

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

弹道

Recognizing normal PDFs

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

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

$$2\alpha x + \beta = 0$$

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

$$\mu = 3$$

$$\frac{1}{2\sigma^2} = 8 \Rightarrow \sigma^2 = \frac{1}{16}$$

$$c = \frac{1}{\frac{1}{4}\sqrt{2\pi}}$$

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

$$\alpha x^2 + \beta x + \gamma = \alpha \left(x^2 + \frac{\beta}{\alpha}x + \frac{\gamma}{\alpha} \right) = \alpha \left(\left(x + \frac{\beta}{2\alpha} \right)^2 - \frac{\beta^2}{4\alpha^2} + \frac{\gamma}{\alpha} \right)$$

$$f_X(x) = c \underbrace{e^{-\alpha \left(x + \frac{\beta}{2\alpha} \right)^2} e^{-\alpha \left(-\frac{\beta^2}{4\alpha^2} + \frac{\gamma}{\alpha} \right)}}_{\text{constant}}$$

$$\mu = -\frac{\beta}{2\alpha}$$

$$\frac{1}{2\sigma^2} = \alpha \Rightarrow \sigma^2 = 1/2\alpha$$

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

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

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

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

$$f_{X|\Theta}(x | \theta) : X = \theta + W \quad N(\theta, 1)$$

$$f_{\Theta|X}(\theta | x) = \frac{1}{f_X(x)} c e^{-\frac{1}{2}\theta^2} c e^{-\frac{1}{2}(x-\theta)^2} = \underline{\underline{c(x)}} e^{-\text{quadratic}(\theta)}$$

Fix x $\min_{\theta} \left[\frac{1}{2}\theta^2 + \frac{1}{2}(x-\theta)^2 \right]$ $\theta + (\theta - x) = 0$ Normal!

$$\hat{\theta}_{\text{MAP}} = \hat{\theta}_{\text{LMS}} = \mathbf{E}[\Theta | X = x] = x/2$$

$$\hat{\Theta}_{\text{MAP}} = \mathbf{E}[\Theta | X] = x/2$$

Estimating a normal parameter in the presence of additive normal noise

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

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

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

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

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

The case of multiple observations

$$X_1 = \Theta + W_1 \quad \Theta \sim N(x_0, \sigma_0^2) \quad W_i \sim N(0, \sigma_i^2)$$

$$\vdots$$
$$X_n = \Theta + W_n \quad \Theta, W_1, \dots, W_n \text{ independent}$$

$$f_{X_i|\Theta}(x_i|\theta) = c_i e^{-(x_i - \theta)^2 / 2\sigma_i^2}$$

$$\text{given } \Theta = \theta: X_i = \theta + W_i \sim N(\theta, \sigma_i^2)$$

$$f_{X|\Theta}(x|\theta) = f_{x_1, \dots, x_n|\Theta}(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f_{x_i|\Theta}(x_i|\theta)$$

$$\text{given } \Theta = \theta: W_i \text{ independent} \Rightarrow X_i \text{ independent}$$

$$f_{\Theta|X}(\theta|x) = \frac{1}{f_X(x)} \cdot c_0 e^{-(\theta - x_0)^2 / 2\sigma_0^2} \prod_{i=1}^n c_i e^{-(x_i - \theta)^2 / 2\sigma_i^2}$$

Normal!

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

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

The case of multiple observations

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

find peak

$$\frac{d}{d\theta} \text{quad}(\theta) = 0: \quad \sum_{i=0}^n \frac{(\theta - x_i)}{\sigma_i^2} = 0 \Rightarrow \theta \sum_{i=0}^n \frac{1}{\sigma_i^2} = \sum_{i=0}^n \frac{x_i}{\sigma_i^2}$$

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

The case of multiple observations

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

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

σ_i^2 large
 x_i very noisy
 \Rightarrow small weight

The mean squared error

$$X_i = \Theta + W_i$$

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

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

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

- Performance measures:

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

$$\mathbf{E}[(\Theta - \hat{\Theta})^2] = \int \underbrace{\mathbf{E}[(\Theta - \hat{\Theta})^2 | X = x]}_{\text{MSE}} f_X(x) dx$$

