

## LECTURE 15: Linear models with normal noise

$$X_i = \sum_{j=1}^m a_{ij} \Theta_j + W_i \quad W_i, \Theta_j: \text{independent, normal}$$

- Very common and convenient model
- Bayes' rule: normal posteriors
- MAP and LMS estimates coincide
  - simple formulas  
(linear in the observations)
- Many nice properties
- Trajectory estimation example

## Recognizing normal PDFs

$$X \sim N(\mu, \sigma^2)$$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$$c \cdot e^{-8(x-3)^2}$$

$$f_X(x) = c \cdot e^{-(\alpha x^2 + \beta x + \gamma)} \quad \alpha > 0 \quad \text{Normal with mean } -\beta/2\alpha \text{ and variance } 1/2\alpha$$

## Estimating a normal random variable in the presence of additive normal noise

$$X = \Theta + W \quad \Theta, W : N(0, 1), \quad \text{independent}$$

$$f_{\Theta|X}(\theta | x) = \frac{f_{\Theta}(\theta) f_{X|\Theta}(x | \theta)}{f_X(x)}$$

$$f_X(x) = \int f_{\Theta}(\theta) f_{X|\Theta}(x | \theta) d\theta$$

$$f_{X|\Theta}(x | \theta) :$$

$$f_{\Theta|X}(\theta | x) =$$

$$\hat{\theta}_{\text{MAP}} = \hat{\theta}_{\text{LMS}} = \mathbb{E}[\Theta | X = x] =$$

$$\widehat{\Theta}_{\text{MAP}} = \mathbb{E}[\Theta | X] =$$

## Estimating a normal random variable in the presence of additive normal noise

$$X = \Theta + W \quad \Theta, W : N(0, 1), \text{ independent}$$

$$\widehat{\Theta}_{\text{MAP}} = \widehat{\Theta}_{\text{LMS}} = \mathbb{E}[\Theta | X] = \frac{X}{2}$$

$$f_{\Theta|X}(\theta | x) = \frac{f_{\Theta}(\theta) f_{X|\Theta}(x | \theta)}{f_X(x)}$$

$$f_X(x) = \int f_{\Theta}(\theta) f_{X|\Theta}(x | \theta) d\theta$$

- Even with general means and variances:
  - posterior is normal
  - LMS and MAP estimators coincide
  - these estimators are “linear,” of the form  $\widehat{\Theta} = aX + b$

## The case of multiple observations

$$\begin{aligned} X_1 &= \Theta + W_1 & \Theta &\sim N(x_0, \sigma_0^2) & W_i &\sim N(0, \sigma_i^2) \\ &\vdots \\ X_n &= \Theta + W_n & \Theta, W_1, \dots, W_n &\text{ independent} \end{aligned}$$

$$f_{X_i|\Theta}(x_i | \theta) =$$

$$f_{X|\Theta}(x | \theta) =$$

$$f_{\Theta|X}(\theta | x) =$$

$$f_{\Theta|X}(\theta | x) = \frac{f_{\Theta}(\theta) f_{X|\Theta}(x | \theta)}{f_X(x)}$$

$$f_X(x) = \int f_{\Theta}(\theta) f_{X|\Theta}(x | \theta) d\theta$$

## The case of multiple observations

$$f_{\Theta|X}(\theta|x) = c \cdot \exp \left\{ -\text{quad}(\theta) \right\} \quad \text{quad}(\theta) = \frac{(\theta - x_0)^2}{2\sigma_0^2} + \frac{(\theta - x_1)^2}{2\sigma_1^2} + \cdots + \frac{(\theta - x_n)^2}{2\sigma_n^2}$$

$$\hat{\theta}_{\text{MAP}} = \hat{\theta}_{\text{LMS}} = \mathbb{E}[\Theta | X = x] = \frac{\sum_{i=0}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^n \frac{1}{\sigma_i^2}}$$

## The case of multiple observations

- Key conclusions:
  - posterior is normal
  - LMS and MAP estimates coincide
  - these estimates are “linear,” of the form  $\hat{\theta} = a_0 + a_1x_1 + \cdots + a_nx_n$
- Interpretations:
  - estimate  $\hat{\theta}$ : weighted average of  $x_0$  (prior mean) and  $x_i$  (observations)
  - weights determined by variances

$$\hat{\theta}_{\text{MAP}} = \hat{\theta}_{\text{LMS}} = \mathbf{E}[\Theta | X = x] = \frac{\sum_{i=0}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^n \frac{1}{\sigma_i^2}}$$

