Chapter 6

Elements of parameter estimation

6.1 Bayesian parameter estimation

Throughout this chapter we assume that the data are samples of a parametrically known process $\{\mathcal{P}_\theta; \theta \in \boldsymbol{\tau}\}$, where $\mathcal{P}_\theta$ denotes a distribution on the observation space $(\Gamma, \mathcal{G})$:

$$\boldsymbol{X} \sim \mathcal{P}_{\boldsymbol{\theta}}(\boldsymbol{x}) \tag{6.1}$$

The goal of the parameter estimation problem is to find a function $\widehat{\boldsymbol{\theta}}(\boldsymbol{x}) : \Gamma \mapsto \boldsymbol{\tau}$ such that $\widehat{\boldsymbol{\theta}}(\boldsymbol{x})$ is the best guess of the true value of $\boldsymbol{\theta}$. Of course, the solution depends on a goodness criterion. As in the hypothesis testing problems, we have to define a cost function $c[\widehat{\boldsymbol{\theta}}(\boldsymbol{X}), \boldsymbol{\theta}] : \boldsymbol{\tau} \times \boldsymbol{\tau} \mapsto \mathbb{R}^+$ such that $c[\boldsymbol{a}, \boldsymbol{\theta}]$ is the cost of estimating the true value of $\boldsymbol{\theta}$ by $\boldsymbol{a}$.

Then, as in the hypothesis testing problems, we can define the conditional risk function

$$\begin{aligned} R_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) &= \mathrm{E}_{\boldsymbol{\theta}}\left\{c[\widehat{\boldsymbol{\theta}}(\boldsymbol{X}), \boldsymbol{\theta}]\right\} = \mathrm{E}\left[c[\widehat{\boldsymbol{\theta}}(\boldsymbol{X}), \boldsymbol{\Theta}] \,|\, \boldsymbol{\Theta} = \boldsymbol{\theta}\right] \\ &= \int_{\boldsymbol{\Gamma}} c[\widehat{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{\theta}] f_{\boldsymbol{\theta}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \end{aligned} \tag{6.2}$$

and the Bayes risk

$$\begin{aligned} r(\widehat{\boldsymbol{\theta}}) &= \mathrm{E}\left[R_{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\theta}}(\boldsymbol{X}))\right] \\ &= \int_{\boldsymbol{\tau}} \int_{\boldsymbol{\Gamma}} c[\widehat{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{\theta}] f_{\boldsymbol{\theta}}(\boldsymbol{x}) \pi(\boldsymbol{\theta}) \,\mathrm{d}\boldsymbol{x} \,\mathrm{d}\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\Gamma}} \int_{\boldsymbol{\tau}} c[\widehat{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{\theta}] \pi(\boldsymbol{\theta}|\boldsymbol{x}) \,\mathrm{d}\boldsymbol{\theta} \,\mathrm{d}\boldsymbol{x} \\ &= \mathrm{E}\left[\mathrm{E}\left[c[\widehat{\boldsymbol{\theta}}(\boldsymbol{X}), \boldsymbol{\Theta}] \,|\, \boldsymbol{X} = \boldsymbol{x}\right]\right] \end{aligned} \tag{6.3}$$

From this relation, and the fact that in general the cost function is positive, we see that $r(\widehat{\boldsymbol{\theta}})$ is minimized over $\widehat{\boldsymbol{\theta}}$, when for any $\boldsymbol{x} \in \boldsymbol{\Gamma}$, the mean posterior cost

$$\bar{c}[\boldsymbol{x}] = \mathrm{E}\left[c[\widehat{\boldsymbol{\theta}}(\boldsymbol{X}), \boldsymbol{\Theta}] \,|\, \boldsymbol{X} = \boldsymbol{x}\right] = \int_{\boldsymbol{\tau}} c[\widehat{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{\theta}] \pi(\boldsymbol{\theta}|\boldsymbol{x}) \,\mathrm{d}\boldsymbol{\theta} \tag{6.4}$$

is minimized.

It is clear that the resulting estimate depends on the choice of the cost function. In the following section we first consider the case of a scalar parameter and then extend it to the vector parameter case.

6.1.1 Minimum-Mean-Squared-Error

In case where $\boldsymbol{\tau} = \mathbb{R}$, a commonly used cost function is

$$c[a,\theta] = c(a-\theta) = (a-\theta)^2, \quad (a,\theta) \in \mathbb{R}^2 \tag{6.5}$$

The corresponding Bayes risk is $\mathrm{E}\left[(\widehat{\theta}(\boldsymbol{X}) - \Theta)^2\right]$, a quantity which is known as the *Mean-Squared-Error* (MSE). The corresponding Bayes estimate is called the *Minimum-Mean-Squared-Error* (MMSE) *estimator*.