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

$$\alpha = \frac{1}{2\sigma_0^2} + \dots + \frac{1}{2\sigma_n^2}$$

some σ_i^2 small \rightarrow MSE small
all σ_i^2 large \rightarrow MSE large

The mean squared error

$$\mathbb{E}[(\Theta - \hat{\Theta})^2 \mid X = \underline{x}] \stackrel{!}{=} \mathbb{E}[(\Theta - \hat{\Theta})^2] = 1 / \sum_{i=0}^n \frac{1}{\sigma_i^2}$$

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

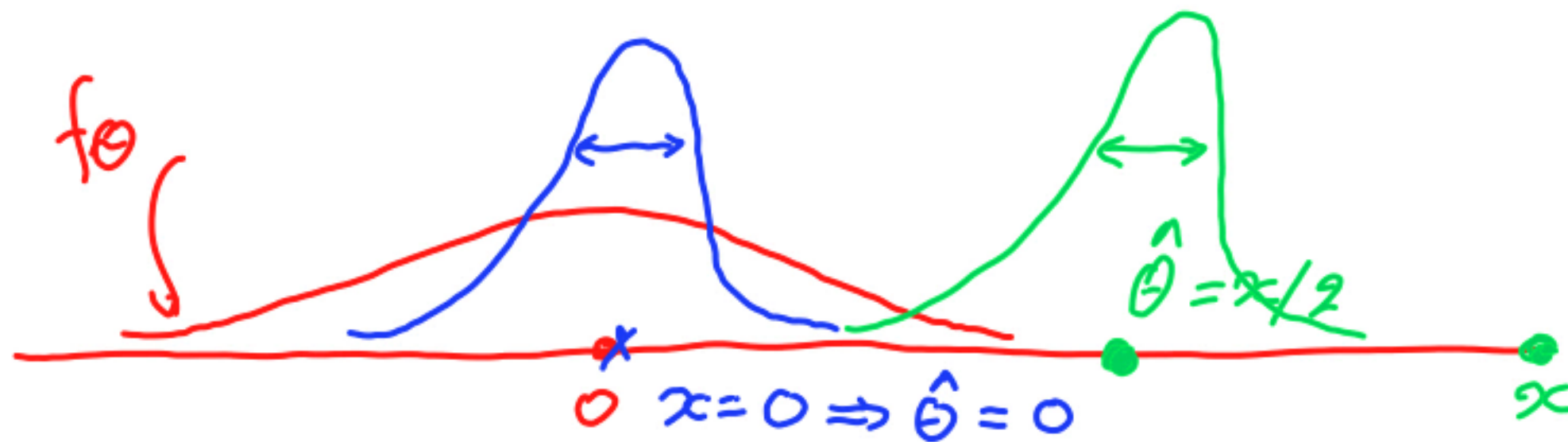
- Example: $\sigma_0^2 = \sigma_1^2 = \dots = \sigma_n^2 = \sigma^2$ $\frac{1}{(n+1)\sigma^2} = \frac{\sigma^2}{n+1}$
- conditional mean squared error same for all x

