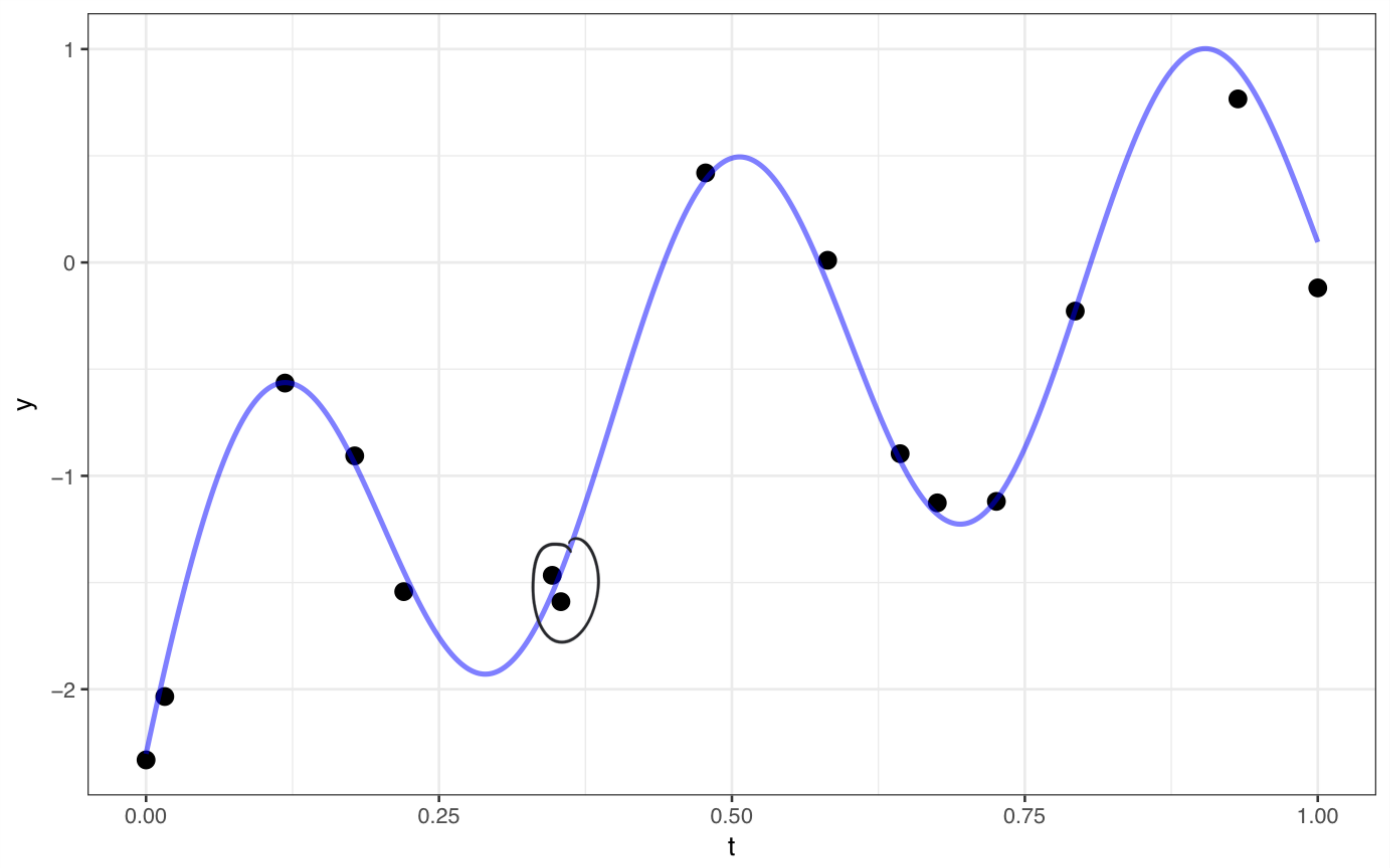
Powered Exponential Covariance ($p \in (0, 2]$):

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

Covariance Function Decay



Example



Prediction

Our example has 15 observations which we would like to use as the basis for predicting Y_t at other values of t (say a grid of values from 0 to 1).

Prediction

Our example has 15 observations which we would like to use as the basis for predicting Y_t at other values of t (say a grid of values from 0 to 1).

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

Prediction

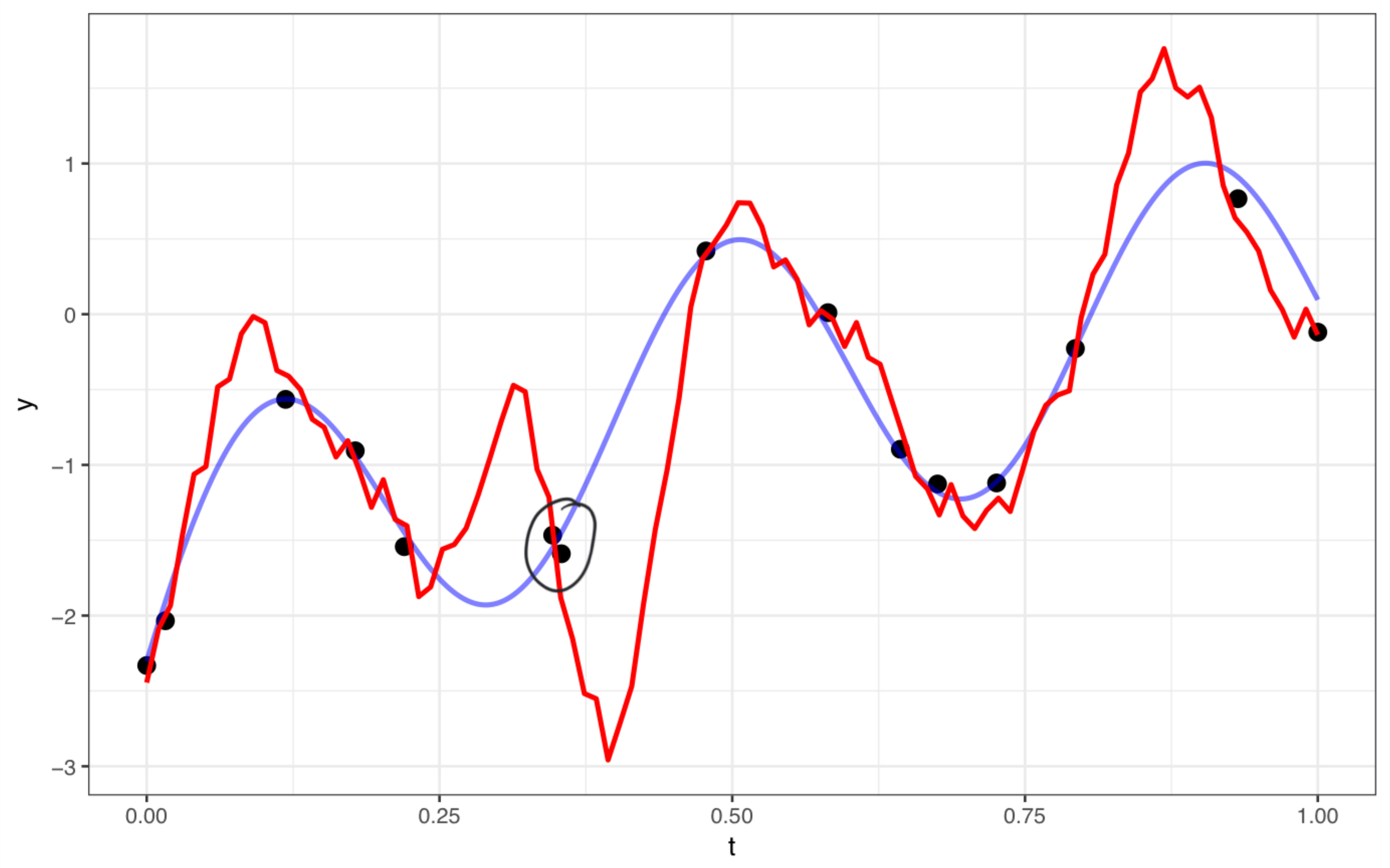
Our example has 15 observations which we would like to use as the basis for predicting Y_t at other values of t (say a grid of values from 0 to 1).

