MATH 569 Statistical Learning

Part VII: Model Inference

Maggie Cheng

ILLINOIS INSTITUTE OF TECHNOLOGY

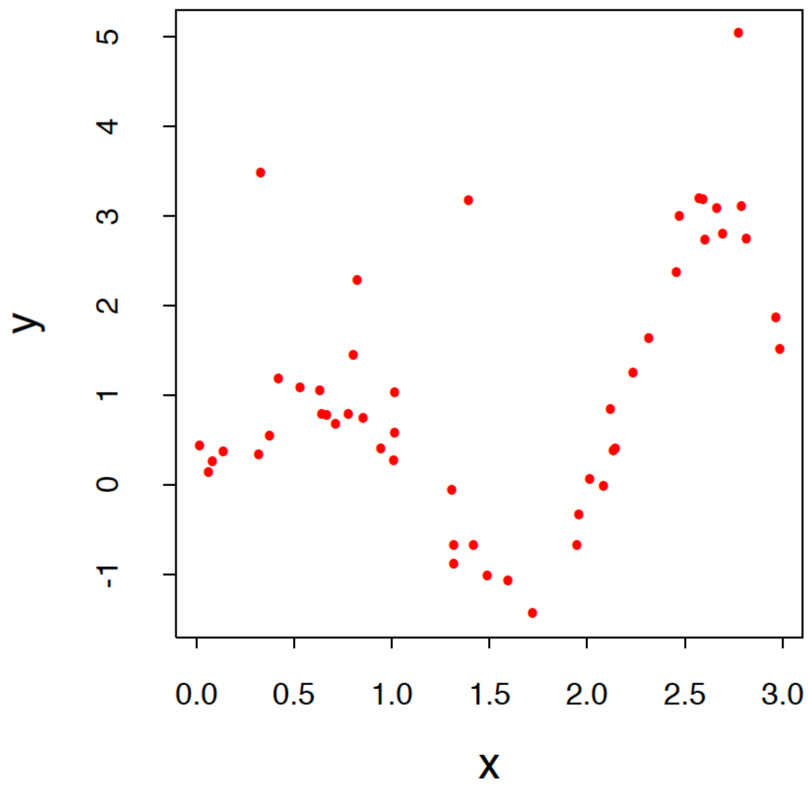
Fig 8.1

Use cubic splines to fit the data (3 knots)

