LECTURE 15: Linear models with normal noise

$$X_i = \sum_{j=1}^{m} a_{ij}\Theta_j + W_i$$

W_i, Θ_j: 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

$X \sim N(\mu, \sigma^2)$ $\qquad$ $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ $\qquad$ $2\alpha x + \beta = 0$

$c \cdot e^{-8(x-3)^2}$ $\qquad$ $\mu = 3$ $\qquad$ $\frac{1}{2\sigma^2} = 8 \Rightarrow \sigma^2 = \frac{1}{16}$ $\qquad$ $c = \frac{1}{\frac{1}{4}\sqrt{2\pi}}$

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

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

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

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

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

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

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

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

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

找到theta是constant的时候，W的pdf是什么

$$f_{X|\Theta}(x \mid \theta) : \quad X = \underset{\text{constant}}{\theta} + W \qquad W \sim N(\theta, 1)$$

$$f_{\Theta|X}(\theta \mid x) = \frac{1}{f_X(x)}\, c\, e^{-\frac{1}{2}\theta^2} \underbrace{c\, e^{-\frac{1}{2}(x-\theta)^2}}_{\text{已知theta的情况下，W的pdf用X来表示}} = \underline{\underline{c(x)}}\, e^{-\text{quadratic}(\theta)}$$

Normal!

最大整个式子，所以把所有的指数项相加，求极值

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

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

$$\widehat{\Theta}_{\text{MAP}} = \mathbf{E}[\Theta \mid X] = X/2$$

Estimating a normal parameter in the presence of additive normal noise

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

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

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

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

只要theta和W都是Normal

- 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 $\boxed{\text{multiple}}$ observations

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

$$\vdots$$

$$X_n = \Theta + W_n \qquad \Theta, W_1, \ldots, W_n \text{ independent}$$

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

X和x是vector

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

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

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

$$f_{X|\Theta}(x \mid \theta) = f_{X_1,\ldots,X_n|\Theta}(x_1,\ldots,x_n \mid \theta) = \prod_{i=1}^{n} f_{X_i|\Theta}(x_i \mid \theta)$$

given $\Theta=\theta$: W_i independent $\Rightarrow$ X_i independent

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

Normal!

The case of multiple observations

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

find peak

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

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

The case of multiple observations

- Key conclusions:
  - posterior is normal
  - LMS and MAP estimates coincide
  - these estimates are "linear," of the form $\hat{\theta} = a_0 + a_1 x_1 + \cdots + a_n x_n$
- Interpretations:
  - estimate $\hat{\theta}$: weighted average of x_0 (prior mean) and x_i (observations)
  - weights determined by variances

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

σ_i^2 large

X_i very noisy

$\Rightarrow$ small weight

The mean squared error

$X_i = \Theta + W_i$

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

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

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

- Performance measures:

$$\mathbf{E}\left[(\Theta - \widehat{\Theta})^2 \mid X = x\right] = \mathbf{E}\left[(\Theta - \hat{\theta})^2 \mid X = x\right] = \operatorname{var}(\Theta \mid X = x) = \boxed{1 \Big/ \sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$$

因为知道了X的值是x，所有estimator的值已经决定了

$$\mathbf{E}\left[(\Theta - \widehat{\Theta})^2\right] = \int \underbrace{E\left[(\Theta - \hat{\Theta})^2 \mid X = x\right]}\, \underbrace{f_X(x)\,dx}_{=1}$$