## The mean squared error

$$f_{\Theta|X}(\theta | x) = c \cdot \exp \left\{ -\text{quad}(\theta) \right\}$$

$$\text{quad}(\theta) = \frac{(\theta - x_0)^2}{2\sigma_0^2} + \frac{(\theta - x_1)^2}{2\sigma_1^2} + \dots + \frac{(\theta - x_n)^2}{2\sigma_n^2}$$

$$\hat{\theta} = \frac{\sum_{i=0}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^n \frac{1}{\sigma_i^2}}$$

- Performance measures:

$$\mathbf{E}[(\Theta - \widehat{\Theta})^2 | X = x] = \mathbf{E}[(\Theta - \hat{\theta})^2 | X = x] = \text{var}(\Theta | X = x) = 1 / \sum_{i=0}^n \frac{1}{\sigma_i^2}$$

$$\mathbf{E}[(\Theta - \widehat{\Theta})^2] =$$

$$f_X(x) = c \cdot e^{-(\alpha x^2 + \beta x + \gamma)} \quad \alpha > 0 \quad \text{Normal with mean } -\beta/2\alpha \text{ and variance } 1/2\alpha$$

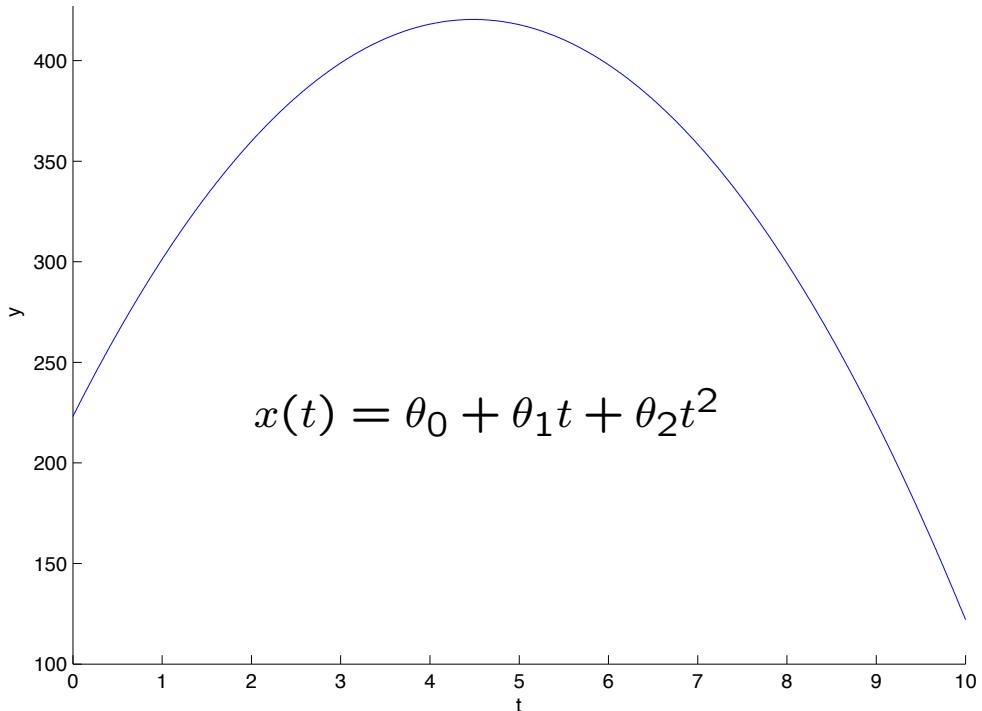
## The mean squared error

$$\mathbb{E}[(\Theta - \widehat{\Theta})^2 | X = x] = \mathbb{E}[(\Theta - \widehat{\Theta})^2] = 1 / \sum_{i=0}^n \frac{1}{\sigma_i^2}$$

$$\widehat{\theta} = \frac{\sum_{i=0}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^n \frac{1}{\sigma_i^2}}$$

- Example:  $\sigma_0^2 = \sigma_1^2 = \dots = \sigma_n^2 = \sigma^2$
- conditional mean squared error same for all  $x$
- Example:  $X = \Theta + W$      $\Theta \sim N(0, 1)$ ,     $W \sim N(0, 1)$   
independent  $\Theta, W$        $\widehat{\Theta} = X/2$        $\mathbb{E}[(\Theta - \widehat{\Theta})^2 | X = x] =$