The posterior cost is given by

$$\begin{aligned}
\mathrm{E}\left[(\widehat{\theta}(\boldsymbol{X}) - \Theta)^2 \,|\, \boldsymbol{X} = \boldsymbol{x}\right] &= \mathrm{E}\left[\widehat{\theta}^2(\boldsymbol{X}) \,|\, \boldsymbol{X} = \boldsymbol{x}\right] - 2\,\mathrm{E}\left[\widehat{\theta}(\boldsymbol{X})\Theta \,|\, \boldsymbol{X} = \boldsymbol{x}\right] + \mathrm{E}\left[\Theta^2 \,|\, \boldsymbol{X} = \boldsymbol{x}\right] \\
&= [\widehat{\theta}(\boldsymbol{X})]^2 - 2\,\widehat{\theta}(\boldsymbol{X})\,\mathrm{E}\left[\Theta \,|\, \boldsymbol{X} = \boldsymbol{x}\right] + \mathrm{E}\left[\Theta^2 \,|\, \boldsymbol{X} = \boldsymbol{x}\right] \qquad (6.6)
\end{aligned}$$

This expression is a quadratic function of $\widehat{\theta}(\boldsymbol{X})$ and its minimum is obtained for

$$\widehat{\theta}_{MMSE}(\boldsymbol{X}) = \mathrm{E}\left[\Theta \,|\, \boldsymbol{X} = \boldsymbol{x}\right] \tag{6.7}$$

Thus the MMSE estimate is the mean of the posterior probability density function. This estimate is also called *posterior mean* (PM) estimate.

6.1.2 Minimum-Mean-Absolute-Error

In case where $\boldsymbol{\tau} = \mathbb{R}$, another commonly used cost function is

$$c[a,\theta] = c(a-\theta) = |a-\theta|, \quad (a,\theta) \in \mathbb{R}^2 \tag{6.8}$$

The corresponding Bayes risk is $\mathrm{E}\left[|\widehat{\theta}(\boldsymbol{X}) - \Theta|\right]$, a quantity which is known as the *Mean-Absolute-Error* (MAE). The corresponding Bayes estimate is called the *Minimum-Mean-Absolute-Error* (MMAE) *estimator*.

The posterior cost is given by

$$\begin{aligned}
\mathrm{E}\left[|\widehat{\theta}(\boldsymbol{X}) - \Theta| \,|\, \boldsymbol{X} = \boldsymbol{x}\right] &= \int_0^\infty \mathrm{Pr}\left\{|\widehat{\theta}(\boldsymbol{x}) - \Theta| > z \,|\, \boldsymbol{X} = \boldsymbol{x}\right\} \mathrm{d}z \\
&= \int_0^\infty \mathrm{Pr}\left\{\Theta > z + \widehat{\theta}(\boldsymbol{x}) \,|\, \boldsymbol{X} = \boldsymbol{x}\right\} \mathrm{d}z \\
&\quad + \int_0^\infty \mathrm{Pr}\left\{\Theta < -z + \widehat{\theta}(\boldsymbol{x}) \,|\, \boldsymbol{X} = \boldsymbol{x}\right\} \mathrm{d}z
\end{aligned} \tag{6.9}$$

Doing the variable change $t = z + \widehat{\theta}(\boldsymbol{x})$ in the first integral and $t = -z + \widehat{\theta}(\boldsymbol{x})$ in the second one we obtain

$$\begin{aligned}
\mathrm{E}\left[|\widehat{\theta}(\boldsymbol{x}) - \Theta| \,|\, \boldsymbol{X} = \boldsymbol{x}\right] &= \int_{\widehat{\theta}(\boldsymbol{x})}^\infty \mathrm{Pr}\left\{\Theta > t \,|\, \boldsymbol{X} = \boldsymbol{x}\right\} \mathrm{d}t \\
&\quad + \int_{-\infty}^{\widehat{\theta}(\boldsymbol{x})} \mathrm{Pr}\left\{\Theta < t \,|\, \boldsymbol{X} = \boldsymbol{x}\right\} \mathrm{d}t \qquad (6.10)
\end{aligned}$$