//

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

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

some σ_i^2 small $\rightarrow$ MSE small

all σ_i^2 large $\rightarrow$ MSE large

The mean squared error

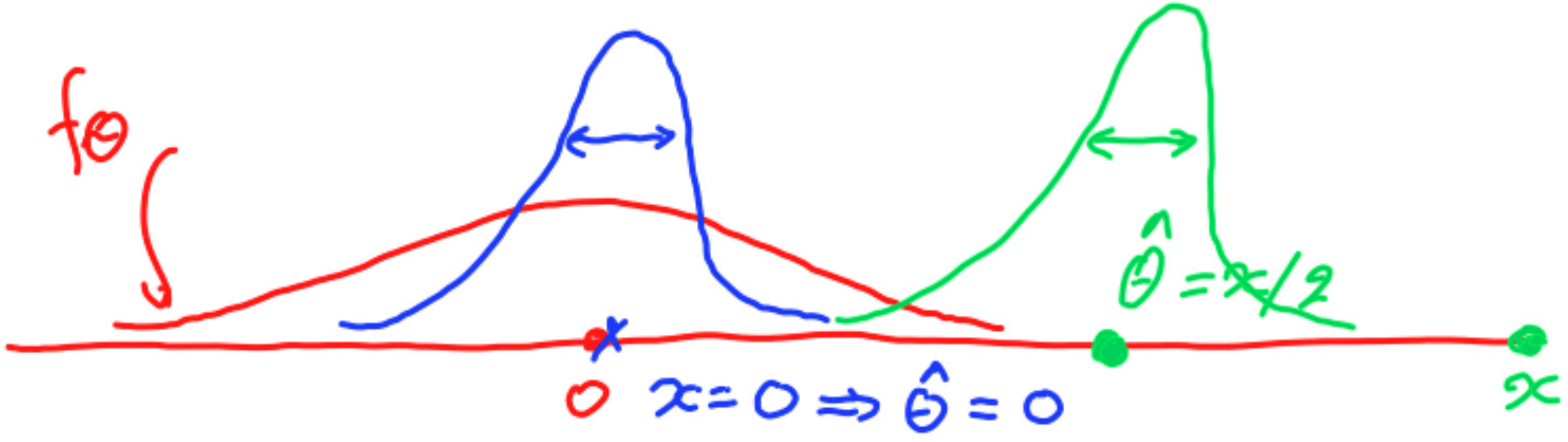
$$\mathbf{E}\big[(\Theta - \widehat{\Theta})^2 \mid X = x\big] = \mathbf{E}\big[(\Theta - \widehat{\Theta})^2\big] = 1 \Big/ \sum_{i=0}^{n} \frac{1}{\sigma_i^2}$$

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

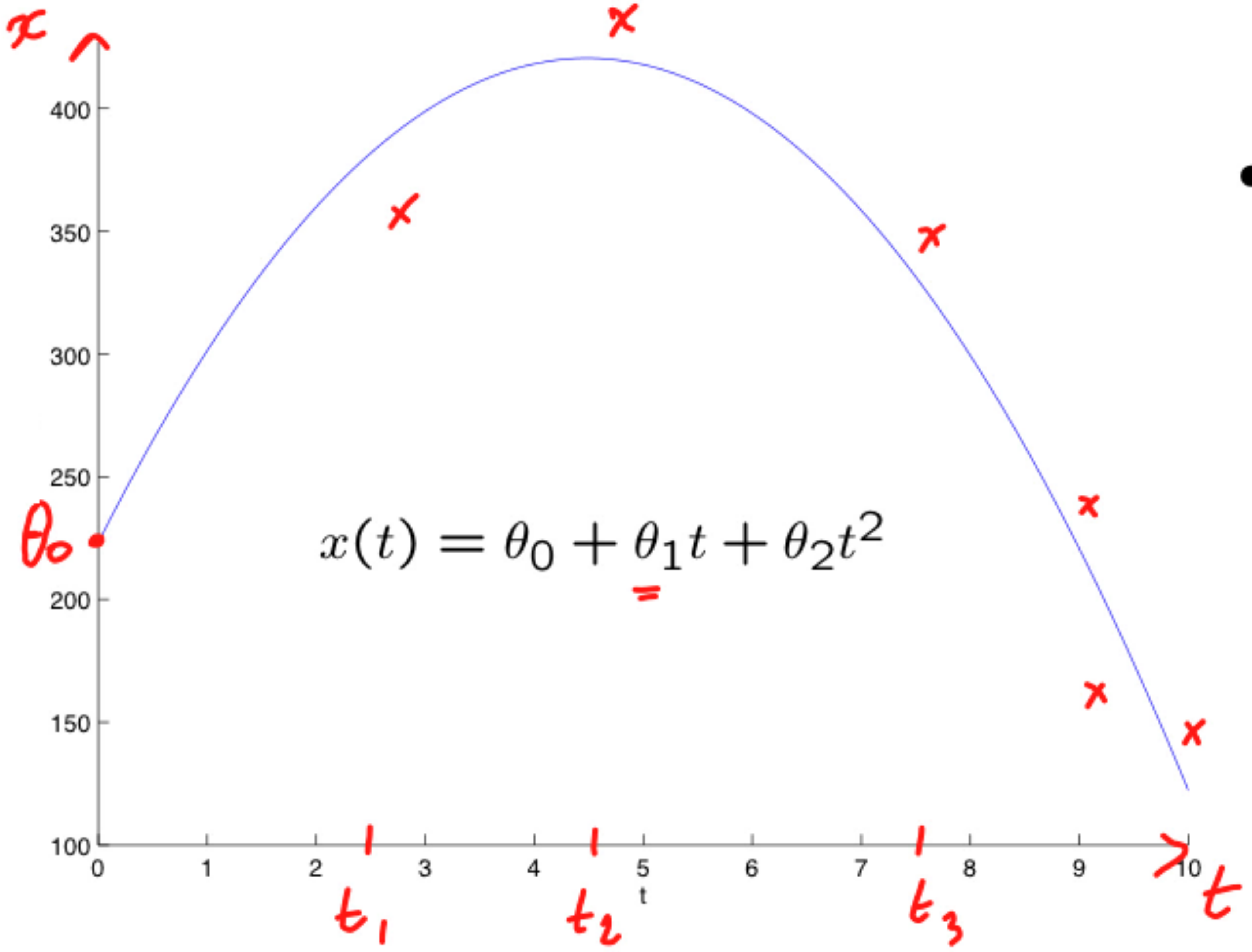
- Example: $\sigma_0^2 = \sigma_1^2 = \cdots = \sigma_n^2 = \sigma^2$ $\quad \dfrac{1}{(n+1)\frac{1}{\sigma^2}} = \dfrac{\sigma^2}{n+1}$
- conditional mean squared error same for all x 无论观测到的x值是什么，estimate的表现都一样
- Example: $X = \Theta + W \qquad \Theta \sim N(0, 1), \qquad W \sim N(0, 1)$