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

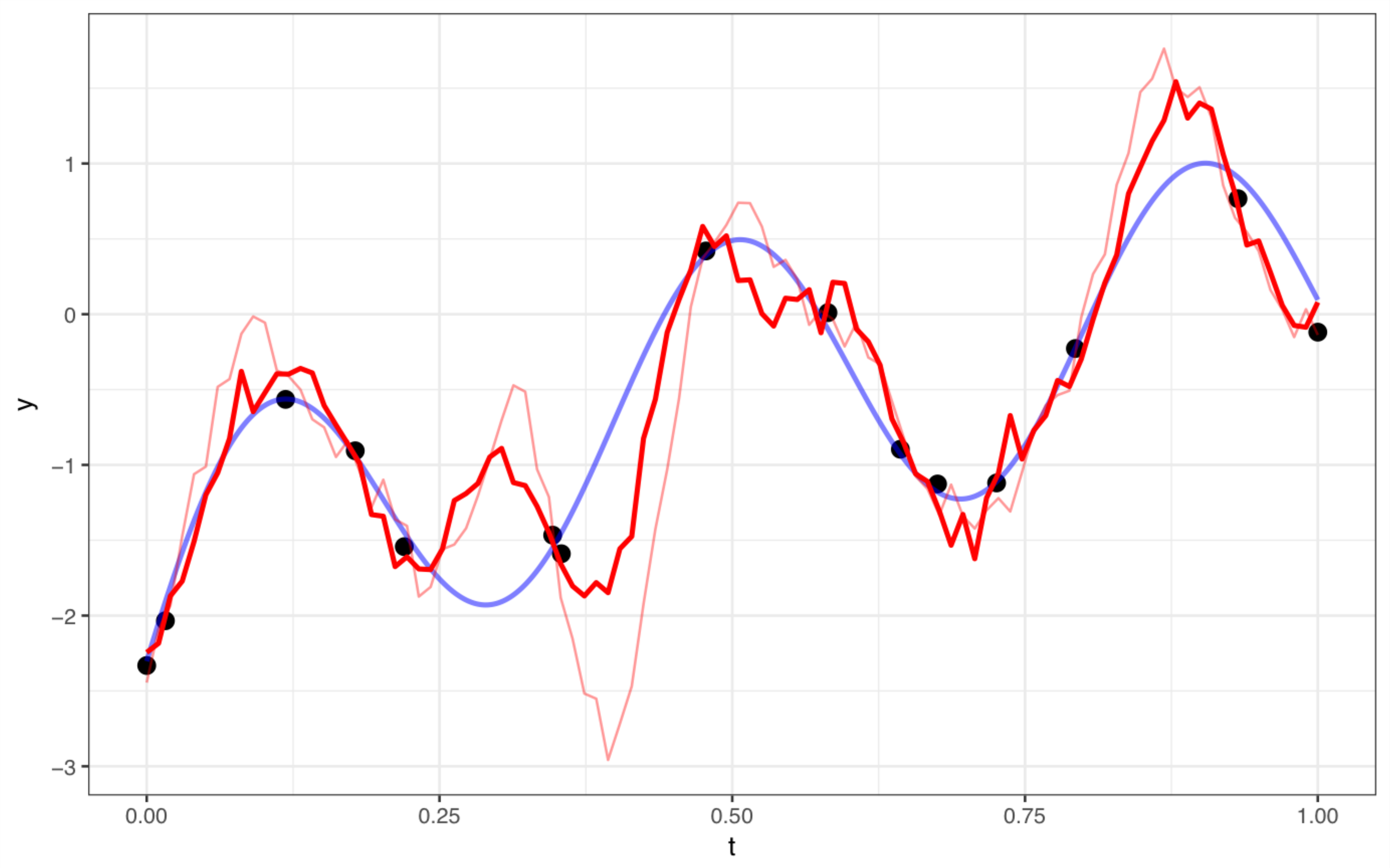
We therefore want to sample from $Y_{pred} | Y_{obs}$.

$$Y_{pred} \mid Y_obs = y \sim \mathcal{N}(\boldsymbol{\Sigma}_{po}\, \boldsymbol{\Sigma}_{obs}^{-1}\, y,\ \boldsymbol{\Sigma}_{pred} - \boldsymbol{\Sigma}_{po}\, \boldsymbol{\Sigma}_{pred}^{-1}\, \boldsymbol{\Sigma}_{op})$$