4x4-3x3=7

$$\mu(x) = \sum_{j=1}^{7} \beta_j h_j(x)$$

Data to fit



Set of seven B-spline basis functions

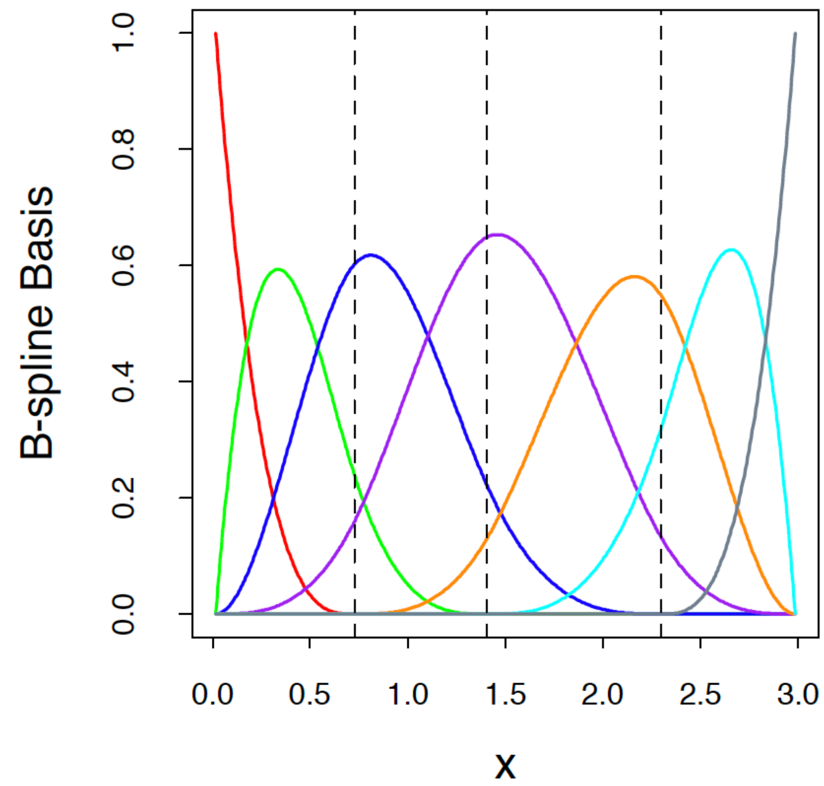


Fig 8.2

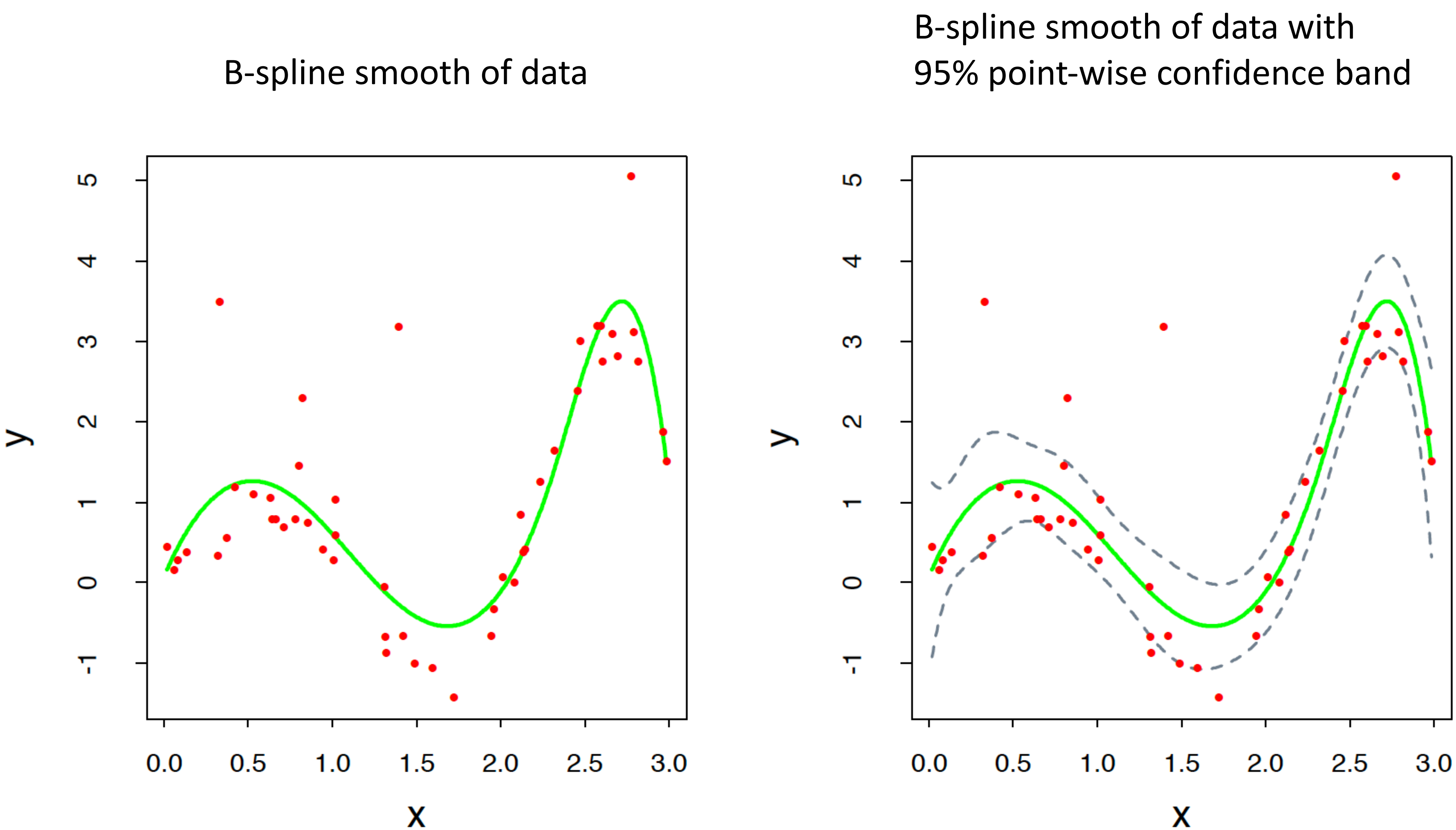

B-spline smooth of data
B-spline smooth of data with
95% point-wise confidence band
y
x
0.0 0.5 1.0 1.5 2.0 2.5 3.0
-1 0 1 2 3 4 5

Fig 8.2

B-spline smooth of each of the 10 bootstrap datasets (B=10)