- Example: $X = \Theta + W$ $\Theta \sim N(0, \underline{1})$, $W \sim N(0, 1)$

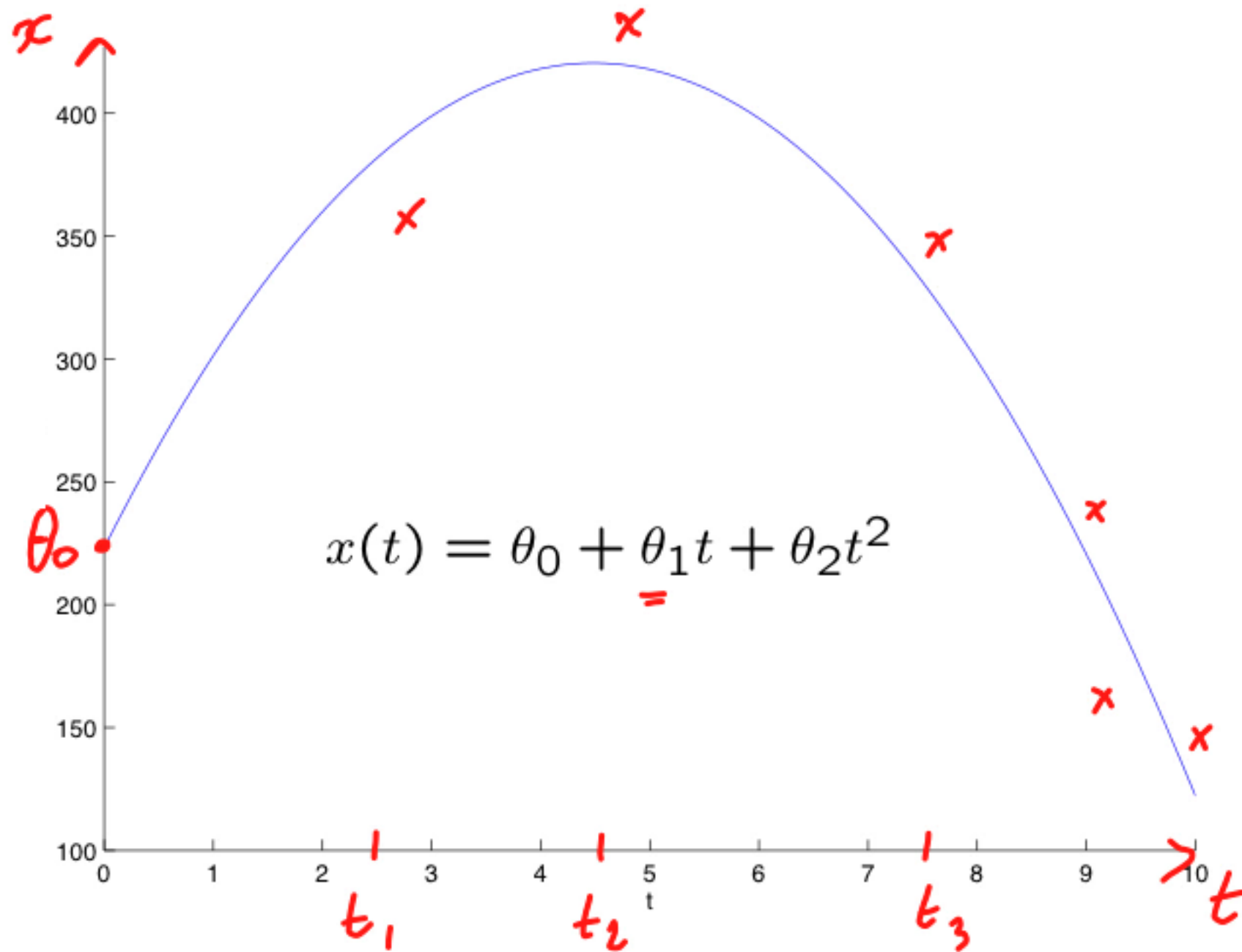
independent Θ, W

$$\hat{\Theta} = X/2$$

$$\mathbb{E}[(\Theta - \hat{\Theta})^2 \mid X = \underline{x}] = \underline{\underline{1/2}}$$



The case of multiple parameters: trajectory estimation



- Random variables $\Theta_0, \Theta_1, \Theta_2$
independent; priors f_{Θ_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 Θ_j

A model with normality assumptions

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

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

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

- 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: $N(\theta_0 + \theta_1 t_i + \theta_2 t_i^2, \sigma^2)$

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

- posterior: $f_{\Theta|X}(\theta | x) = \frac{1}{f_X(x)} \prod_{j=0}^2 f_{\Theta_j}(\theta_j) \prod_{i=1}^n f_{X_i|\Theta}(x_i | \theta)$

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

A model with normality assumptions

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

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

$$\frac{\partial}{\partial \theta_j} (\text{quad}(\theta)) = 0$$

3 equations, 3 unknowns.
↑
linear

Linear normal models •

- Θ_j and X_i are linear functions of independent normal random variables
- $f_{\Theta|X}(\theta | x) = c(x) \exp \{ - \text{quadratic}(\theta_1, \dots, \theta_m) \}$ *linear regression*
- MAP estimate: maximize over $(\theta_1, \dots, \theta_m)$; *linear equations*
(minimize quadratic function)