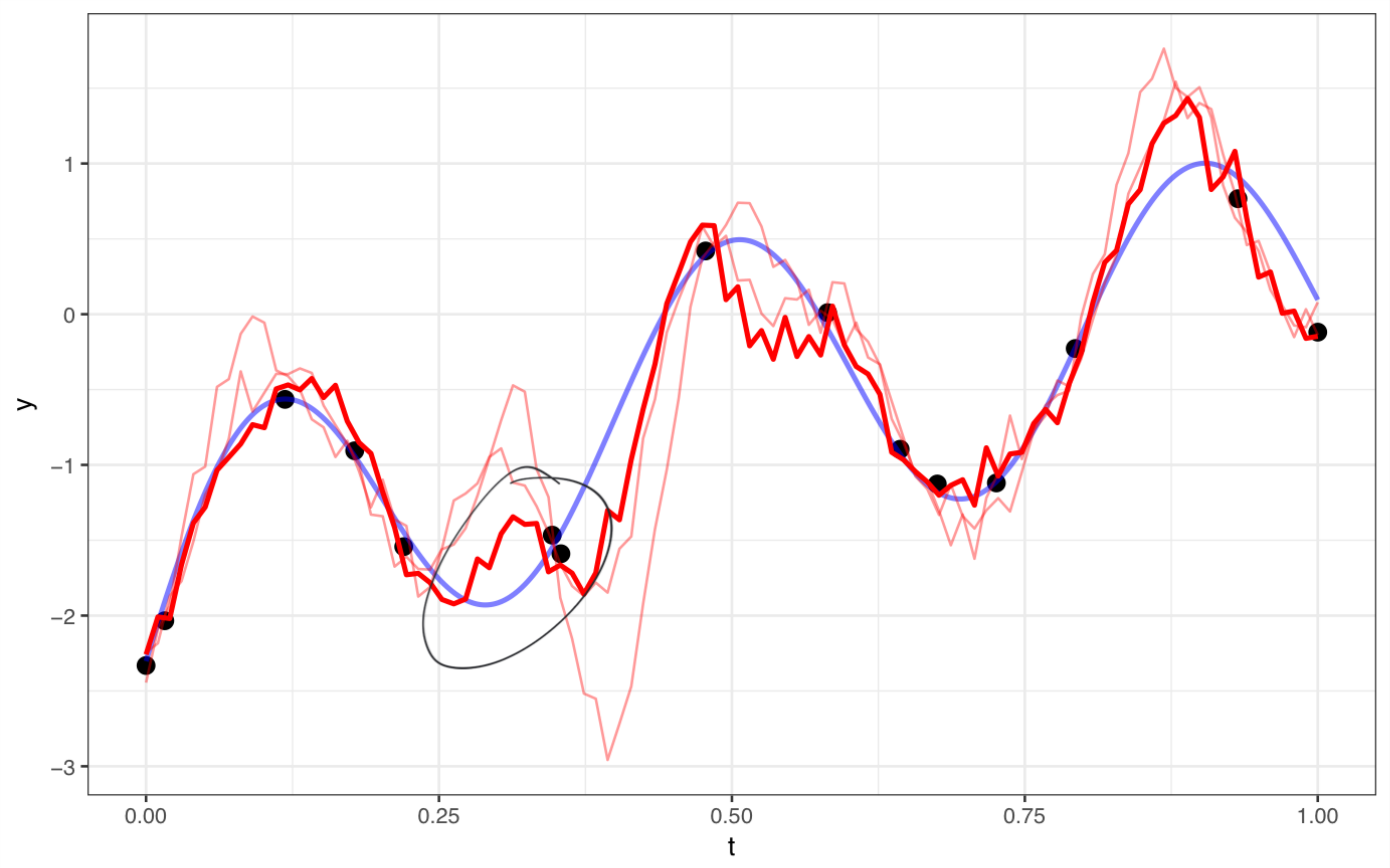
Draw 1



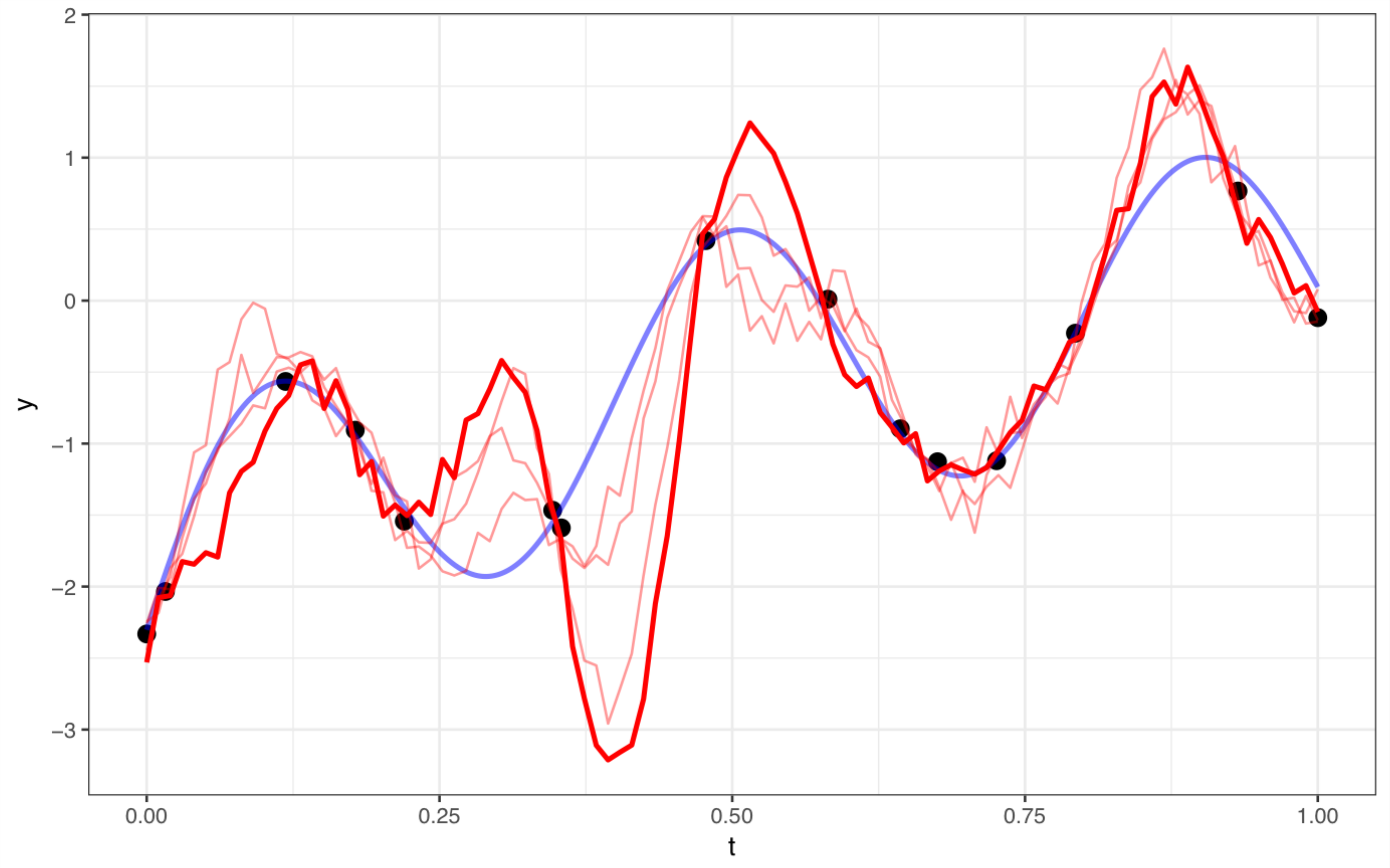
Draw 2



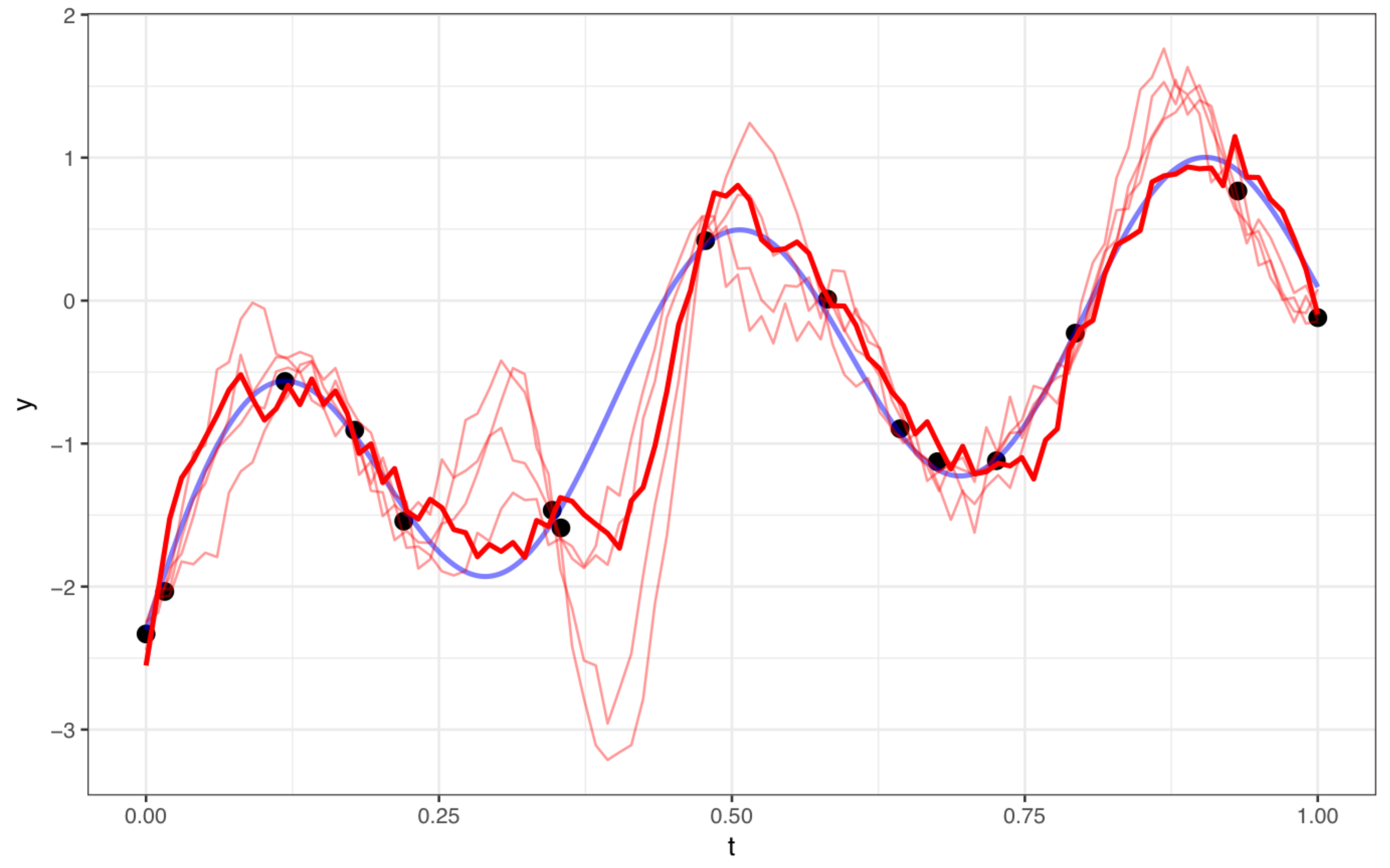
Draw 3