  independent Θ, W $\qquad \boxed{\widehat{\Theta} = X/2}$ $\qquad \mathbf{E}\big[(\Theta - \widehat{\Theta})^2 \mid X = x\big] = 1/2$

uncertainty



The case of multiple parameters: trajectory estimation



- Random variables $\Theta_0, \Theta_1, \Theta_2$

  independent; priors f_{Θ_j}

- Measurements at times $t_1, \ldots, t_n$

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

  noise model: f_{W_i}

  independent W_i; independent from Θ_j

A model with normality assumptions

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

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

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

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

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

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

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

A model with normality assumptions

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

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

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

3 equations, 3 unknowns.
linear

分别对不同的参数求偏导

Linear normal models

- Θ_j and X_i and are linear functions of independent normal random variables

- $f_{\Theta|X}(\theta \,|\, x) = c(x) \exp\left\{ - \text{quadratic}(\theta_1, \ldots, \theta_m) \right\}$ linear regression

- MAP estimate: maximize over $(\theta_1, \ldots \theta_m)$; (minimize quadratic function) linear equations

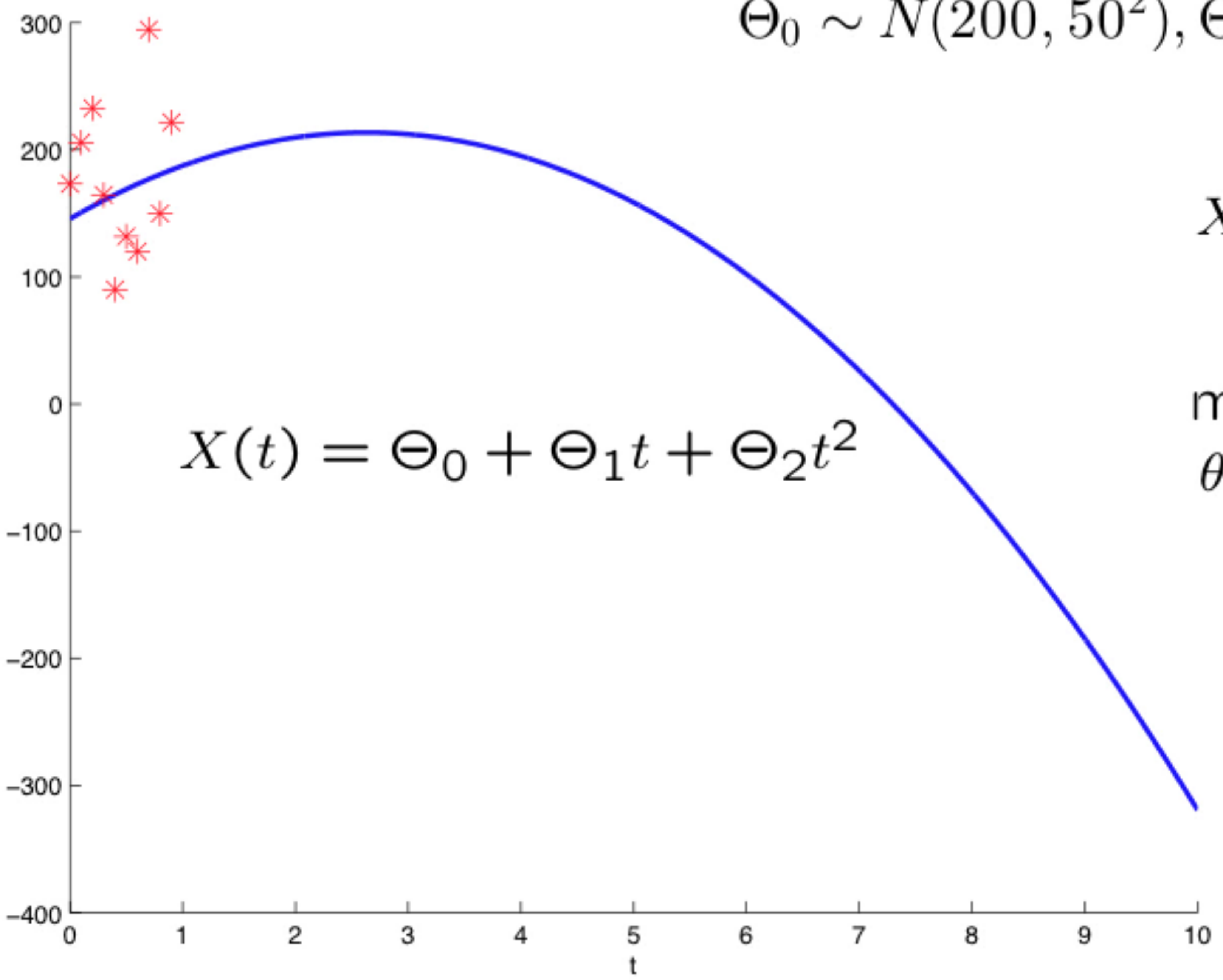
  $\widehat{\Theta}_{\text{MAP},j}$: linear function of $X = (X_1, \ldots, X_n)$

