Lecture 12

Gaussian Process Models

3/01/2018

Multivariate Normal

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

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

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

Density

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

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

and its log density is given by

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

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

Sampling

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

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

Sampling

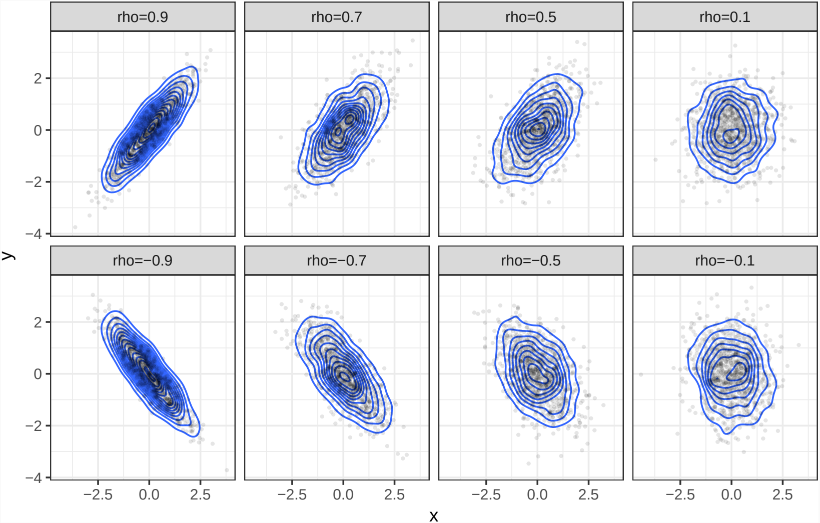
To generate draws from an n-dimensional multivate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$,

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

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

Bivariate Example

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



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

Marginal distributions

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

For a univariate marginal distribution,

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

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

For a univariate marginal distribution,

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

For a bivariate marginal distribution,

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

Marginal distributions

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

For a univariate marginal distribution,

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

For a bivariate marginal distribution,

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

For a k-dimensional marginal distribution,

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

If we partition the n-dimensions into two pieces such that $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)^t$ then

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

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

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

If we partition the n-dimensions into two pieces such that $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)^t$ then

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

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

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

then the conditional distributions are given by

$$\mathbf{Y_1} \mid \mathbf{Y}_2 = \mathbf{a} \sim \mathcal{N}(\boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}}\,\boldsymbol{\Sigma_{22}}^{-1}\,(\mathbf{a} - \boldsymbol{\mu_2}),\ \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}}\,\boldsymbol{\Sigma_{22}}^{-1}\,\boldsymbol{\Sigma_{21}})$$

$$\mathbf{Y_2} \mid \mathbf{Y}_1 = \mathbf{b} \sim \mathcal{N}(\boldsymbol{\mu_2} + \boldsymbol{\Sigma_{21}}\,\boldsymbol{\Sigma_{11}}^{-1}\,(\mathbf{b} - \boldsymbol{\mu_1}),\ \boldsymbol{\Sigma_{22}} - \boldsymbol{\Sigma_{21}}\,\boldsymbol{\Sigma_{11}}^{-1}\,\boldsymbol{\Sigma_{21}})$$

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 $t_1, t_2, \dots, t_n$, and every positive integer n, have a multivariate normal distribution.*

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 $\mathbb{R}$).

Gaussian Process Regression

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

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 $Y(t)$s.

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 $Y(t)$s.
- We will only have a (small) finite number of observations $Y(\boldsymbol{t_1}), \ldots, Y(\boldsymbol{t_n})$ with which to say something useful about this infinite dimension process.