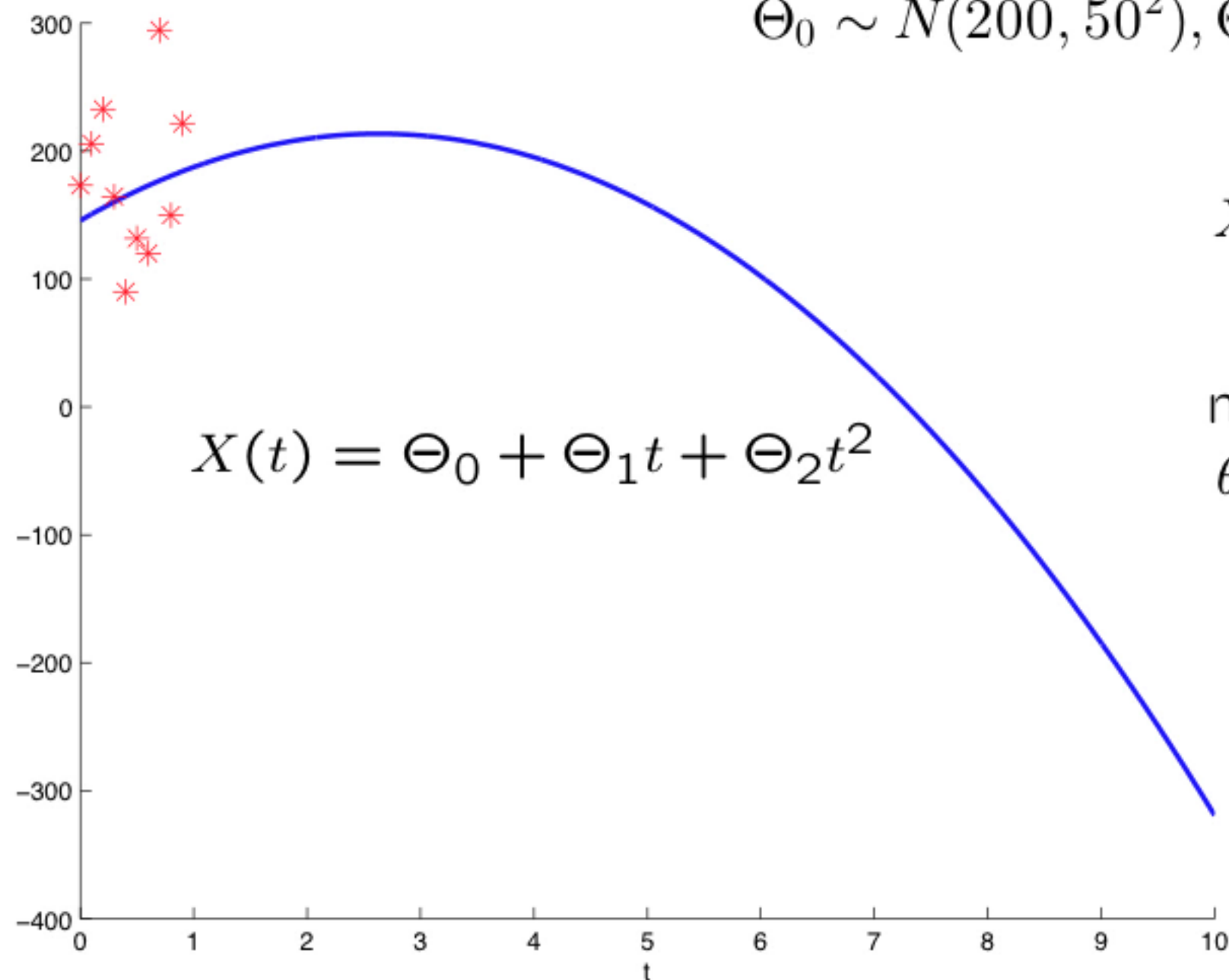
$\hat{\Theta}_{\text{MAP},j}$: linear function of $X = (X_1, \dots, X_n)$

- Facts:
 - $\hat{\Theta}_{\text{MAP},j} = \mathbf{E}[\Theta_j | X]$
 - marginal posterior PDF of Θ_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
 - $\mathbf{E}[(\hat{\Theta}_{i,\text{MAP}} - \Theta_i)^2 | X = x]$: same for all x