- Facts:
  - $\widehat{\Theta}_{\text{MAP},j} = \mathbf{E}[\Theta_j \,|\, X]$
  - 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
  - $\mathbf{E}\Big[(\widehat{\Theta}_{i,\text{MAP}} - \Theta_i)^2 \,|\, X = x\Big]$: same for all x

An illustration

Estimating the trajectory of a free-falling object

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



$$X(t) = \Theta_0 + \Theta_1 t + \Theta_2 t^2$$

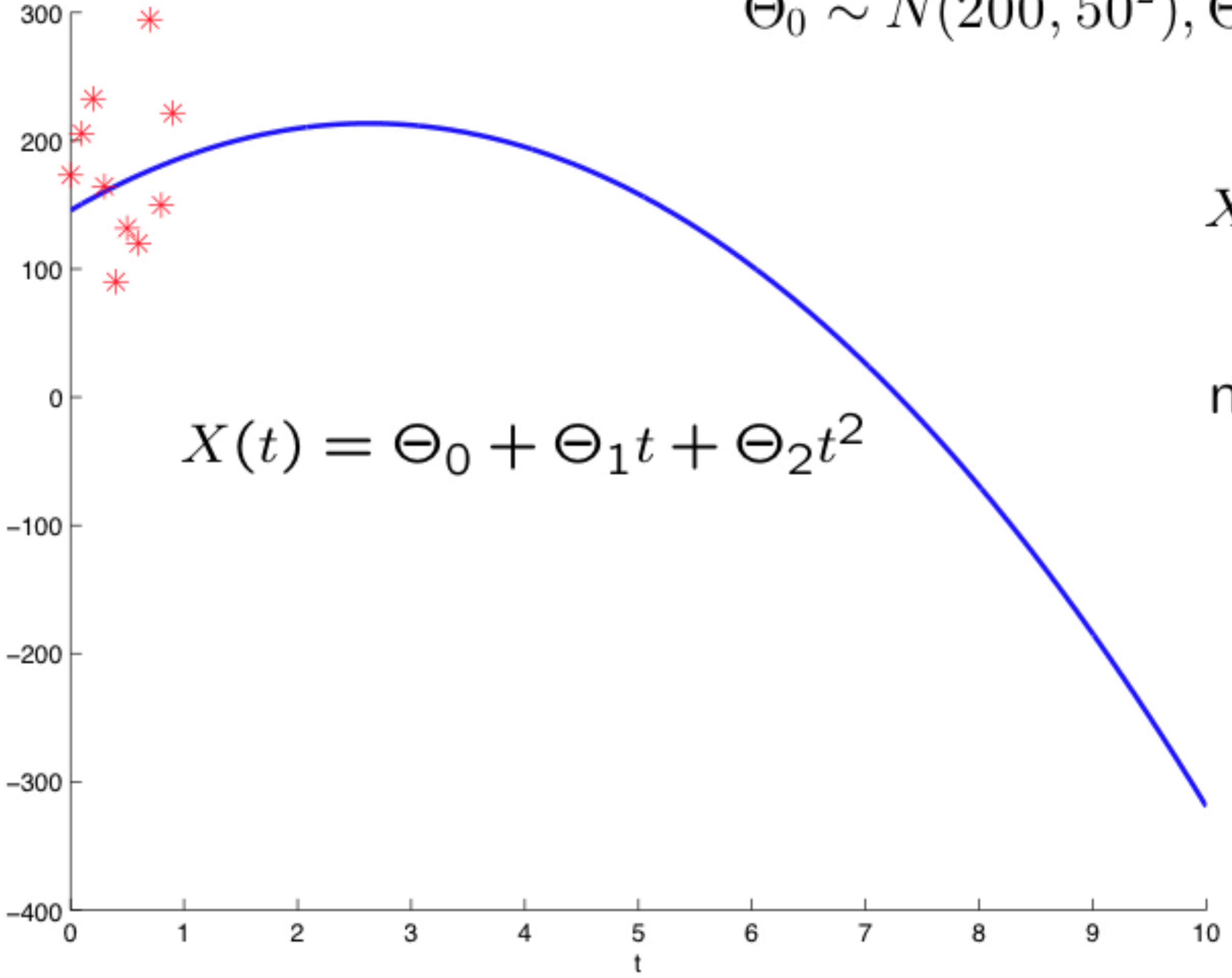
$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2 + W_i$$