Parameterizing a Gaussian Process

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

- We now have an uncountably infinite set of possible $Y(t)$s.
- We will only have a (small) finite number of observations $Y(t_1), \ldots, Y(t_n)$ 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*

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 $Y(t)$s.
- We will only have a (small) finite number of observations $Y(t_1), \ldots, Y(t_n)$ 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*
- Necessary to make some simplifying assumptions:
    - Stationarity
    - Simple parameterization of Σ

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

Covariance Functions

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

Exponential Covariance:

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

Covariance Functions

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

Exponential Covariance:

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

Squared Exponential Covariance:

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

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

Exponential Covariance:

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

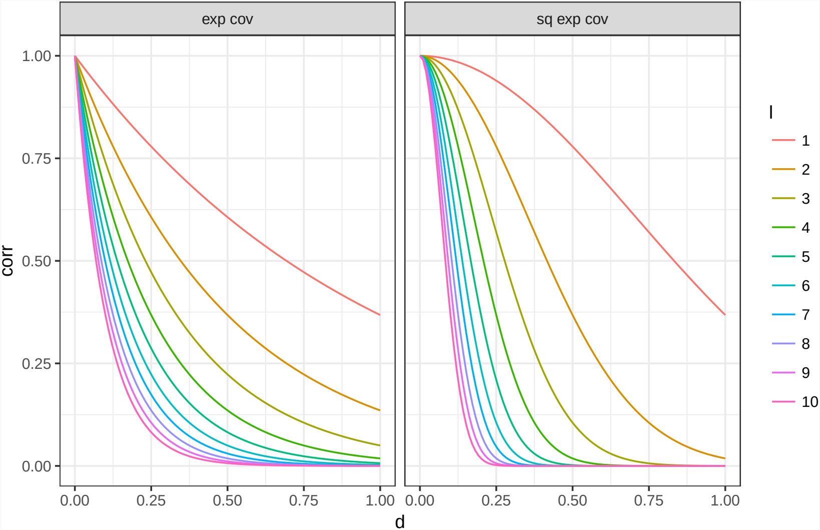
Squared Exponential Covariance:

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

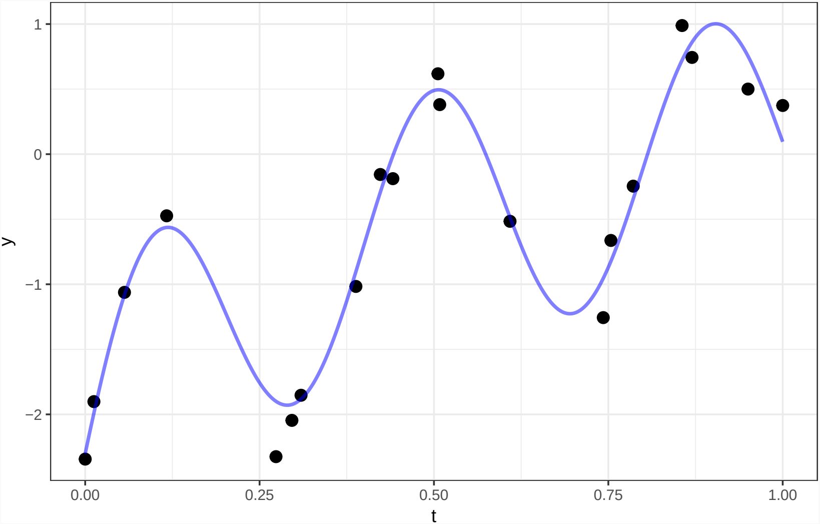
Powered Exponential Covariance ($p \in (0, 2]$):

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

Covariance Function - Correlation 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

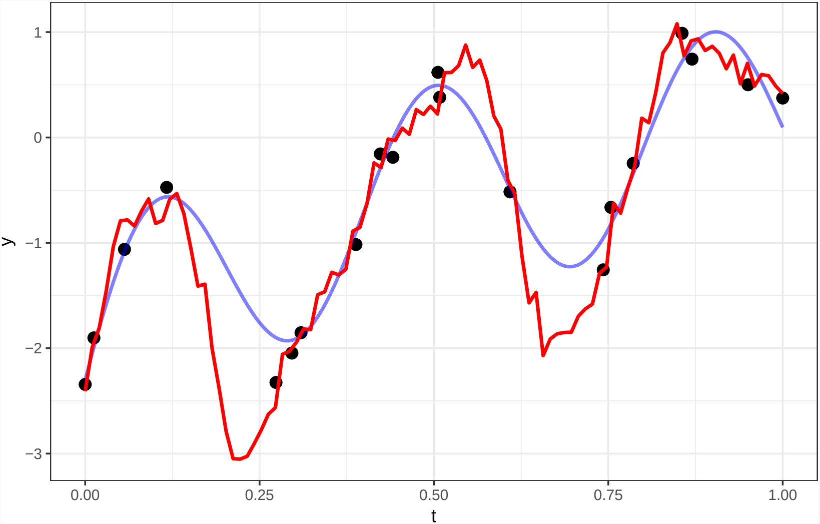
Our example has 15 observations which we would like to use as the basis for predicting Y_t at other values of t (say a grid of values from 0 to 1).