An illustration

Estimating the trajectory of a free-falling object

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



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

minimize
 $\theta_0, \theta_1, \theta_2$

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

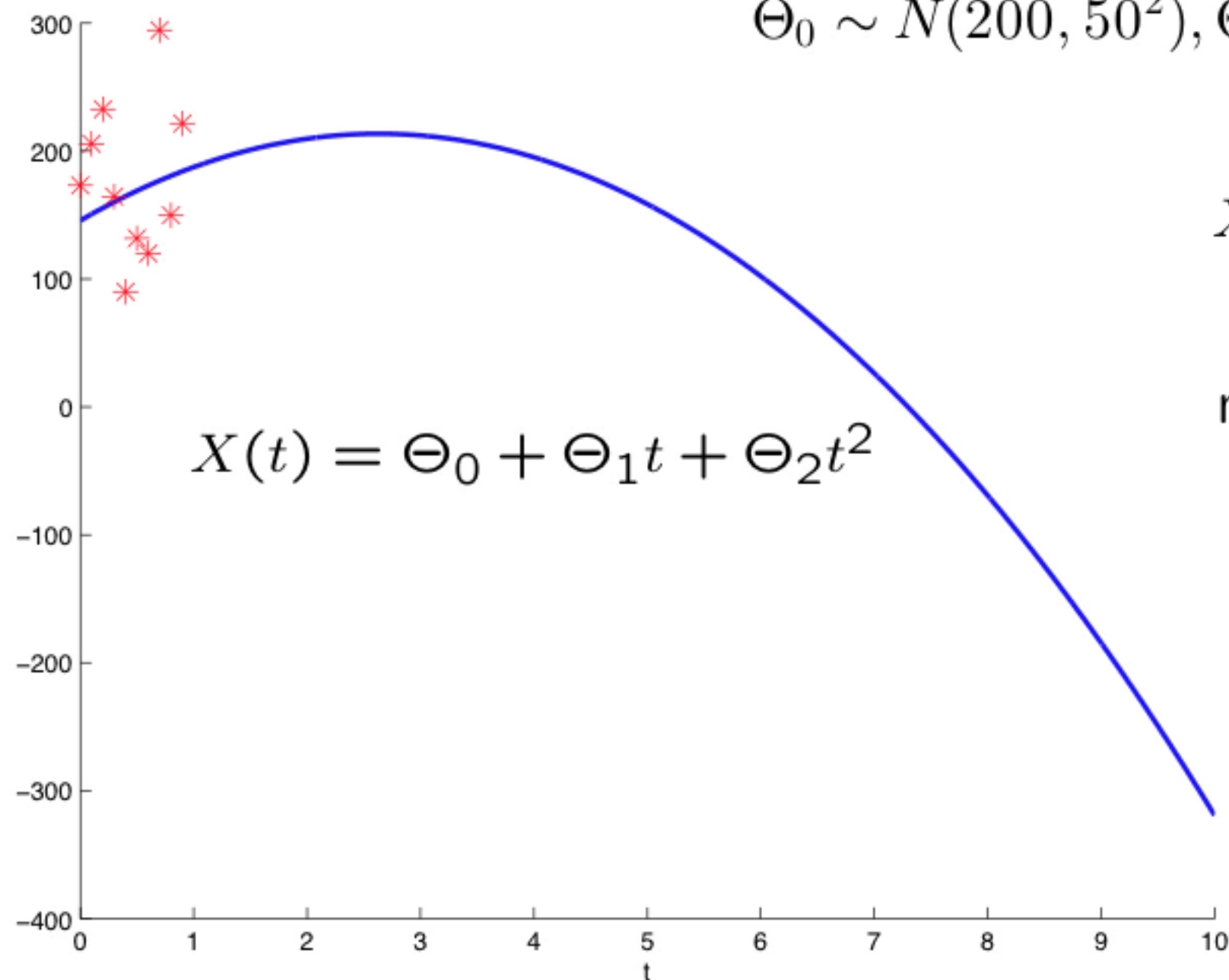
An illustration

Estimating the trajectory of a free-falling object

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

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