95% standard error bands computed from the bootstrap distribution

With B=200, 2.5%*200=5
The band: The 5th largest and smallest at each x

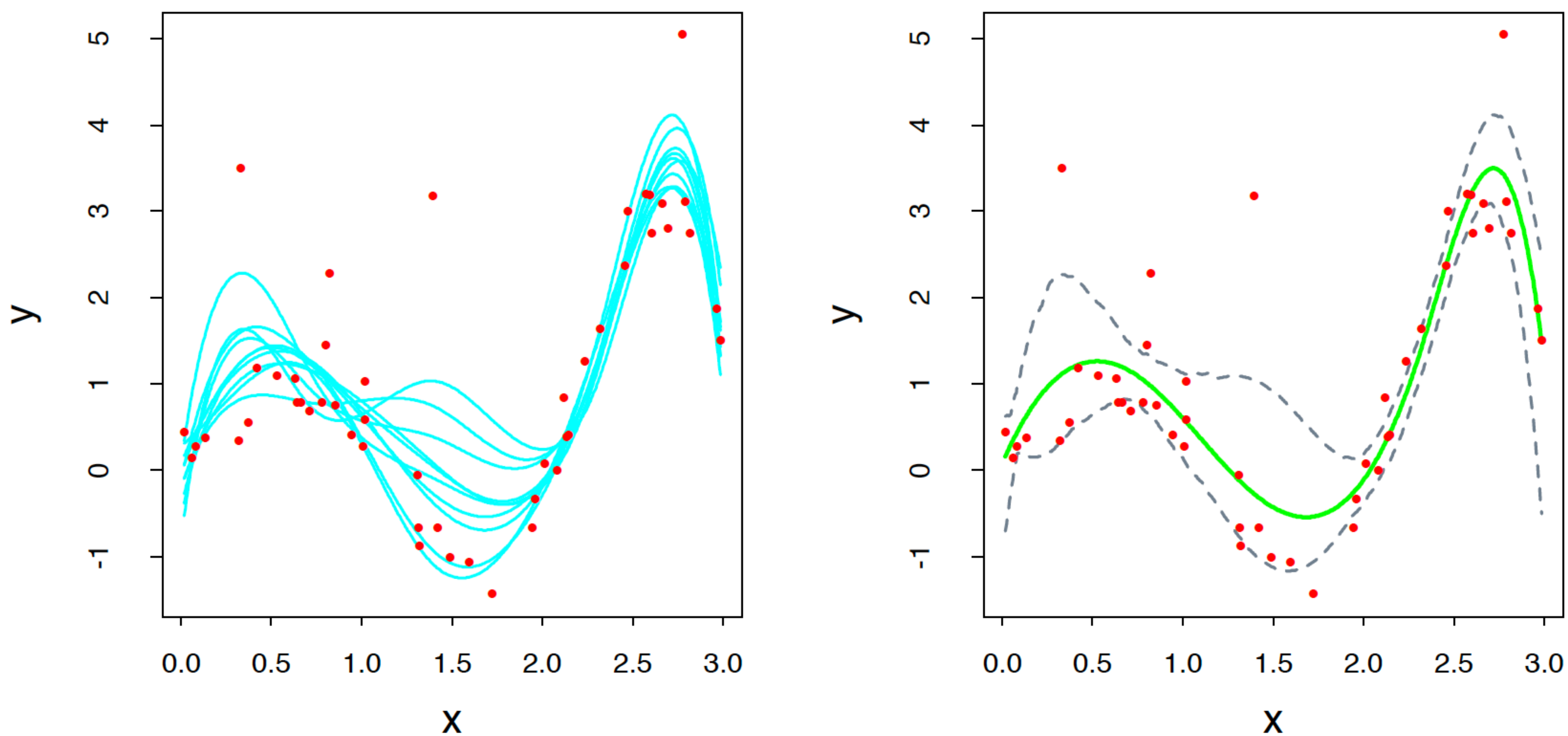


Fig 8.3

Smoothing example: Ten draws from the Gaussian prior distribution for the function μ(x).