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

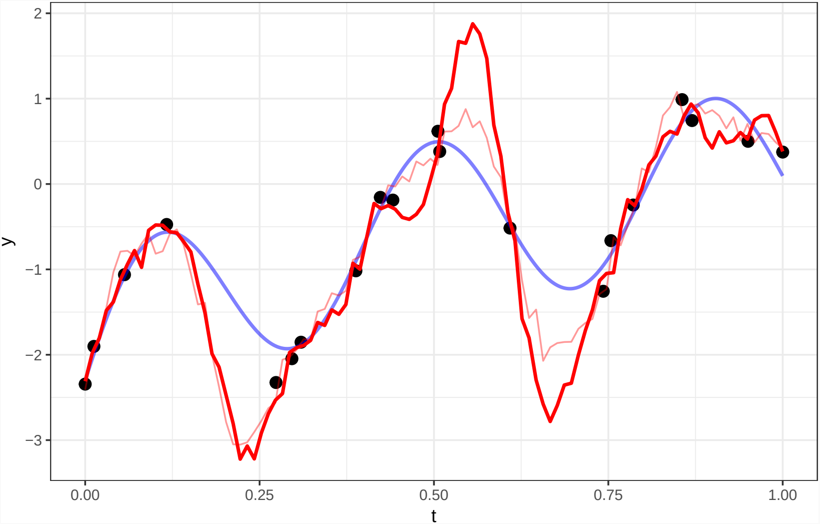
We therefore want to sample from $\mathbf{Y}_{pred}|\mathbf{Y}_{obs}$.