## The case of multiple parameters: trajectory estimation



- Random variables  $\Theta_0, \Theta_1, \Theta_2$  independent; priors  $f_{\Theta_j}$
- Measurements at times  $t_1, \dots, t_n$   
 $X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$   
noise model:  $f_{W_i}$   
independent  $W_i$ ; independent from  $\Theta_j$

## A model with normality assumptions

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i \quad i = 1, \dots, n$$

$$f_{\Theta|X}(\theta | x) = \frac{f_{\Theta}(\theta) f_{X|\Theta}(x | \theta)}{f_X(x)}$$

$$f_X(x) = \int f_{\Theta}(\theta) f_{X|\Theta}(x | \theta) d\theta$$

- assume  $\Theta_j \sim N(0, \sigma_j^2)$ ,  $W_i \sim N(0, \sigma^2)$ ; independent
- Given  $\Theta = \theta = (\theta_0, \theta_1, \theta_2)$ ,  $X_i$  is:

$$f_{X_i|\Theta}(x_i | \theta) = c \cdot \exp \left\{ - (x_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2)^2 / 2\sigma^2 \right\}$$

- posterior:  $f_{\Theta|X}(\theta | x) =$

$$c(x) \exp \left\{ - \frac{1}{2} \left( \frac{\theta_0^2}{\sigma_0^2} + \frac{\theta_1^2}{\sigma_1^2} + \frac{\theta_2^2}{\sigma_2^2} \right) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2)^2 \right\}$$

## A model with normality assumptions

$$f_{\Theta|X}(\theta|x) = c(x) \exp \left\{ -\frac{1}{2} \left( \frac{\theta_0^2}{\sigma_0^2} + \frac{\theta_1^2}{\sigma_1^2} + \frac{\theta_2^2}{\sigma_2^2} \right) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2)^2 \right\}$$

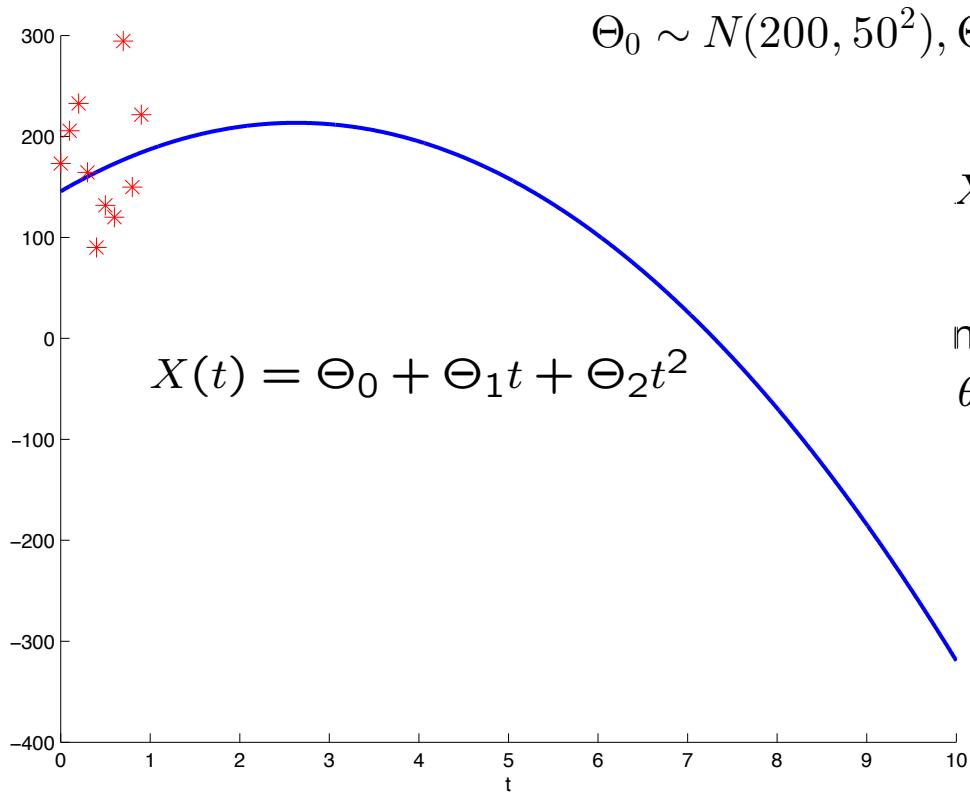
- MAP estimate: maximize over  $(\theta_0, \theta_1, \theta_2)$ ;  
(minimize quadratic function)