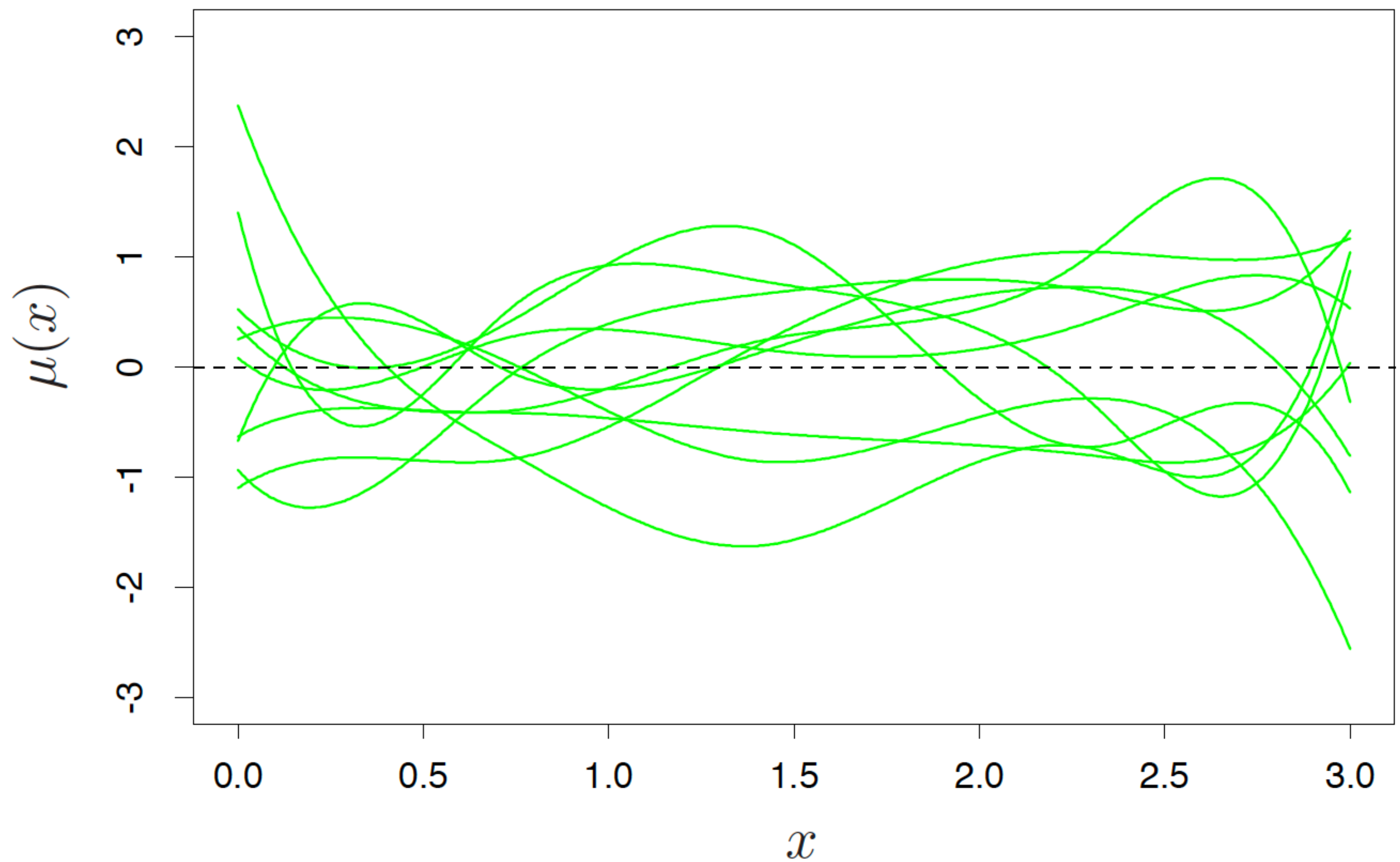


Fig 8.4 Ten draws from the posterior distribution for the function μ(x).

The purple curves are the posterior means.