This expression is differentiable with respect to $\widehat{\theta}(\boldsymbol{x})$ and

$$\begin{aligned}\frac{\partial \mathrm{E}\left[|\widehat{\theta}(\boldsymbol{X})-\Theta|\,|\,\boldsymbol{X}=\boldsymbol{x}\right]}{\partial\widehat{\theta}(\boldsymbol{x})} &= \Pr\left\{\Theta<\widehat{\theta}(\boldsymbol{x})\,|\,\boldsymbol{X}=\boldsymbol{x}\right\}\\ &\quad -\Pr\left\{\Theta>\widehat{\theta}(\boldsymbol{x})\,|\,\boldsymbol{X}=\boldsymbol{x}\right\} \end{aligned} \tag{6.11}$$

This derivative is a nondecreasing function of $\widehat{\theta}(\boldsymbol{x})$ which approaches -1 as $\widehat{\theta}(\boldsymbol{x}) \longrightarrow -\infty$ and $+1$ as $\widehat{\theta}(\boldsymbol{x}) \longrightarrow +\infty$. The minimum of (6.11) is achieved at the point $\widehat{\theta}(\boldsymbol{x})$ where the derivative vanishes. Consequently, the Bayes estimate satisfies

$$\begin{aligned}\Pr\{\Theta<t\,|\,\boldsymbol{X}=\boldsymbol{x}\} &\leq \Pr\{\Theta>t\,|\,\boldsymbol{X}=\boldsymbol{x}\}, \quad t<\widehat{\theta}(\boldsymbol{x})\\ &\text{and}\\ \Pr\{\Theta<t\,|\,\boldsymbol{X}=\boldsymbol{x}\} &\geq \Pr\{\Theta>t\,|\,\boldsymbol{X}=\boldsymbol{x}\}, \quad t>\widehat{\theta}(\boldsymbol{x})\end{aligned}$$

or

$$\Pr\left\{\Theta<\widehat{\theta}(\boldsymbol{x})\,|\,\boldsymbol{X}=\boldsymbol{x}\right\}=\Pr\left\{\Theta>\widehat{\theta}(\boldsymbol{x})\,|\,\boldsymbol{X}=\boldsymbol{x}\right\}. \tag{6.12}$$

$\widehat{\theta}(\boldsymbol{x})$ is the *median* of the posterior distribution of Θ given $\boldsymbol{X}=\boldsymbol{x}$:

$$\widehat{\theta}_{MMAE}(\boldsymbol{X}) = \text{median of} \quad \pi(\theta\,|\,\boldsymbol{X}=\boldsymbol{x}) \tag{6.13}$$

6.1.3 Maximum *A Posteriori* (MAP) estimation

Another commonly used cost function in the cases where $\boldsymbol{\tau}=\mathbb{R}$ is

$$c[a,\theta]=c(a-\theta)=\begin{cases}0 & \text{if} \quad |a-\theta|\leq\Delta\\ 1 & \text{if} \quad |a-\theta|>\Delta\end{cases} \tag{6.14}$$

where Δ is a positive real number. The corresponding Bayes risk is given by

$$\begin{aligned}\mathrm{E}\left[c[\widehat{\theta}(\boldsymbol{X})-\Theta]\,|\,\boldsymbol{X}=\boldsymbol{x}\right] &= \Pr\left\{|\widehat{\theta}(\boldsymbol{x})-\Theta|>\Delta\,|\,\boldsymbol{X}=\boldsymbol{x}\right\}\\ &= 1-\Pr\left\{|\widehat{\theta}(\boldsymbol{x})-\Theta|\leq\Delta\,|\,\boldsymbol{X}=\boldsymbol{x}\right\}\end{aligned} \tag{6.15}$$

To minimize this expression we consider two cases:

- Θ is a discrete random variable taking its values in a finite set $\boldsymbol{\tau}=\{\theta_1,\ldots,\theta_M\}$ such that $|\theta_i-\theta_j|>\Delta$ for any $i\neq j$. Then we have

  $$\mathrm{E}\left[c[\widehat{\theta}(\boldsymbol{X}),\Theta]\,|\,\boldsymbol{X}=\boldsymbol{x}\right]=1-\Pr\left\{\Theta=\widehat{\theta}(\boldsymbol{x})\,|\,\boldsymbol{X}=\boldsymbol{x}\right\}=1-\pi(\widehat{\theta}(\boldsymbol{x})\,|\,\boldsymbol{x}) \tag{6.16}$$

  where $\pi(\theta\,|\,\boldsymbol{x})$ is the posterior distribution of Θ given $\boldsymbol{X}=\boldsymbol{x}$. The estimate is the value of Θ which has the maximum *a posteriori* probability:

  $$\widehat{\theta}_{MAP}=\arg\max_{\theta\in\mathcal{T}}\{\pi(\theta\,|\,\boldsymbol{x})\} \tag{6.17}$$