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

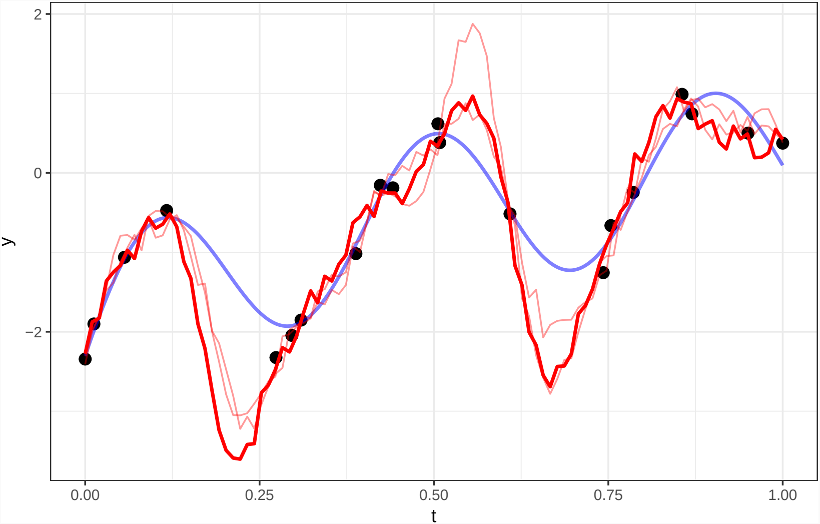
Draw 1



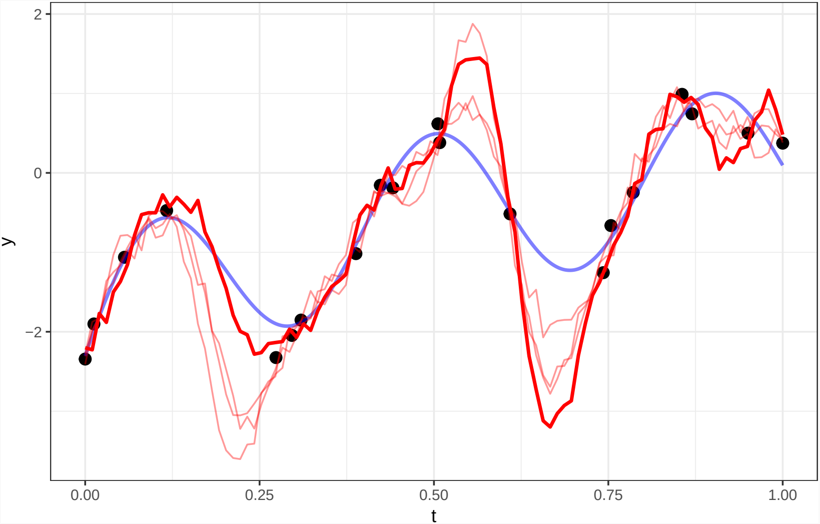
Draw 2



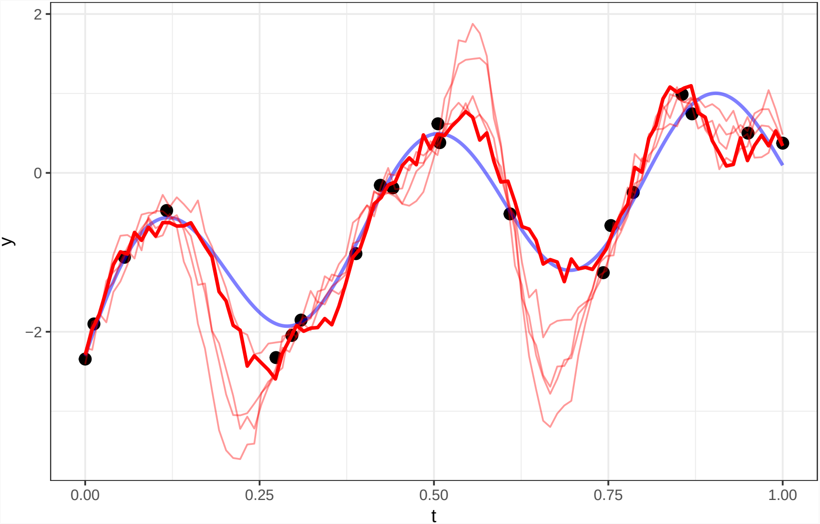
Draw 3



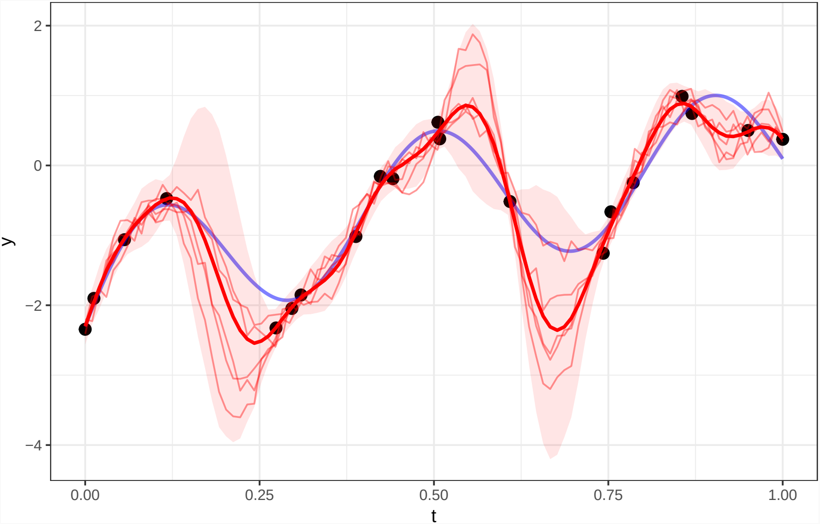
Draw 4



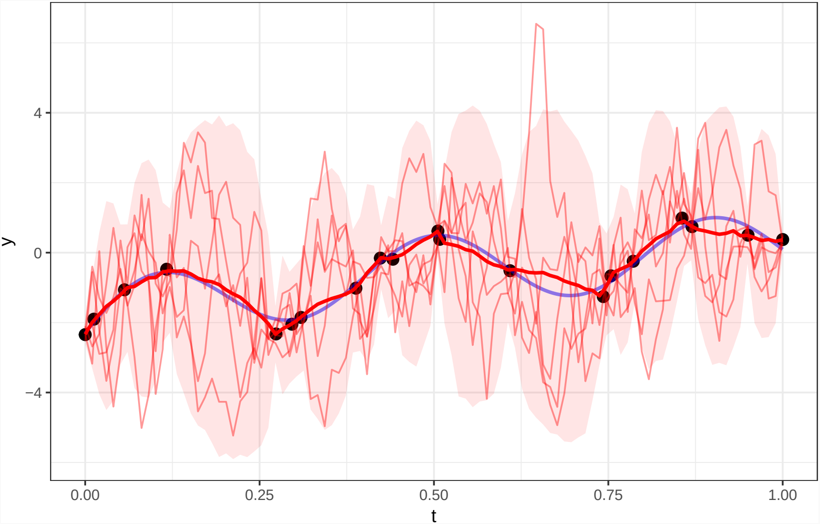
Draw 5