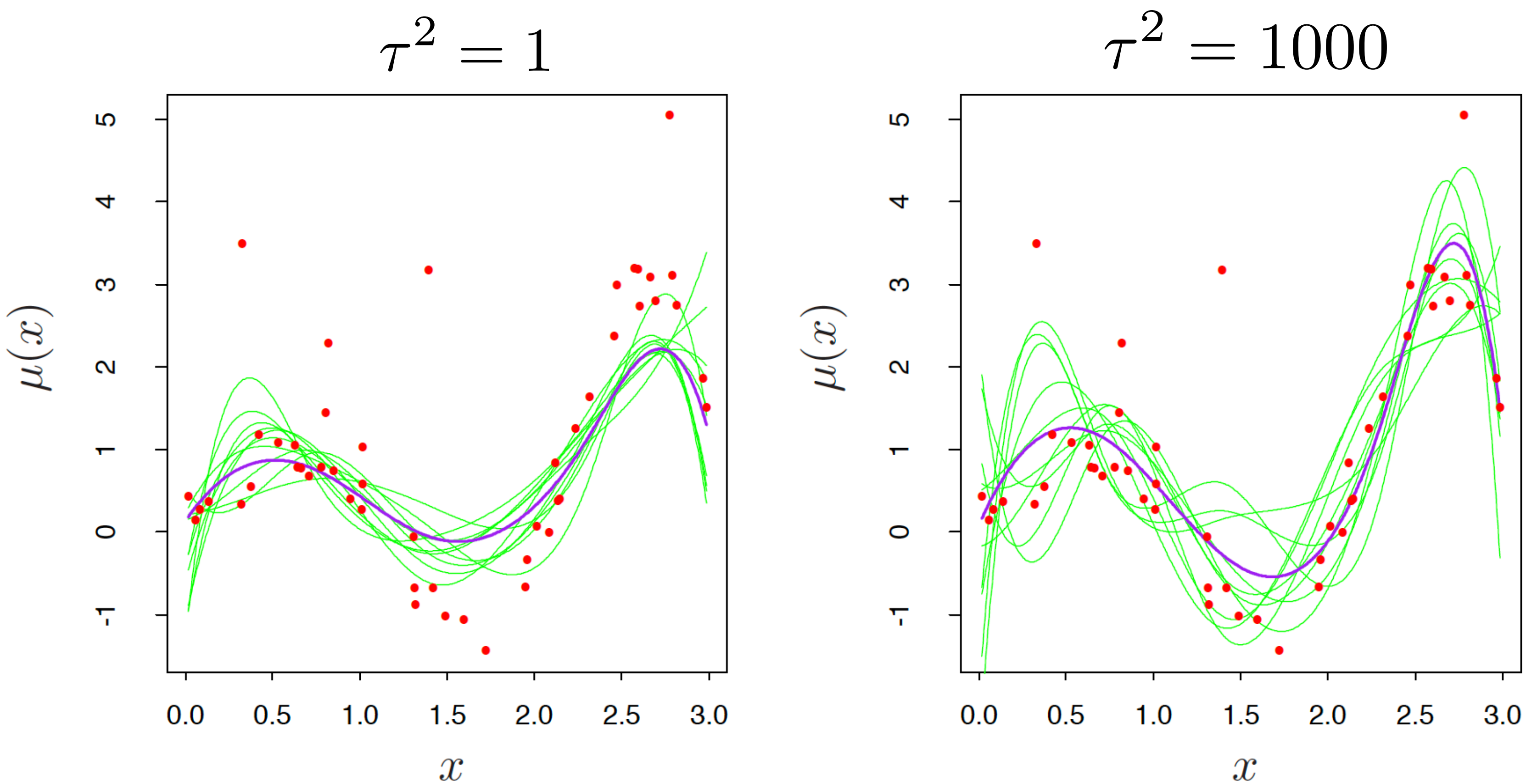