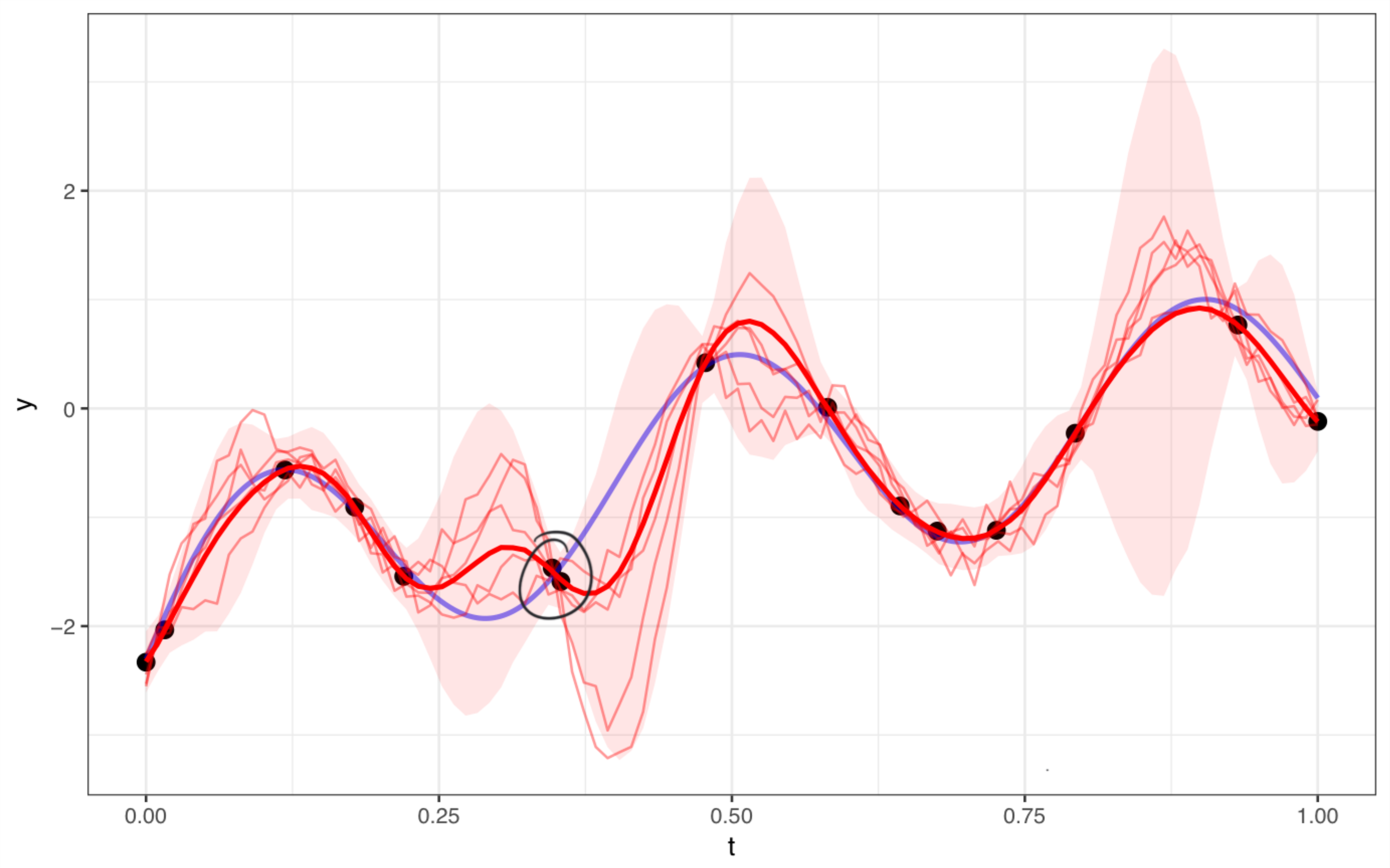
Draw 4



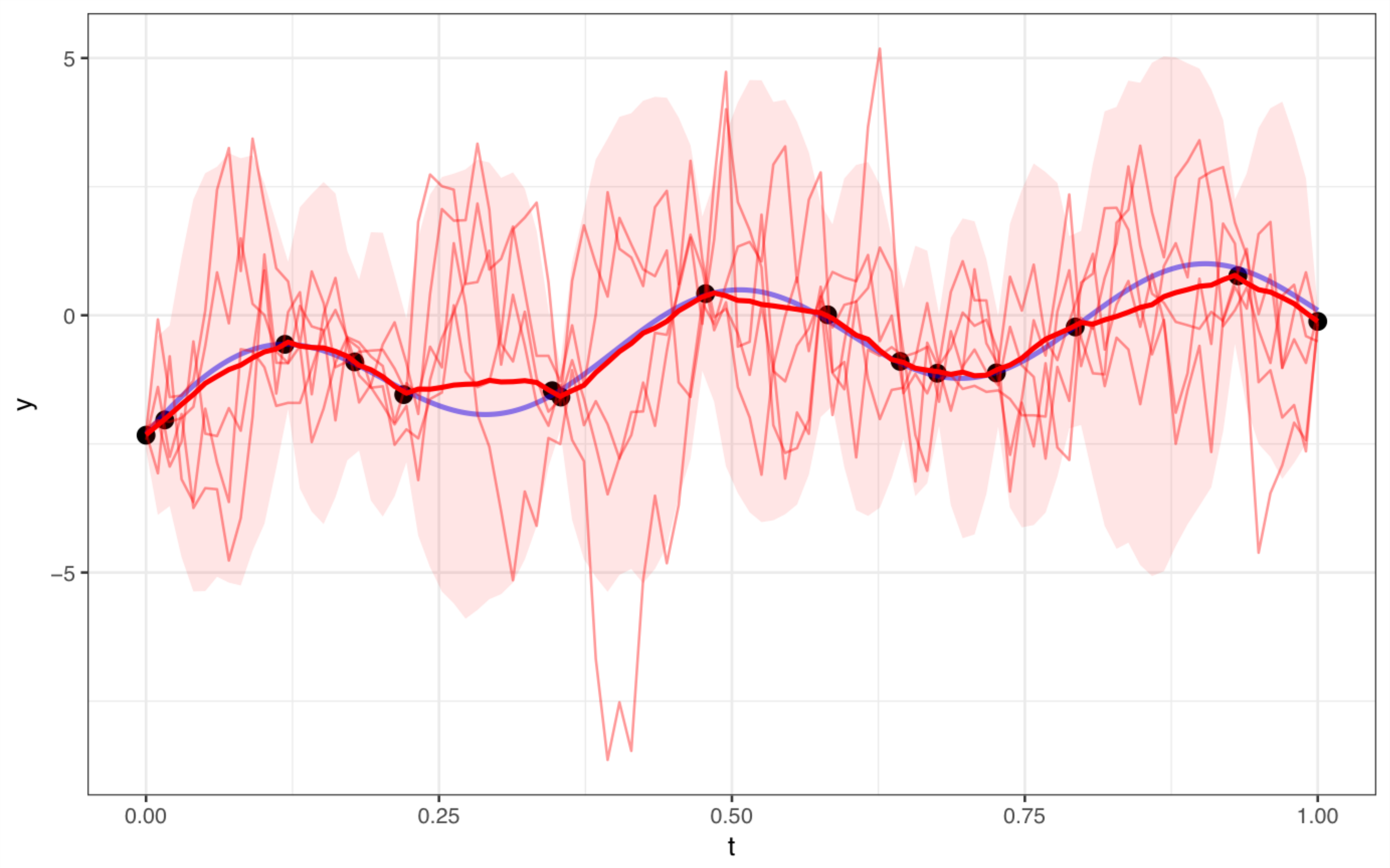
Draw 5



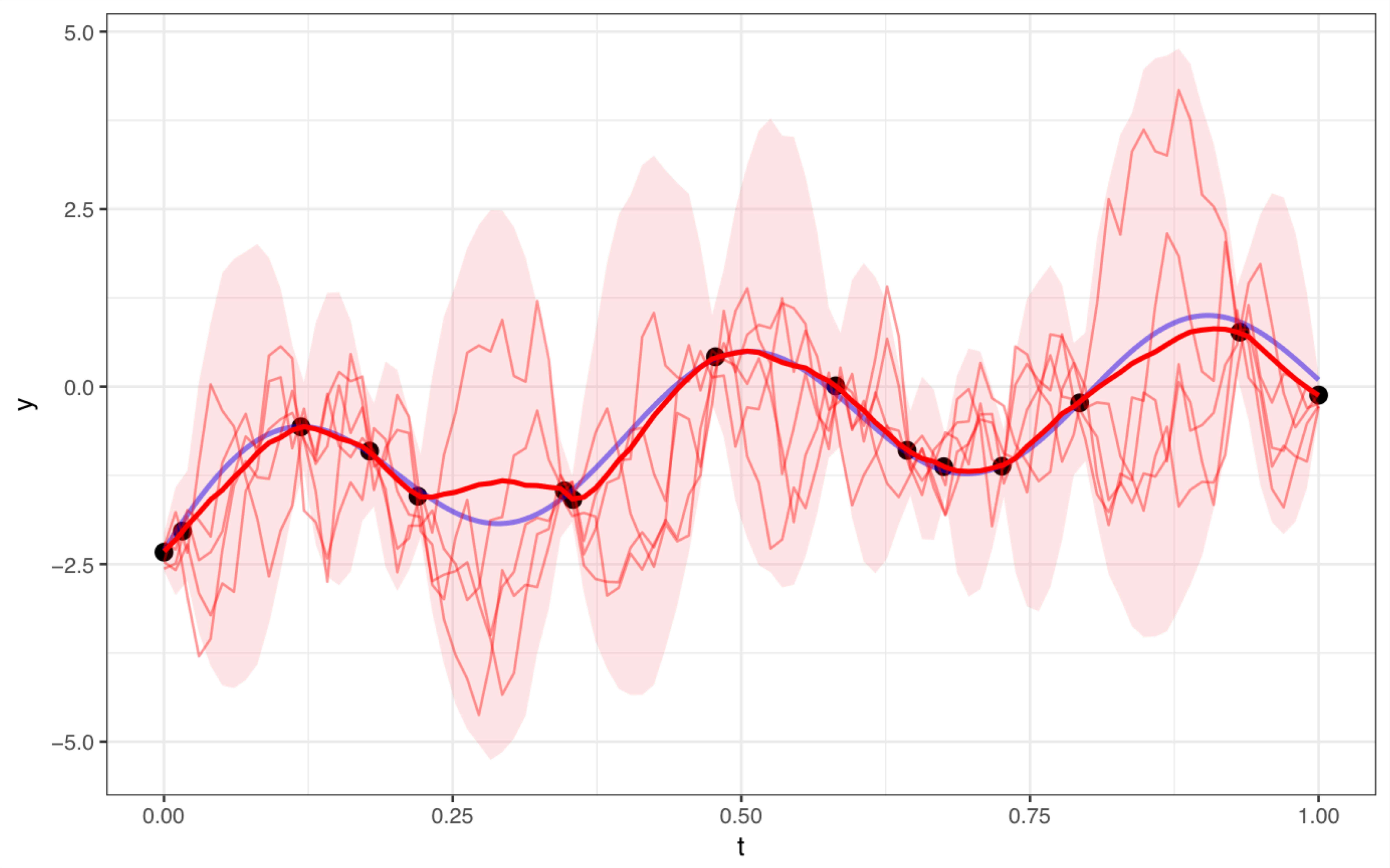
Many draws later



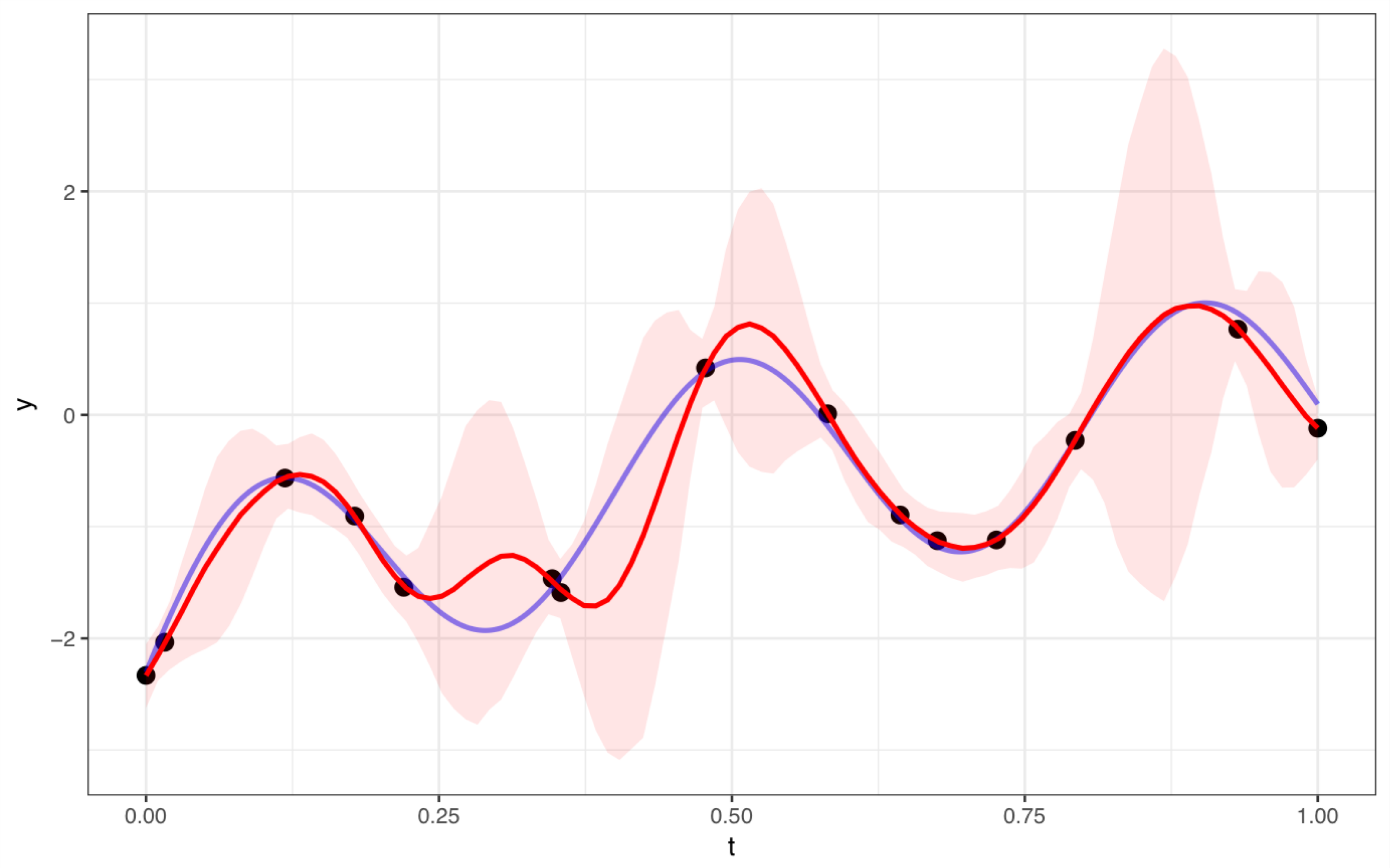