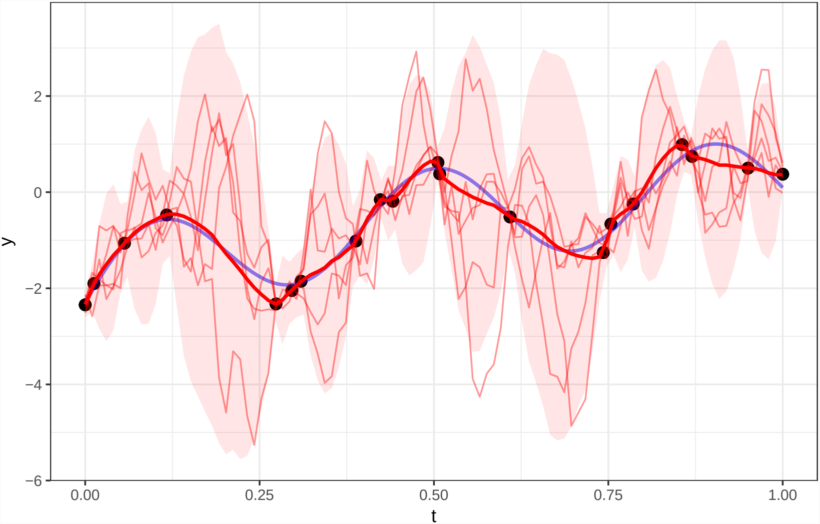
Many draws later



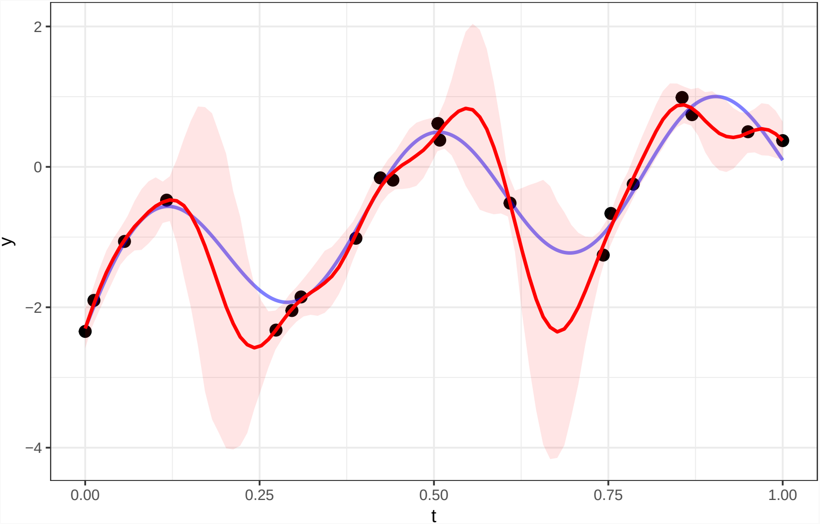
Exponential Covariance



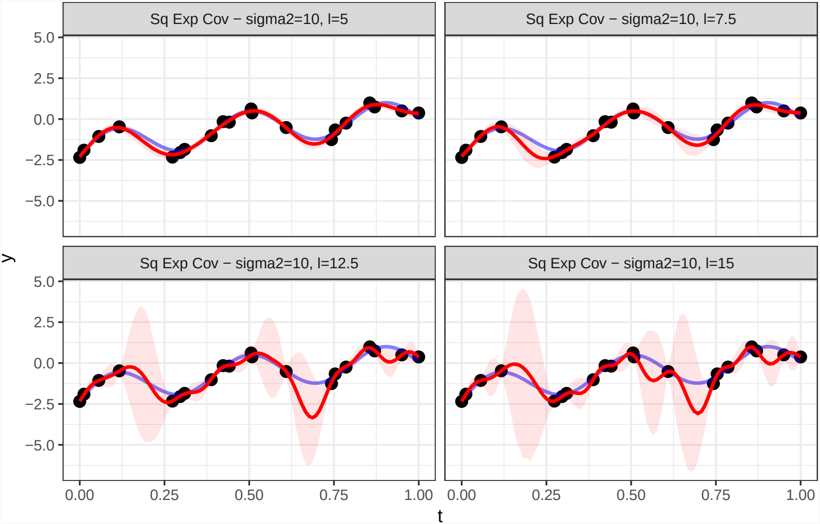
Powered Exponential Covariance ($p = 1.5$)



Back to the square exponential



Changing the range (l)

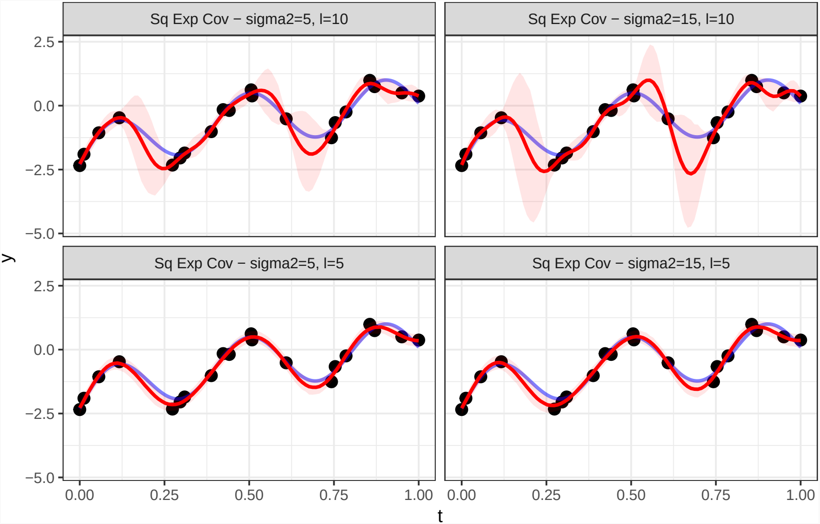


For the square exponential covariance