- Θ is a continuous random variable. In this case, we have

$$\mathrm{E}\left[c[\widehat{\theta}(\boldsymbol{x}), \Theta] \,|\, \boldsymbol{X} = \boldsymbol{x}\right] = 1 - \int_{\widehat{\theta}(\boldsymbol{x})-\Delta}^{\widehat{\theta}(\boldsymbol{x})+\Delta} \pi(\theta \,|\, \boldsymbol{X} = \boldsymbol{x}) \, \mathrm{d}\theta \tag{6.18}$$

If we assume that the posterior probability distribution $\pi(\theta \,|\, \boldsymbol{x})$ is a continuous and smooth function and Δ is sufficiently small, then we can write

$$\mathrm{E}\left[c[\widehat{\theta}(\boldsymbol{x}), \Theta] \,|\, \boldsymbol{X} = \boldsymbol{x}\right] = 1 - 2\Delta \; \pi(\widehat{\theta}(\boldsymbol{x}) \,|\, \boldsymbol{X} = \boldsymbol{x}) \tag{6.19}$$

and again we have

$$\widehat{\theta}_{MAP} = \arg\max_{\theta \in \mathcal{T}} \{\pi(\theta \,|\, \boldsymbol{x})\} \tag{6.20}$$

Example 1: Estimation of the parameter of an exponential distribution.
Suppose both distributions $f_\theta(x)$ and $\pi(\theta)$ are exponential:

$$f_\theta(x) = \begin{cases} \theta \exp\left[-\theta x\right] & \text{if } \; x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.21}$$

and

$$\pi(\theta) = \begin{cases} \alpha \exp\left[-\alpha\theta\right] & \text{if } \; \theta > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.22}$$

Note that

$$\begin{cases} \mathrm{E}\left[X\right] = \theta \\ \mathrm{Var}\left\{X\right\} = \mathrm{E}\left[(X-\theta)^2\right] = \theta^2 \end{cases} \tag{6.23}$$

and

$$\begin{cases} \mathrm{E}\left[\Theta\right] = \alpha \\ \mathrm{Var}\left\{\Theta\right\} = \mathrm{E}\left[(\Theta-\alpha)^2\right] = \alpha^2 \end{cases} \tag{6.24}$$

Then, we can calculate the joint distribution $\phi(x, \theta)$

$$\phi(x,\theta) = \begin{cases} \alpha\,\theta \, \exp\left[-\theta x - \alpha\theta\right] & \text{if } \; \theta > 0, x > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{6.25}$$

The marginal distribution $m(x)$ is given by

$$m(x) = \begin{cases} \int_0^\infty \alpha\,\theta \, \exp\left[-(\alpha+x)\theta\right] \, \mathrm{d}\theta = \frac{\alpha}{(\alpha+x)^2} & \text{if } \; x > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{6.26}$$

and the posterior distribution $\pi(\theta|x)$ is given by

$$\pi(\theta|x) = \begin{cases} \frac{\alpha\theta\exp[-(\alpha+x)\theta]}{m(x)} = (\alpha+x)^2\theta\exp\left[-(\alpha+x)\theta\right] & \text{if } \; \theta > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{6.27}$$

Note that we have

$$\begin{cases} \mathrm{E}\left[\Theta \,|\, X = x\right] = \frac{2}{\alpha+x} \\ \mathrm{Var}\left\{\Theta \,|\, X = x\right\} = \frac{2}{(\alpha+x)^2} \end{cases} \tag{6.28}$$

The MMSE estimator is given by

$$\begin{aligned}\widehat{\theta}_{MMSE}(x) &= \mathrm{E}\left[\Theta \,|\, X = x\right] \\ &= \int_0^\infty \theta \pi(\theta|x)\,\mathrm{d}\theta \\ &= \int_0^\infty (\alpha+x)^2\theta^2 \exp\left[-(\alpha+x)\theta\right]\,\mathrm{d}\theta \\ &= \frac{2}{\alpha+x} \end{aligned} \tag{6.29}$$

and the corresponding MMSE is:

$$\begin{aligned}MMSE &= r(\widehat{\theta}_{MMSE}) = \mathrm{E}\left[\mathrm{Var}\left\{\Theta \,|\, X\right\}\right] \\ &= \int_0^\infty \frac{2}{(\alpha+x)^2}\, m(x)\,\mathrm{d}x \\ &= \int_0^\infty \frac{2\alpha}{(\alpha+x)^4}\,\mathrm{d}x \\ &= \frac{2}{3\alpha^2} \end{aligned} \tag{6.30}$$