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

An illustration

Estimating the trajectory of a free-falling object

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



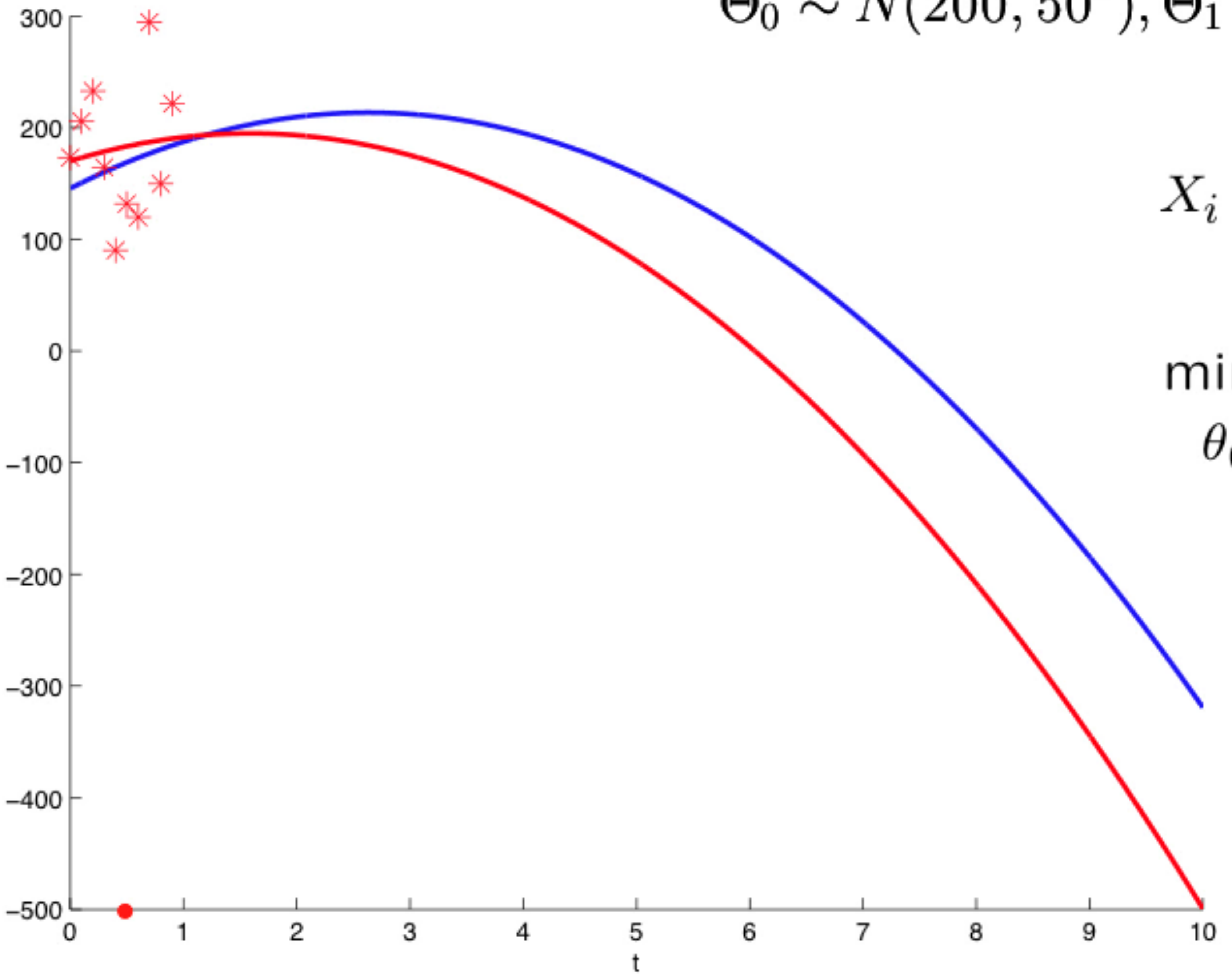
$$X(t) = \Theta_0 + \Theta_1 t + \Theta_2 t^2$$