## Linear normal models

- $\Theta_j$  and  $X_i$  are linear functions of independent normal random variables
- $f_{\Theta|X}(\theta|x) = c(x) \exp\left\{-\text{quadratic}(\theta_1, \dots, \theta_m)\right\}$
- MAP estimate: maximize over  $(\theta_1, \dots, \theta_m)$ ;  
(minimize quadratic function)  
 $\widehat{\Theta}_{\text{MAP},j}$ : linear function of  $X = (X_1, \dots, X_n)$
- Facts:
  - $\widehat{\Theta}_{\text{MAP},j} = \mathbf{E}[\Theta_j | X]$
  - marginal posterior PDF of  $\Theta_j$ :  $f_{\Theta_j|X}(\theta_j | x)$ , is normal
  - MAP estimate based on the joint posterior PDF:  
same as MAP estimate based on the marginal posterior PDF
  - $\mathbf{E}[(\widehat{\Theta}_{i,\text{MAP}} - \Theta_i)^2 | X = x]$ : same for all  $x$

## An illustration

Estimating the trajectory of a free-falling object



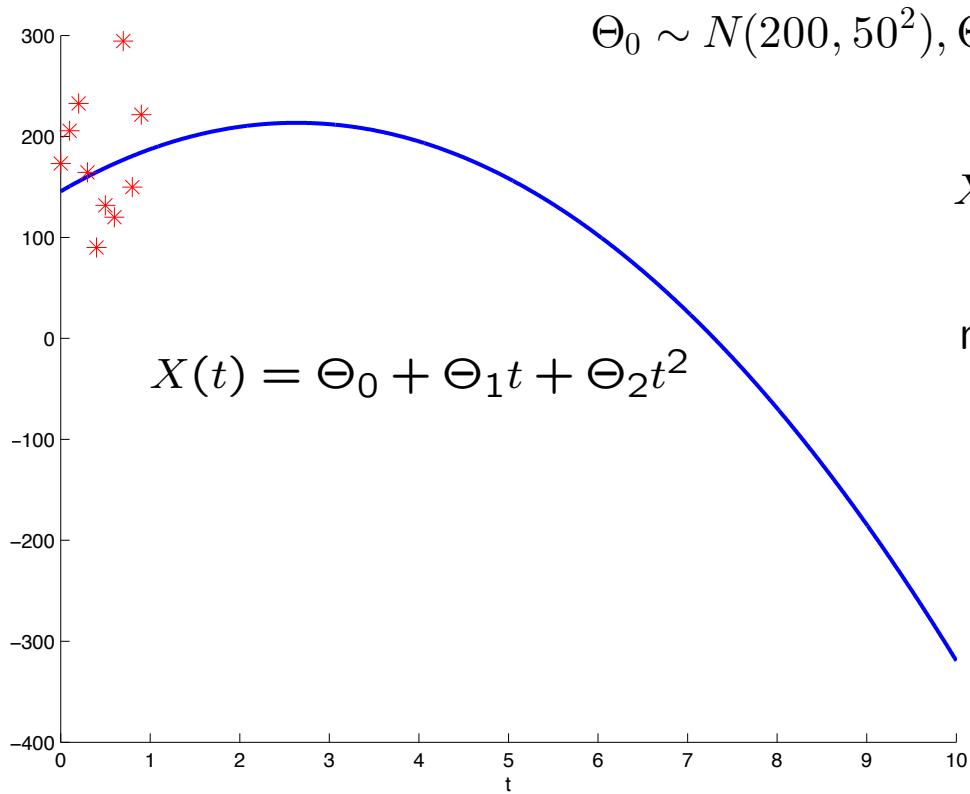
$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

$$\begin{aligned} & \text{minimize}_{\theta_0, \theta_1, \theta_2} \quad \frac{1}{2} \left( \frac{\theta_0^2}{\sigma_0^2} + \frac{\theta_1^2}{\sigma_1^2} + \frac{\theta_2^2}{\sigma_2^2} \right) \\ & + \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2)^2 \end{aligned}$$