The MMAE estimate $\widehat{\theta}_{ABS}(x)$ is such that

$$\int_{\widehat{\theta}_{ABS}}^\infty \pi(\theta|x)\,\mathrm{d}\theta = \left[1+(\alpha+x)\widehat{\theta}_{ABS}\right]\exp\left[-(\alpha+x)\widehat{\theta}_{ABS}\right] = \frac{1}{2} \tag{6.31}$$

It is easily shown that

$$\widehat{\theta}_{ABS}(x) = \frac{T_0}{\alpha+x} \tag{6.32}$$

where T_0 is the solution of

$$(1+T_0)\exp\left[-T_0\right] = \frac{1}{2} \longrightarrow T_0 \simeq 1.68$$

To calculate $\widehat{\theta}_{MAP}(x)$, we can remark that

$$\left\{\begin{aligned} \frac{\partial \log \pi(\theta\,|\,x)}{\partial\theta} &= \frac{1}{\theta} - (\alpha+x) \\ \frac{\partial^2 \log \pi(\theta\,|\,x)}{\partial\theta^2} &= -\frac{1}{\theta^2} < 0 \end{aligned}\right. \tag{6.33}$$

So, we have

$$\widehat{\theta}_{MAP}(x) = \frac{1}{\alpha+x} \tag{6.34}$$

The following table sumarizes theses estimates:

$$\left\{\begin{aligned} \widehat{\theta}_{MMSE}(x) &= \frac{2}{\alpha+x} \\ \widehat{\theta}_{ABS}(x) &= \frac{T_0}{\alpha+x} \\ \widehat{\theta}_{MAP}(x) &= \frac{1}{\alpha+x} \end{aligned}\right. \tag{6.35}$$

Example 2: Estimation of the location parameter of a Gaussian distribution.

Assume $X|\Theta$ and Θ have both Gaussian distributions:

$$\begin{aligned} X|\Theta &\sim f_\theta(x) &= \mathcal{N}\left(\theta, \sigma^2\right) \\ \Theta &\sim \pi(\theta) &= \mathcal{N}\left(\theta_0, \sigma_\theta^2\right) \end{aligned}$$

Then, the joint distribution $\phi(x,\theta)$, the marginal distribution $m(x)$ and the posterior distribution $\pi(\theta\,|\,x)$ are

$$\begin{aligned} X,\Theta &\sim \phi(x,\theta) &= \mathcal{N}\left((\theta,\theta_0), \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_\theta^2 \end{pmatrix}\right) \\ X &\sim m(x) &= \mathcal{N}\left(\theta, \sigma_x^2 = \sigma^2 + \sigma_\theta^2\right) \\ \Theta\,|\,X &\sim \pi(\theta\,|\,x) &= \mathcal{N}\left(\widehat{\theta}, \widehat{\sigma^2}\right) \end{aligned}$$

with

$$\begin{cases} \widehat{\theta} &= \frac{\sigma_\theta^2}{\sigma_x^2}\,\theta_0 + \frac{\sigma^2}{\sigma_x^2}\,x \\ \widehat{\sigma^2} &= \frac{\sigma\,\sigma_\theta}{\sigma_x^2} \end{cases}$$

In this case all the estimators are equal and we have

$$\begin{aligned} \widehat{\theta}_{MMSE}(x) = \widehat{\theta}_{ABS}(x) = \widehat{\theta}_{MAP}(x) &= \widehat{\theta} = \frac{\sigma_\theta^2}{\sigma_x^2}\,\theta_0 + \frac{\sigma^2}{\sigma_x^2}\,x && (6.36) \\ &= \frac{\sigma_\theta^2}{\sigma^2+\sigma_\theta^2}\,\theta_0 + \frac{\sigma^2}{\sigma^2+\sigma_\theta^2}\,x && (6.37) \\ & && (6.38) \end{aligned}$$

They also have the same posterior variance

$$\widehat{\sigma^2} = \frac{\sigma\,\sigma_\theta}{\sigma_x^2} = \frac{\sigma\,\sigma_\theta}{\sigma^2+\sigma_\theta^2} \tag{6.39}$$

Note also the following limit cases:

$$\begin{aligned} &\text{When } \sigma^2 \longrightarrow 0 \text{ then } \widehat{\theta} \longrightarrow \theta_0 \\ &\text{When } \sigma_\theta^2 \longrightarrow 0 \text{ then } \widehat{\theta} \longrightarrow x \end{aligned}$$

Example 3: Estimation of the parameter of a Binomial distribution.