Exponential Covariance



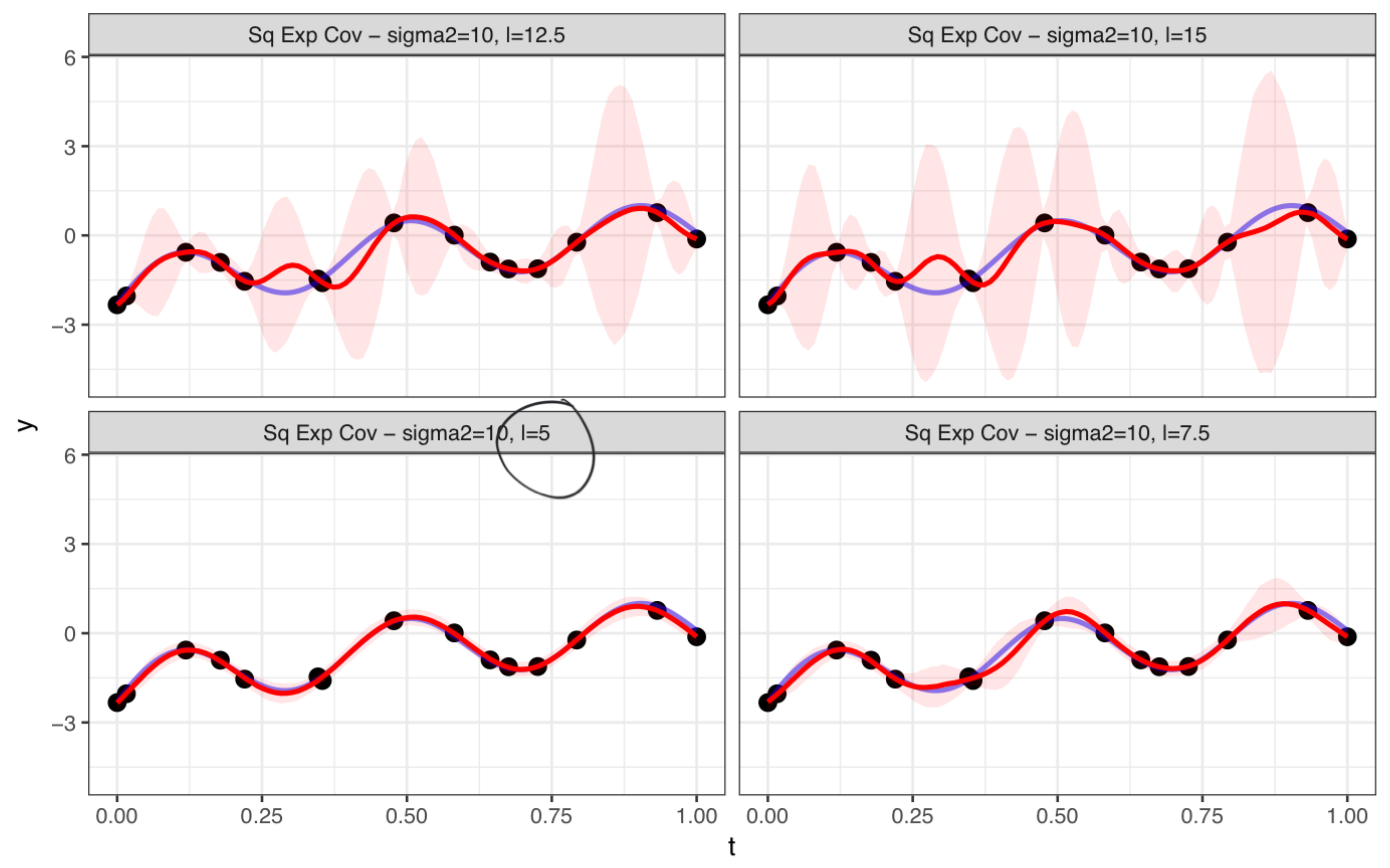
Powered Exponential Covariance ($p = 1.5$)



Back to the square exponential



Changing the range (l)



Effective Range

For the square exponential covariance

$$Cov(d) = \sigma^2 \exp\left(-(l \cdot d)^2\right)$$
$$Corr(d) = \exp\left(-(l \cdot d)^2\right)$$