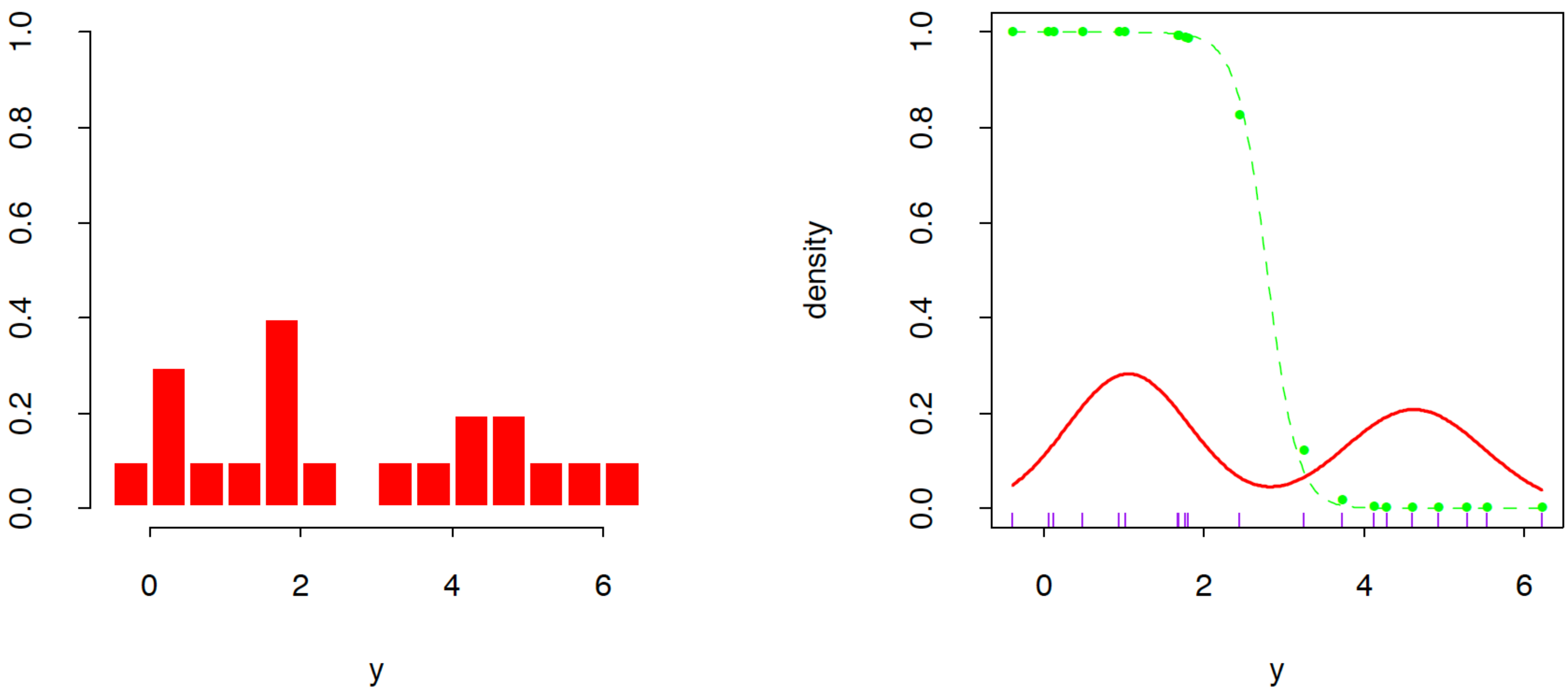