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

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

An illustration

Estimating the trajectory of a free-falling object



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

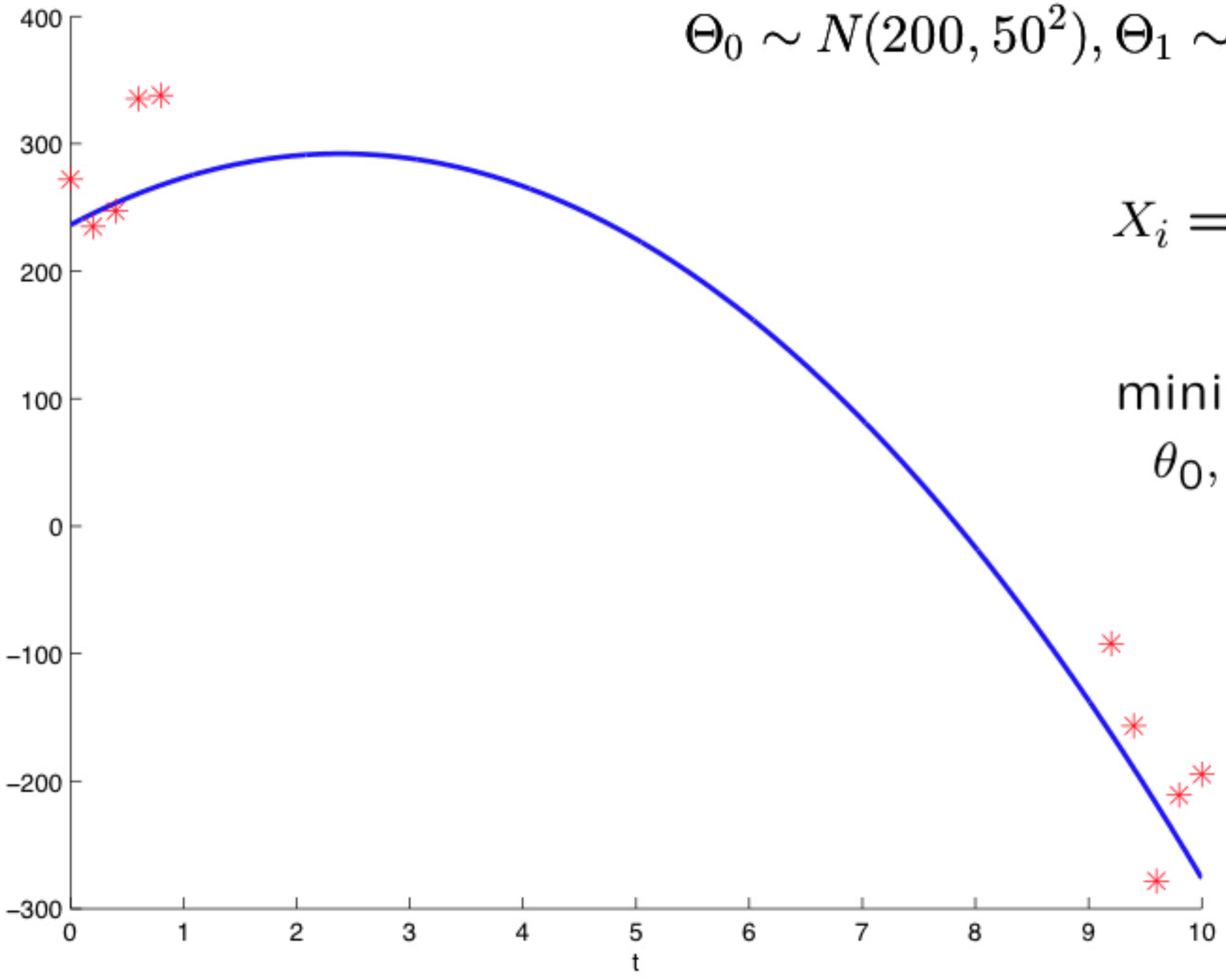
$$X_i = \underline{\Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2} + W_i$$