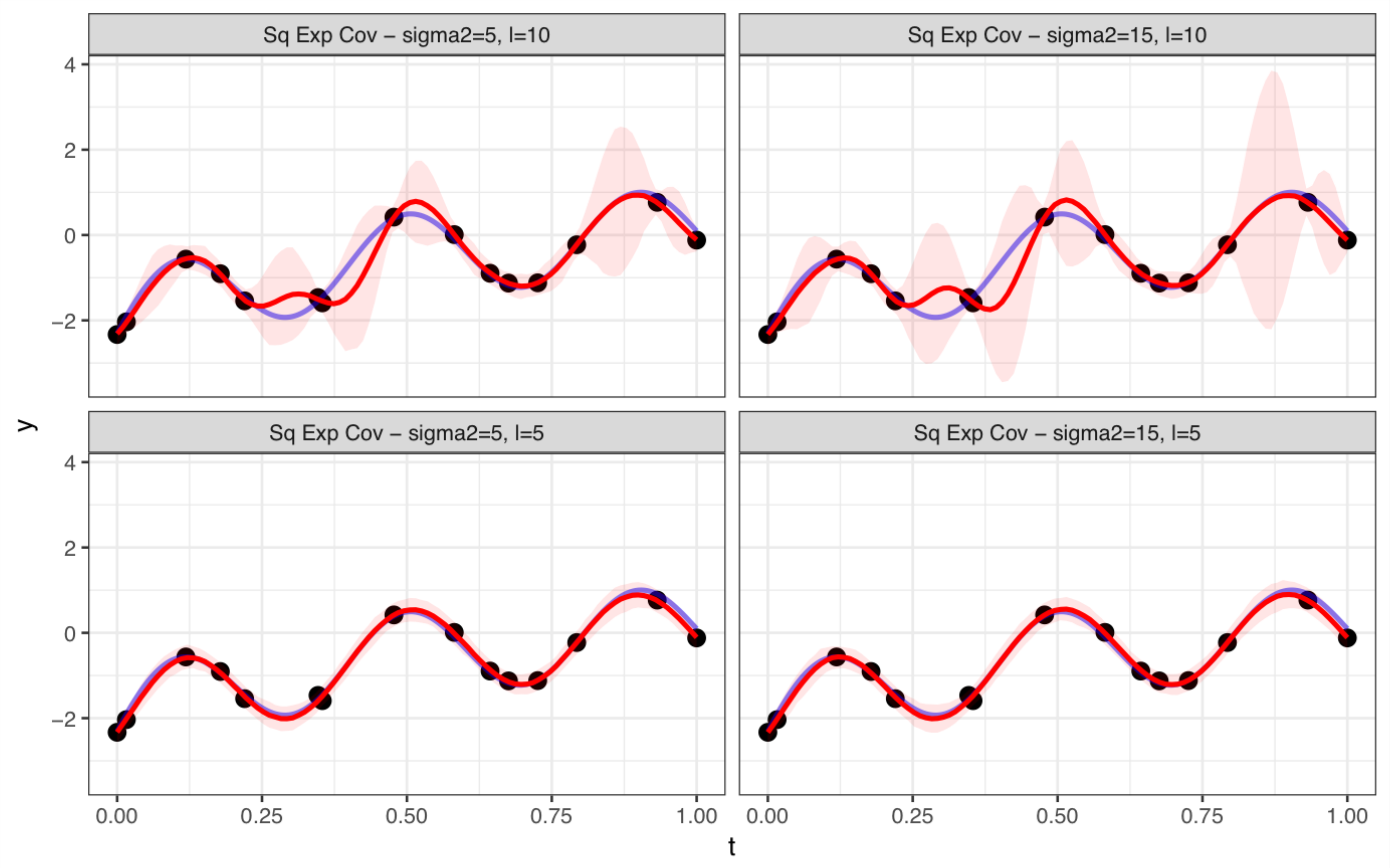
we would like to know, for a given value of l, beyond what distance apart must observations be to have a correlation less than 0.05?

$$e^{-l^2 \cdot d^2} < 0.05$$

$$l^2 d^2 < 3$$

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

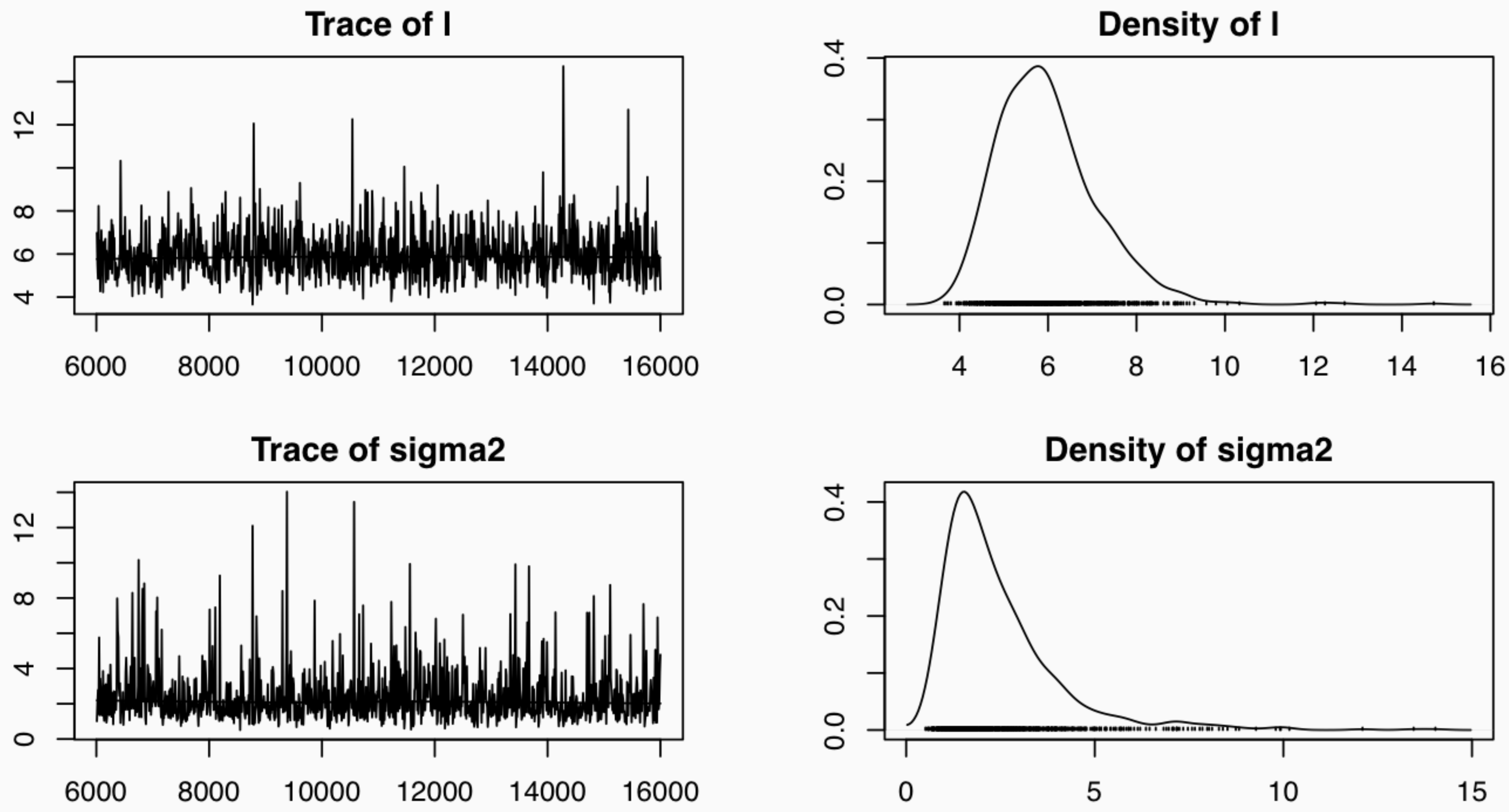
Changing the scale (σ^2)