## An illustration

Estimating the trajectory of a free-falling object



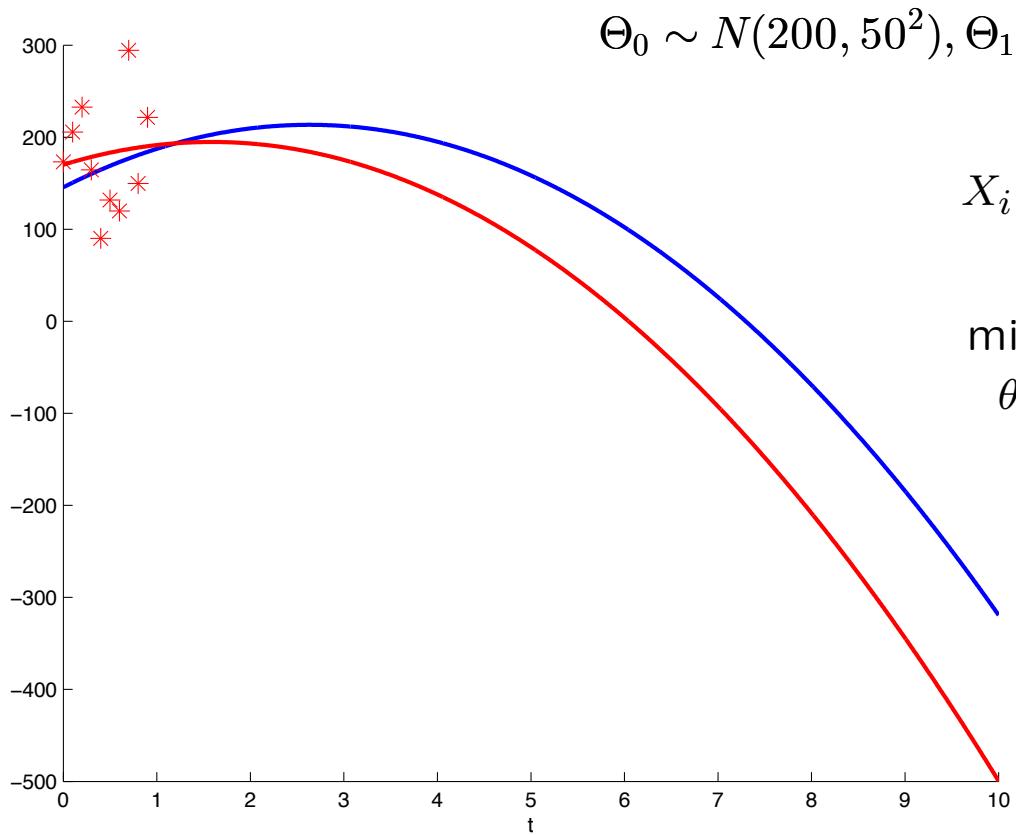
$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

$$\begin{aligned} & \text{minimize}_{\theta_0, \theta_1} && (\theta_0 - 200)^2 + (\theta_1 - 50)^2 \\ & && + \sum_{i=1}^n (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2 \end{aligned}$$