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

蓝线是要预测的函数
你还并不知道
红线是把观察到的data point带进去得到的。

An illustration

Estimating the trajectory of a free-falling object



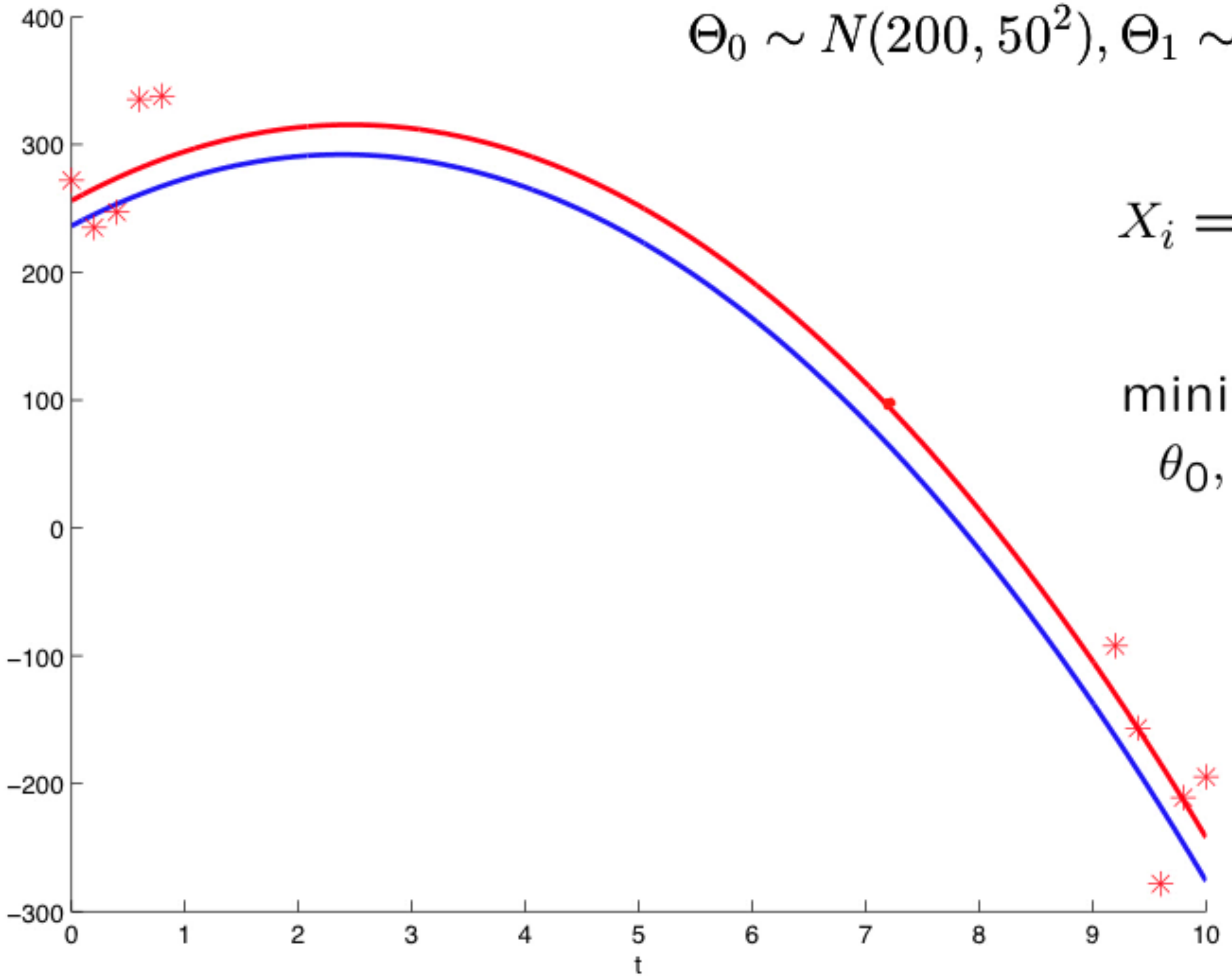
$$\Theta_0 \sim N(200, 50^2), \Theta_1 \sim N(50, 50^2), \Theta_2 = -9.81, W_i \sim N(0, 50^2)$$