Assume that

$$\begin{aligned} X\,|\,\theta &\sim f(x\,|\,\theta) &&= \mathbf{Bin}(x|\theta,n) = C_n^x\,\theta^x\,(1-\theta)^{n-x} \\ \Theta &\sim \pi(\theta) &&= \mathbf{Beta}(\theta|\alpha,\beta) = \tfrac{1}{B(\alpha,\beta)}\,\theta^{\alpha-1}\,(1-\theta)^{\beta-1} \end{aligned}$$

Then, the joint distribution $\phi(x,\theta)$, the marginal distribution $m(x)$ and the posterior distribution $\pi(\theta\,|\,x)$ are

$$\begin{aligned} (X,\Theta) &\sim \phi(x,\theta) &&= \tfrac{C_n^x}{\mathcal{B}(\alpha,\beta)}\,\theta^{\alpha+x-1}\,(1-\theta)^{\beta+n-x-1} \\ X &\sim m(x) &&= \tfrac{C_n^x}{\mathcal{B}(\alpha,\beta)}\,\mathcal{B}(\alpha+x,n+\beta-x) \\ & && = C_n^x\,\tfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\tfrac{\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(\alpha+\beta+n)} \\ \Theta\,|\,X &\sim \pi(\theta|x) &&= \tfrac{1}{B(\alpha+x,\beta+n-x)}\,\theta^{\alpha+x-1}\,(1-\theta)^{\beta+n-x-1} \\ & && = \mathbf{Beta}(\alpha+x,n+\beta-x) \end{aligned}$$

6.2 Other cost functions and related estimators

Here are some other cost functions and corresponding Bayesian estimator expressions.

Name	$C[a,\theta]$	$\widehat{\theta}$
Quadratic	$q(a-\theta)^2$	$\mathrm{E}\left[\Theta\|\boldsymbol{x}\right] = \int \theta \pi(\theta\|\boldsymbol{x})\,\mathrm{d}\theta$
Weighted Quadratic	$\omega(\theta)(a-\theta)^2$	$\frac{\mathrm{E}[\omega(\Theta)\,\Theta\|\boldsymbol{x}]}{\mathrm{E}[\omega(\Theta)\|\boldsymbol{x}]} = \dfrac{\int \omega(\theta)\;\theta\;\pi(\theta\|\boldsymbol{x})\,\mathrm{d}\theta}{\int \omega(\theta)\;\pi(\theta\|\boldsymbol{x})\,\mathrm{d}\theta}.$
Absolute	$\|a-\theta\|$	$\int_{-\infty}^{\widehat{\theta}} \pi(\theta\|\boldsymbol{x})\,\mathrm{d}\theta = \int_{\widehat{\theta}}^{+\infty} \pi(\theta\|\boldsymbol{x})\,\mathrm{d}\theta = \frac{1}{2}$
Nonsymetric Absolute	$\begin{cases} k_2(\theta-a) & \text{si } \theta \leq a, \\ k_1(a-\theta) & \text{si } \theta \geq a. \end{cases}$	$k_1 \int_{-\infty}^{\widehat{\theta}} \pi(\theta\|\boldsymbol{x})\,\mathrm{d}\theta = k_2 \int_{\widehat{\theta}}^{+\infty} \pi(\theta\|\boldsymbol{x})\,\mathrm{d}\theta$
Linex	$\beta\left[\exp\left[-\alpha(a-\theta)\right] - \alpha(\alpha-\theta) - 1\right]$, $\alpha \neq 0, \beta > 0$	$-\frac{1}{\alpha}\log\left(\mathrm{E}\left[\exp\left[-\alpha\Theta\right] \mid \boldsymbol{x}\right]\right)$ $= -\frac{1}{\alpha}\log\left[\int \exp\left[-\alpha\theta\right] p(\theta\|\boldsymbol{x})\,\mathrm{d}\theta\right]$
	$\frac{(a-\theta)^2}{\theta}$	$\frac{1}{\mathrm{E}[1/\Theta\|\boldsymbol{x}]} = \frac{1}{\int \frac{1}{\theta}\pi(\theta\|\boldsymbol{x})\,\mathrm{d}\theta}$
	$\frac{(a-\theta)^2}{a}$	$\sqrt{\mathrm{E}\left[\Theta^2\|\boldsymbol{x}\right]} = \sqrt{\int \theta^2\,\pi(\theta\|\boldsymbol{x})\,\mathrm{d}\theta}$
	$-\ln[a(\theta)]$	$\pi(\theta\|\boldsymbol{x})$

Table 6.1: Relations between the data, *a priori*, marginal and *a posteriori* distributions.