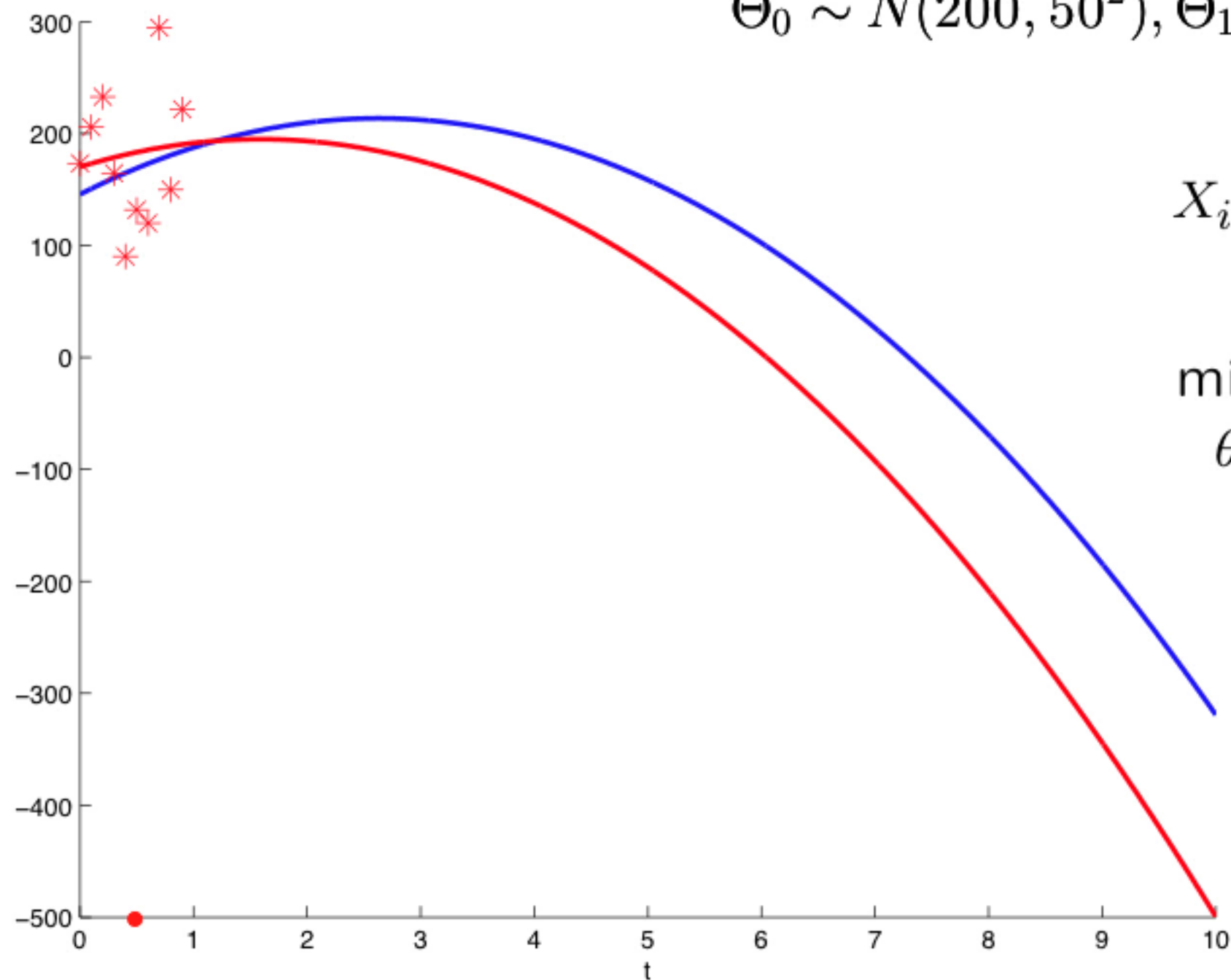
$$\begin{aligned} \text{minimize} \quad & (\theta_0 - 200)^2 + (\theta_1 - 50)^2 \\ & + \sum_{i=1}^n (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2 \end{aligned}$$



An illustration

Estimating the trajectory of a free-falling object

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



$$X_i = \underbrace{\Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2}_{\text{estimated trajectory}} + W_i$$