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

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

An illustration

Estimating the trajectory of a free-falling object



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

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

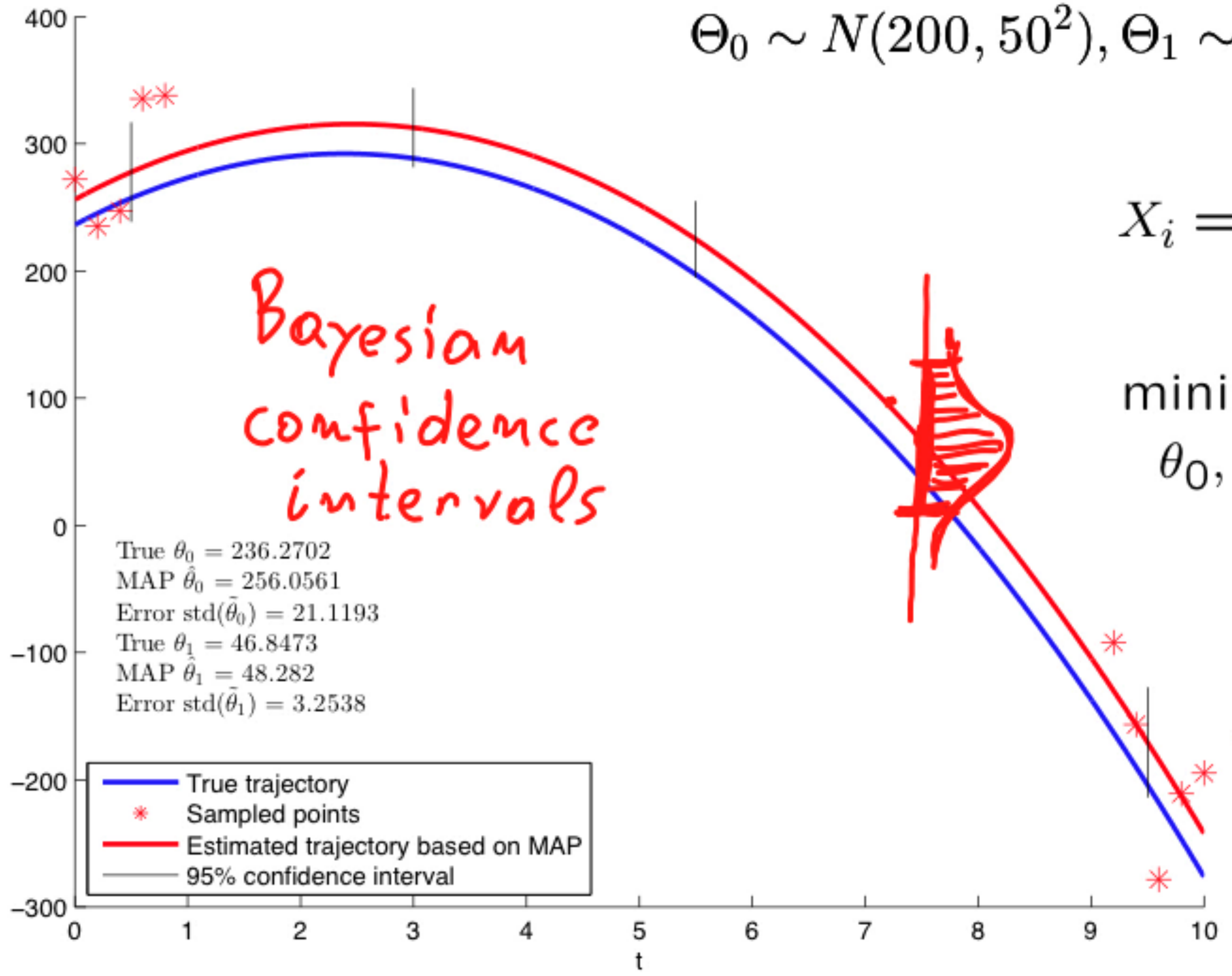
$$\underset{\theta_0,\, \theta_1}{\text{minimize}} \quad (\theta_0 - 200)^2 + (\theta_1 - 50)^2 + \sum_{i=1}^{n} (x_i - \theta_0 - \theta_1 t_i + 9.81 t_i^2)^2$$

你并不知道你的estimate（红线）
有没有接近真实的值

An illustration

Estimating the trajectory of a free-falling object

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



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

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

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