6.3 Examples of posterior calculation

In previous sections, we saw that the computation of Bayesian estimates requires the posterior probability distribution. The following table gives a summary of the expressions of the posterior probability distributions in some classical cases.

Observation law $f(x\|\theta)$	Prior law $\pi(\theta)$	Marginal law $m(x) = \int f(x\|\theta)\,\pi(\theta)\,\mathrm{d}\theta$	Posterior law $\pi(\theta\|x) = \dfrac{f(x\|\theta)\,\pi(\theta)}{m(x)}$
		Discrete variables	
Binomial $\mathbf{Bin}(x\|n,\theta)$	Beta $\mathbf{Bet}(\theta\|\alpha,\beta)$	Binomial-Beta $\mathbf{BinBet}(x\|\alpha,\beta,n)$	Beta $\mathbf{Bet}(\theta\|\alpha+x,\beta+n-x)$
Negative Binomial $\mathbf{NegBin}(x\|n,\theta)$	Beta $\mathbf{Bet}(\theta\|\alpha,\beta)$	Negative Binomial-Beta $\mathbf{NegBinBet}(x\|\alpha,\beta,\theta)$	Beta $\mathbf{Bet}(\theta\|\alpha+n,\beta+x)$
Poisson $\mathbf{Pn}(x\|\theta)$	Gamma $\mathbf{Gam}(\theta\|\alpha,\beta)$	Poisson-Gamma $\mathbf{PnGam}(x\|\alpha,\beta,1)$	Gamma $\mathbf{Gam}(\theta\|\alpha+x,\beta+1)$
		Continuous variables	
Gamma $\mathbf{Gam}(x\|\nu,\theta)$	Gamma $\mathbf{Gam}(\theta\|\alpha,\beta)$	Gamma-Gamma $\mathbf{GamGam}(x\|\alpha,\beta,\nu)$	Gamma $\mathbf{Gam}(\theta\|\alpha+\nu,\beta+x)$
Exponential $\mathbf{Ex}(x\|\theta)$	Gamma $\mathbf{Gam}(\theta\|\alpha,\beta)$	Pareto $\mathbf{Par}(x\|\alpha,\beta)$	Gamma $\mathbf{Gam}(\theta\|\alpha+1,\beta+x)$
Normal $\mathbf{N}(x\|\theta,\sigma^2)$	Normal $\mathbf{N}(\theta\|\mu,\tau^2)$	Normal $\mathbf{N}(x\|\mu+\theta,\tau^2)$?	Normal $\mathbf{N}\left(\mu\|\frac{\mu\sigma^2+\tau^2 x}{\sigma^2+\tau^2},\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)$
Normal $\mathbf{N}(x\|\mu,\lambda\theta)$	Gamma $\mathbf{Gam}(\theta\|\frac{\alpha}{2},\frac{\alpha}{2})$	Student (t) $\mathbf{St}(x\|\mu,\lambda,\alpha)$	Gamma $\mathbf{Gam}\left(\theta\|\frac{\alpha+1}{2},\frac{\alpha}{2}+\frac{1}{2}(\mu-x)^2\right)$

Table 6.2: Relation between the data, *a priori*, marginal and *a posteriori* distributions.

6.4 Estimation of vector parameters

In the case where we have a vector parameter $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_m]^t$ we have to define a cost function $c[\boldsymbol{a}, \boldsymbol{\theta}] : \mathbb{R}^m \times \mathbb{R}^m \mapsto \mathbb{R}^+$. Then it is again possible to define the Bayes risk. In many cases the cost function is of the form

$$c[\boldsymbol{a}, \boldsymbol{\theta}] = \sum_{i=1}^{m} c_i[a_i, \theta_i] \tag{6.40}$$

We then have

$$\mathrm{E}\left[c[\widehat{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{\theta}] \,|\, \boldsymbol{X} = \boldsymbol{x}\right] = \sum_{i=1}^{m} \mathrm{E}\left[c_i[\widehat{\theta}_i(\boldsymbol{x}), \theta_i] \,|\, \boldsymbol{X} = \boldsymbol{x}\right] \tag{6.41}$$