Fitting

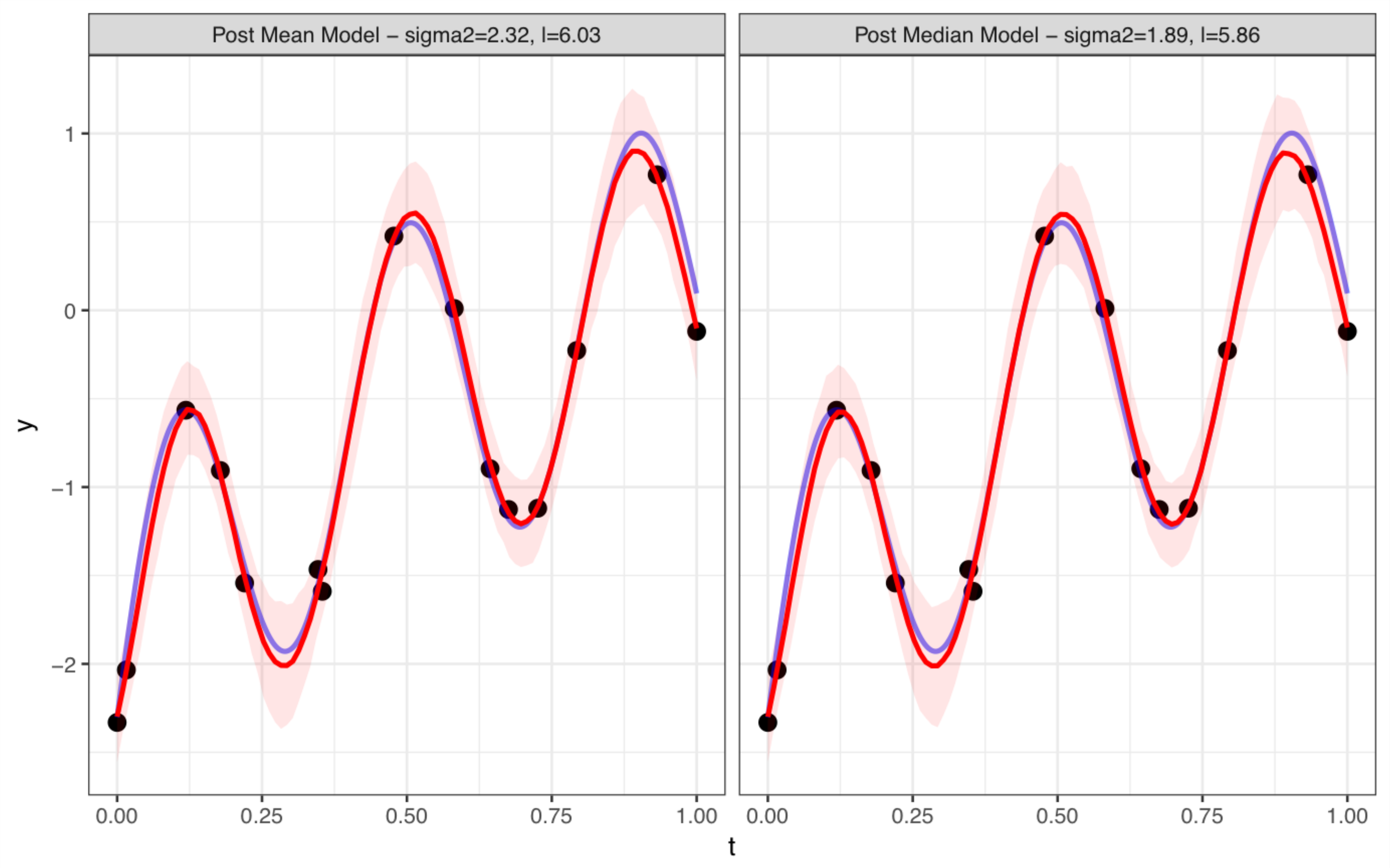
```
## model{
##   y ~ dmnorm(mu, inverse(Sigma))
##
##   for (i in 1:N) {
##     mu[i] <- 0
##   }
##
##   for (i in 1:(N-1)) {
##     for (j in (i+1):N) {
##       Sigma[i,j] <- sigma2 * exp(- pow(l*d[i,j],2))
##       Sigma[j,i] <- Sigma[i,j]
##     }
##   }
##
##   for (k in 1:N) {
##     Sigma[k,k] <- sigma2 + 0.01
##   }
##
##   sigma2   ~ dlnorm(0, 1)
##   l        ~ dt(0, 2.5, 1) T(0,) # Half-cauchy(0,2.5)
## }
```

Trace plots

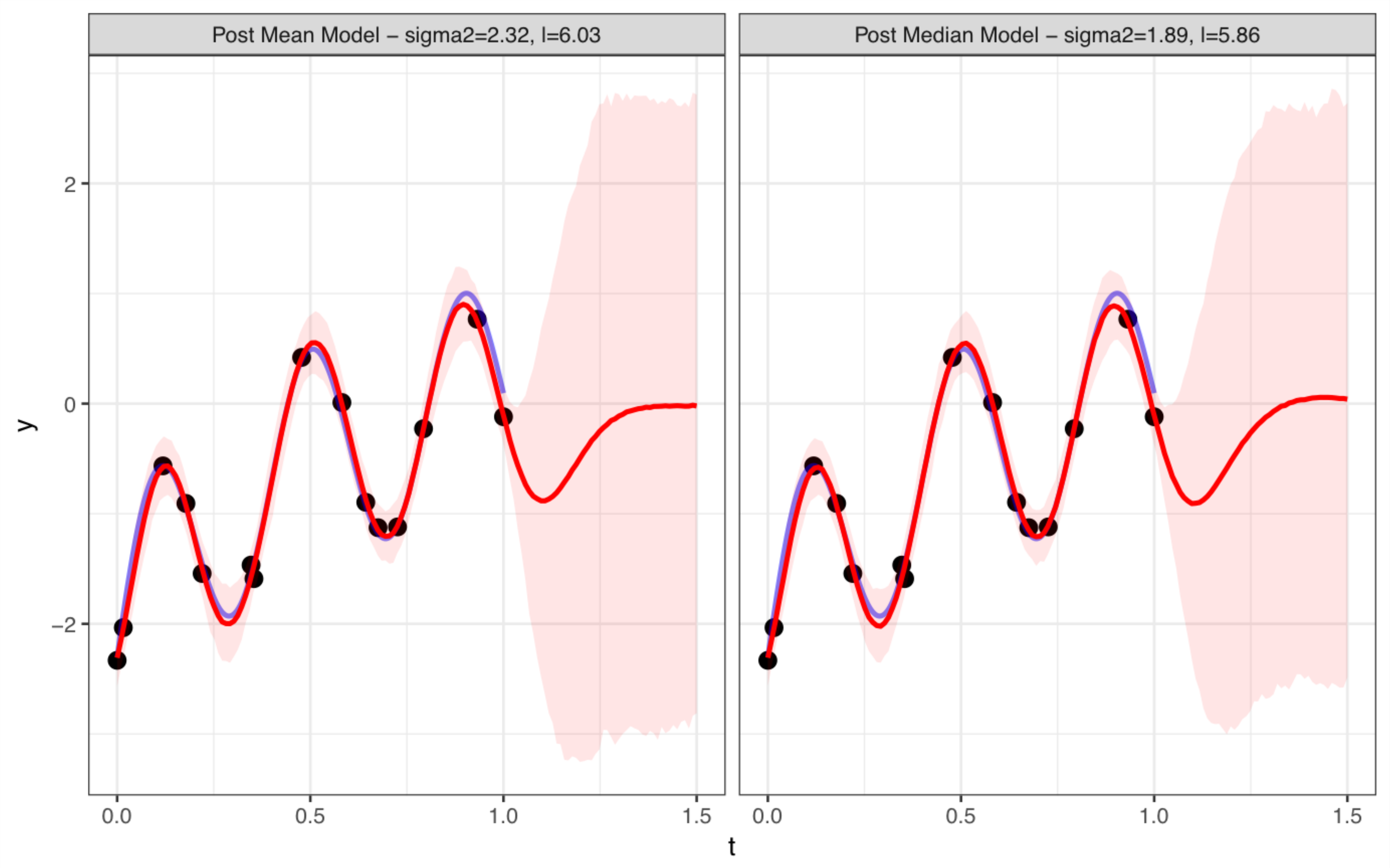


param	post_mean	post_med	post_lower	post_upper
l	5.981289	5.833655	4.2669795	8.456006
sigma2	2.457979	2.032632	0.8173064	7.168197

Fitted models



Forcasting



Post Mean Model – sigma2=2.32, l=6.03
Post Median Model – sigma2=1.89, l=5.86
y
2
0
−2
0.0
0.5
1.0
1.5
t