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

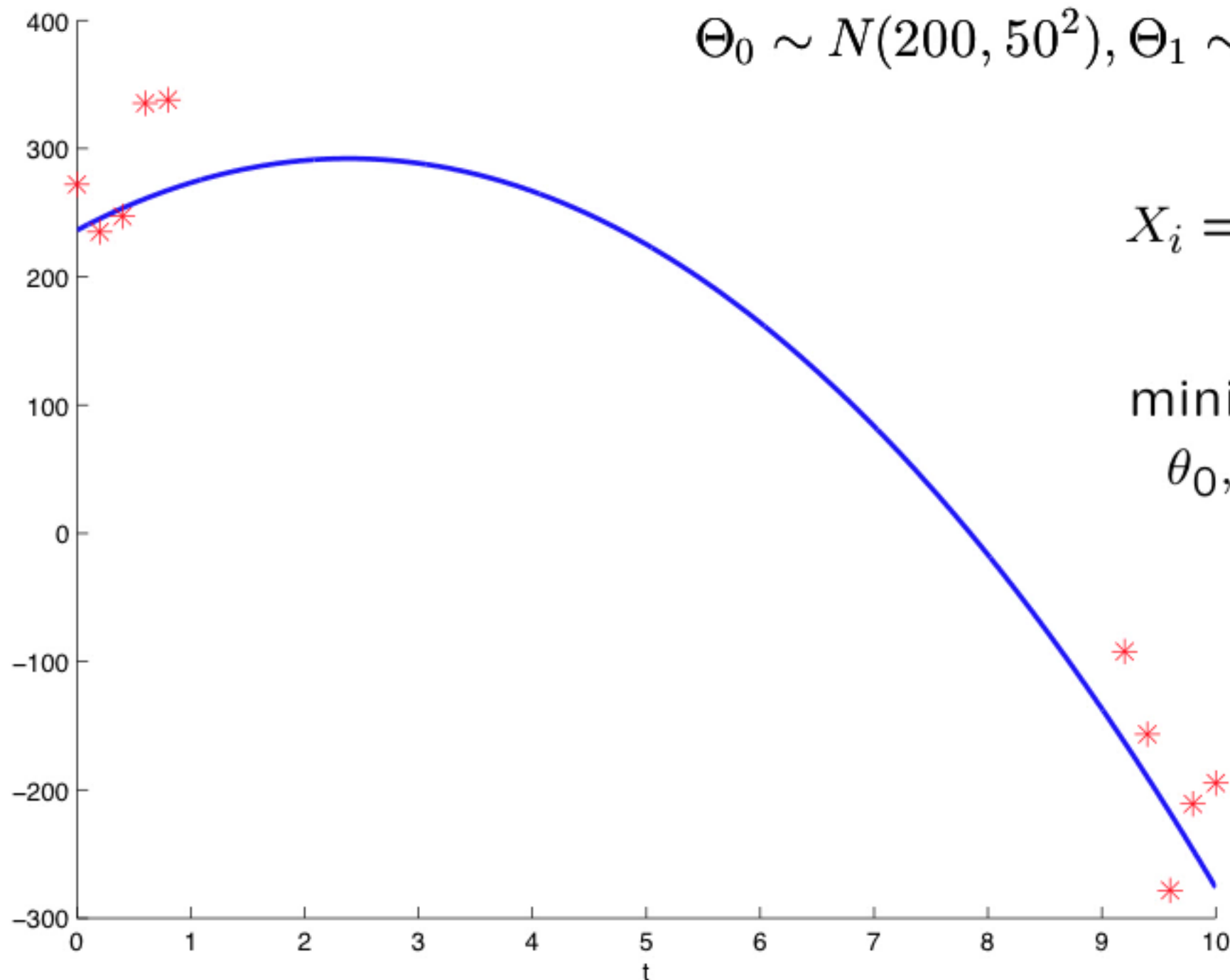
An illustration

Estimating the trajectory of a free-falling object

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

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

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



An illustration

Estimating the trajectory of a free-falling object

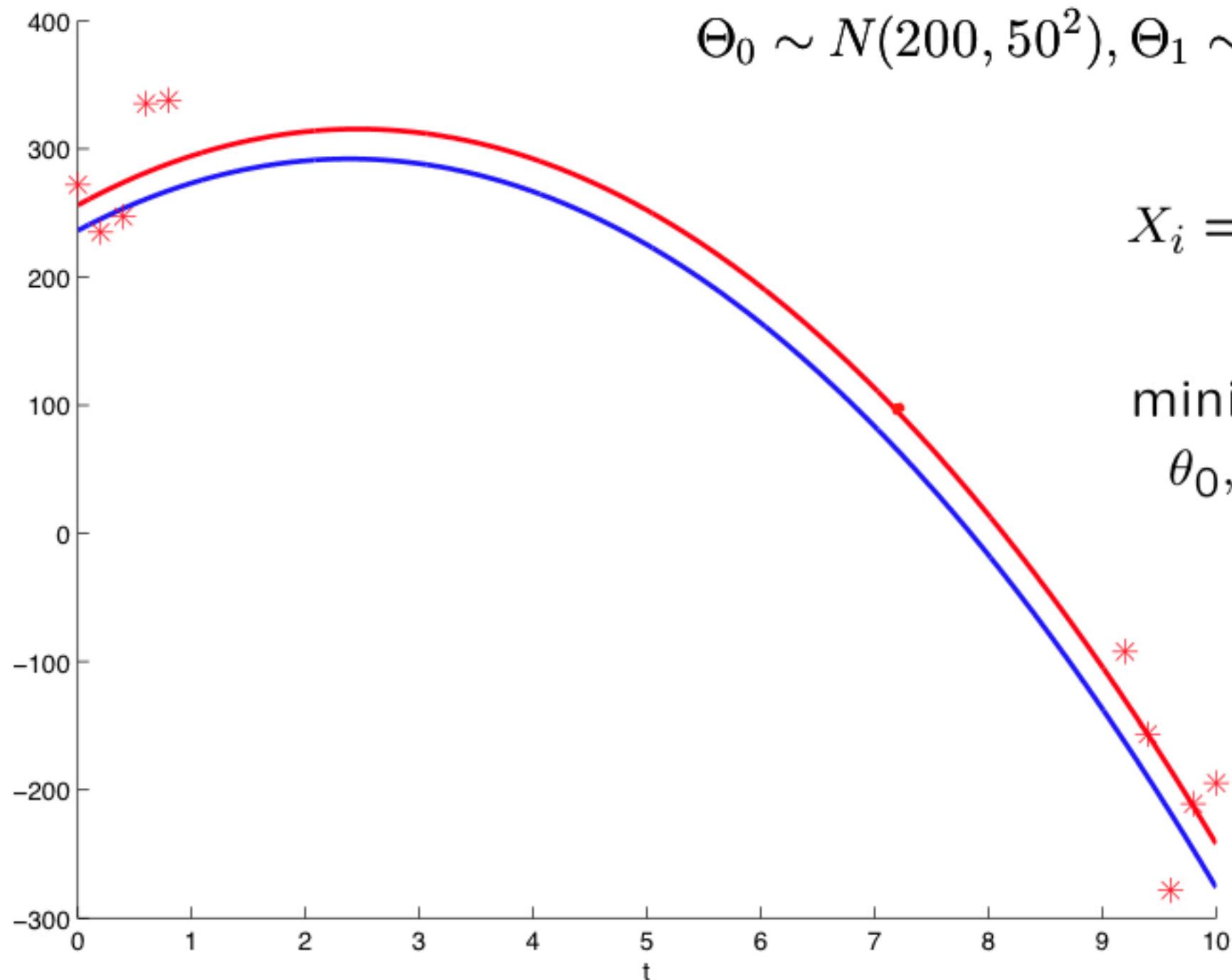
$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

$x(t)$

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

minimize θ_0, θ_1

$$(\theta_0 - 200)^2 + (\theta_1 - 50)^2 + \sum_{i=1}^n (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2$$



An illustration

Estimating the trajectory of a free-falling object

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

$$X_i = \overbrace{\Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2}^{x(t)} + W_i$$

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

$$P(x(t) \in \text{interval} | \text{data}) = 0.95$$