$$\begin{aligned} Cov(d) &= \sigma^2 \exp\left(-(l \cdot d)^2\right) \\ Corr(d) &= \exp\left(-(l \cdot d)^2\right) \end{aligned}$$

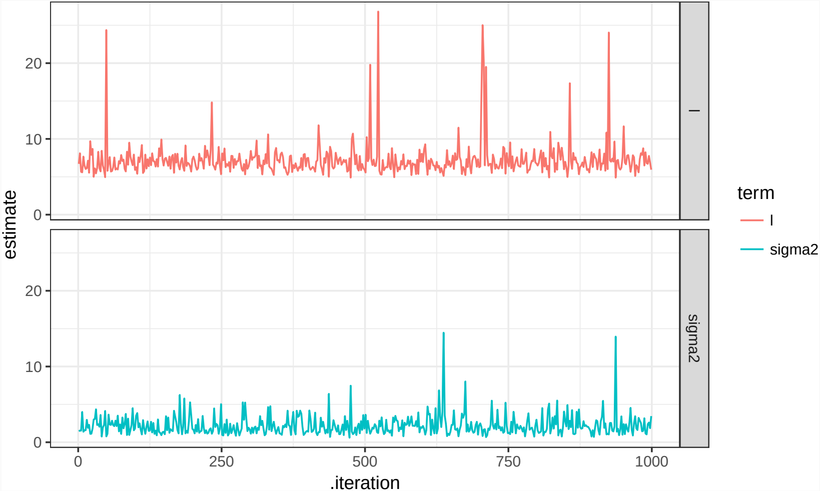
we would like to know, for a given value of l, beyond what distance apart must observations be to have a correlation less than 0.05?