## An illustration

Estimating the trajectory of a free-falling object



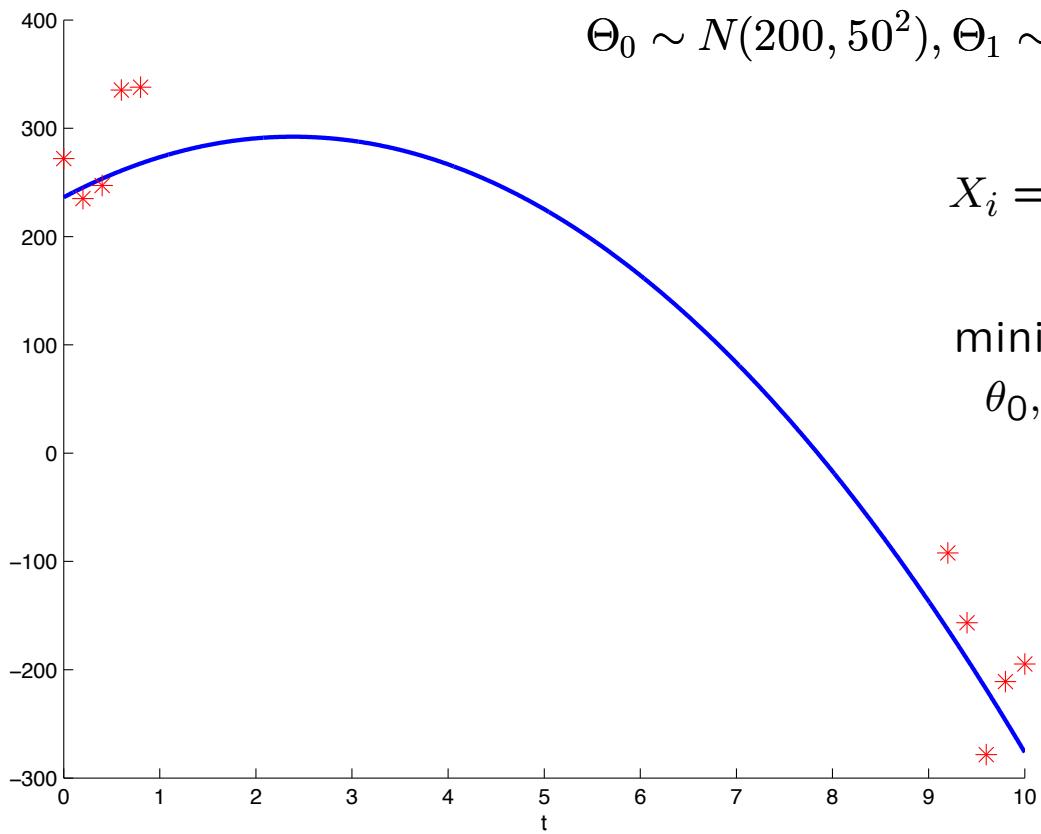
$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

$$\begin{aligned} & \text{minimize}_{\theta_0, \theta_1} && (\theta_0 - 200)^2 + (\theta_1 - 50)^2 \\ & && + \sum_{i=1}^n (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2 \end{aligned}$$

## An illustration

Estimating the trajectory of a free-falling object



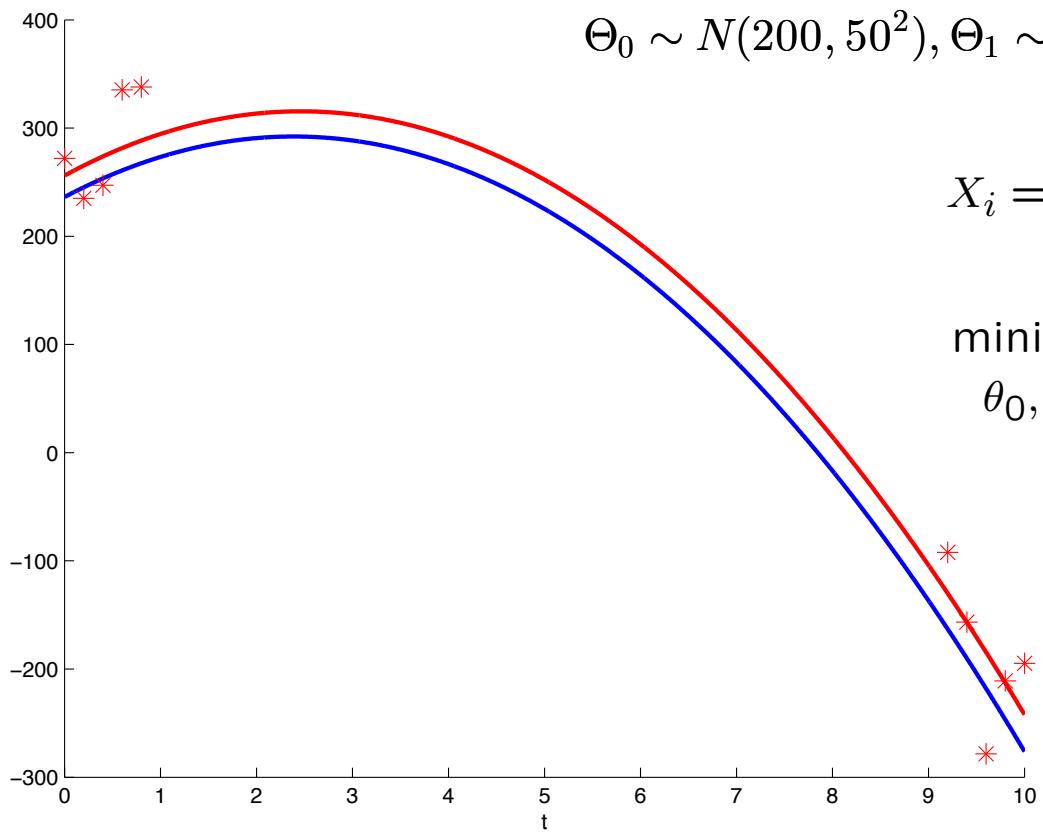
$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