Fig 8.5

Gaussian Mixture Model with Two Components

Histogram

Maximum likelihood fit of Gaussian densities (red)
Responsibility (green)



$$\hat{\mu}_1 = 4.62, \hat{\sigma}_1^2 = 0.87$$

$$\hat{\mu}_2 = 1.06, \hat{\sigma}_2^2 = 0.77$$

$$\hat{\pi} = 0.546$$

Algorithm 8.1 *EM Algorithm for Two-component Gaussian Mixture.*

1. Take initial guesses for the parameters $\hat{\mu}_1, \hat{\sigma}_1^2, \hat{\mu}_2, \hat{\sigma}_2^2, \hat{\pi}$ (see text).

2. *Expectation Step*: compute the responsibilities

$$\hat{\gamma}_i = \frac{\hat{\pi}\phi_{\hat{\theta}_2}(y_i)}{(1-\hat{\pi})\phi_{\hat{\theta}_1}(y_i) + \hat{\pi}\phi_{\hat{\theta}_2}(y_i)}, \quad i = 1, 2, \ldots, N. \tag{8.42}$$

3. *Maximization Step*: compute the weighted means and variances:

$$\hat{\mu}_1 = \frac{\sum_{i=1}^N (1-\hat{\gamma}_i) y_i}{\sum_{i=1}^N (1-\hat{\gamma}_i)}, \qquad \hat{\sigma}_1^2 = \frac{\sum_{i=1}^N (1-\hat{\gamma}_i)(y_i - \hat{\mu}_1)^2}{\sum_{i=1}^N (1-\hat{\gamma}_i)},$$

$$\hat{\mu}_2 = \frac{\sum_{i=1}^N \hat{\gamma}_i y_i}{\sum_{i=1}^N \hat{\gamma}_i}, \qquad \hat{\sigma}_2^2 = \frac{\sum_{i=1}^N \hat{\gamma}_i (y_i - \hat{\mu}_2)^2}{\sum_{i=1}^N \hat{\gamma}_i},$$

   and the mixing probability $\hat{\pi} = \sum_{i=1}^N \hat{\gamma}_i / N$.

4. Iterate steps 2 and 3 until convergence.

The General EM algorithm

Algorithm 8.2 *The EM Algorithm.*

1. Start with initial guesses for the parameters $\hat{\theta}^{(0)}$.
2. *Expectation Step*: at the jth step, compute

$$Q(\theta', \hat{\theta}^{(j)}) = \mathrm{E}(\ell_0(\theta'; \mathbf{T})|\mathbf{Z}, \hat{\theta}^{(j)}) \tag{8.43}$$

   as a function of the dummy argument θ'.
3. *Maximization Step*: determine the new estimate $\hat{\theta}^{(j+1)}$ as the maximizer of $Q(\theta', \hat{\theta}^{(j)})$ over θ'.
4. Iterate steps 2 and 3 until convergence.