Changing the scale (σ^2)



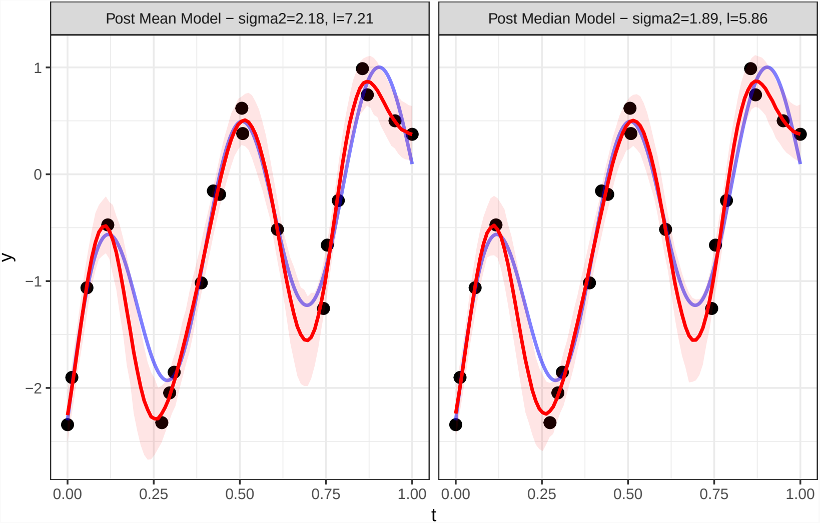
```
gp_sq_exp_model = "model{
  y ~ dmnorm(mu, inverse(Sigma))

  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

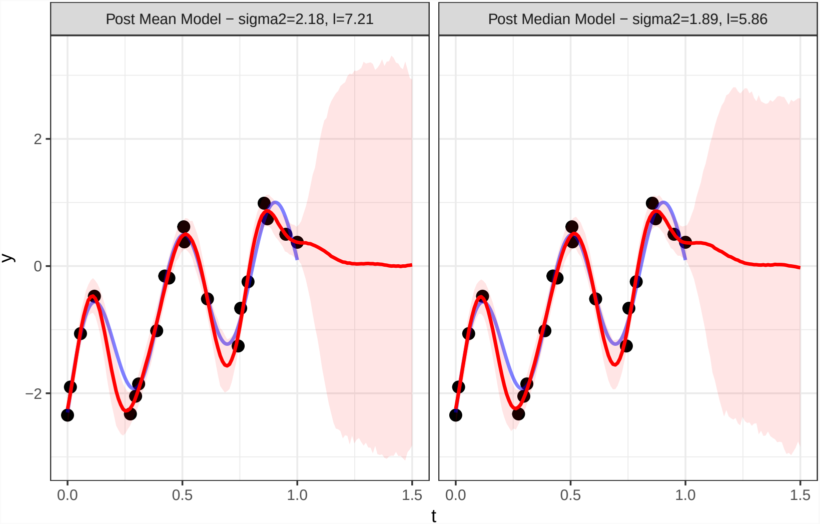


param	post_mean	post_med	post_lower	post_upper
l	7.21	6.79	5.23	10.95
sigma2	2.18	1.86	0.84	5.24

Fitted models



Forcasting


Post Mean Model − sigma2=2.18, l=7.21
Post Median Model − sigma2=1.89, l=5.86
y
2
0
−2
0.0
0.5
1.0
1.5
t