$$\begin{aligned} & \text{minimize}_{\theta_0, \theta_1} && (\theta_0 - 200)^2 + (\theta_1 - 50)^2 \\ & && + \sum_{i=1}^n (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2 \end{aligned}$$

## An illustration

Estimating the trajectory of a free-falling object



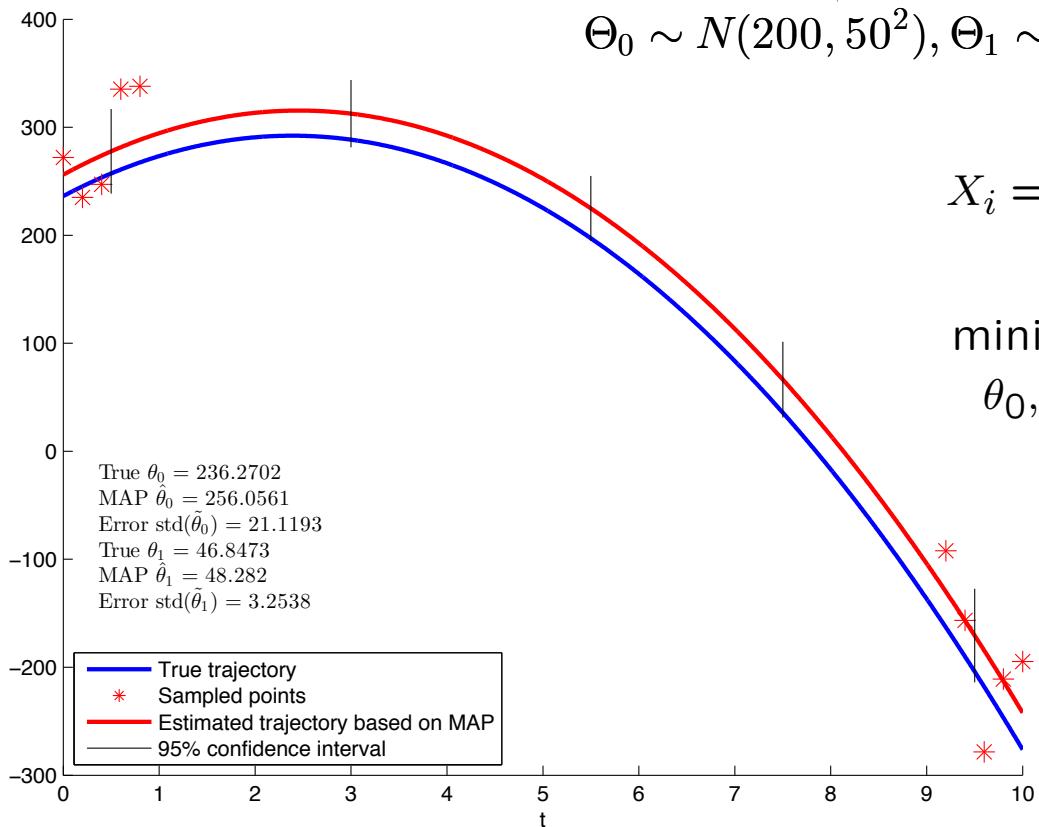
$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

$$\begin{aligned} & \text{minimize} && (\theta_0 - 200)^2 + (\theta_1 - 50)^2 \\ & \theta_0, \theta_1 && + \sum_{i=1}^n (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2 \end{aligned}$$

## An illustration

Estimating the trajectory of a free-falling object



$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

$$\begin{aligned} & \text{minimize}_{\theta_0, \theta_1} && (\theta_0 - 200)^2 + (\theta_1 - 50)^2 \\ & && + \sum_{i=1}^n (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2 \end{aligned}$$