Here after, we consider some common cost functions:

6.4.1 Minimum-Mean-Squared-Error

In case where $\boldsymbol{\tau} = \mathbb{R}^m$, a commonly used cost function is

$$c[\boldsymbol{a}, \boldsymbol{\theta}] = \|\boldsymbol{a} - \boldsymbol{\theta}\|^2 = \sum_{i=1}^{m} (a_i - \theta_i)^2 \tag{6.42}$$

The corresponding Bayes risk is $\mathrm{E}\left[\|\widehat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\Theta}\|^2\right]$ and the corresponding Bayes estimate is the *Minimum-Mean-Squared-Error* (MMSE) *estimator* or the Bayes estimate:

$$\widehat{\boldsymbol{\theta}}_{MMSE}(\boldsymbol{X}) = \mathrm{E}\left[\boldsymbol{\Theta} \,|\, \boldsymbol{X} = \boldsymbol{x}\right] \tag{6.43}$$

Thus the MMSE estimate is the mean of the posterior probability density function. It is also called *posterior mean* (PM) estimate.

Note that, as in the scalar case, the following weighted quadratic cost function

$$c[\boldsymbol{a}, \boldsymbol{\theta}] = \|\boldsymbol{a} - \boldsymbol{\theta}\|^2_{\boldsymbol{Q}} = [\boldsymbol{a} - \boldsymbol{\theta}]^t \boldsymbol{Q} [\boldsymbol{a} - \boldsymbol{\theta}] \tag{6.44}$$

gives the same estimate as in (6.43), *i.e.* the MMSE estimate does not depend on the weighting matrix $\boldsymbol{Q}$. However, the corresponding minimum Bayes risks are different and we have

$$\mathrm{E}\left[\|\widehat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\Theta}\|^2_{\boldsymbol{Q}}\right] = \mathrm{tr}\left\{\boldsymbol{Q}\mathrm{E}\left[\mathrm{Cov}\left\{\boldsymbol{\Theta} \,|\, \boldsymbol{X} = \boldsymbol{x}\right\}\right]\right\} \tag{6.45}$$

.

6.4.2 Minimum-Mean-Absolute-Error

In case where $\boldsymbol{\tau} = \mathbb{R}$, another commonly used cost function is

$$c[\boldsymbol{a}, \boldsymbol{\theta}] = \sum_{i=1}^{m} |a_i - \theta_i|, \tag{6.46}$$

The corresponding estimate is such that

$$\Pr\left\{\Theta_i < \widehat{\theta}_i(\boldsymbol{x}) \,|\, \boldsymbol{X} = \boldsymbol{x}\right\} = \Pr\left\{\Theta_i > \widehat{\theta}_i(\boldsymbol{x}) \,|\, \boldsymbol{X} = \boldsymbol{x}\right\}, \tag{6.47}$$

which means that $\widehat{\theta}_i(\boldsymbol{x})$ is the *median* of the marginal posterior distribution of Θ_i given $\boldsymbol{X} = \boldsymbol{x}$, *i.e.* $\pi(\theta_i \,|\, \boldsymbol{X} = \boldsymbol{x})$

6.4.3 Marginal Maximum *A Posteriori* (MAP) estimation

Another commonly used cost function in the cases where $\boldsymbol{\tau} = \mathbb{R}^m$ is

$$c[\boldsymbol{a}, \boldsymbol{\theta}] = \sum_{i=1}^{m} c[a_i - \theta_i] \quad \text{with} \quad c[a_i - \theta_i] = \begin{cases} 0 & \text{if} \quad |a_i - \theta_i| \leq \Delta \\ 1 & \text{if} \quad |a_i - \theta_i| > \Delta \end{cases} \tag{6.48}$$

where Δ is a positive real number. The corresponding estimate is given by :

$$\widehat{\theta}_i = \arg\max_{\theta_i \in \mathcal{T}} \{\pi(\theta_i \,|\, \boldsymbol{x})\} \tag{6.49}$$

if Δ is sufficiently small.

6.4.4 Maximum *A Posteriori* (MAP) estimation

Two other cost functions which give the same estimates are :

$$c[\boldsymbol{a}, \boldsymbol{\theta}] = \begin{cases} 0 & \text{if} \quad \max_i |a_i - \theta_i| \leq \Delta \\ 1 & \text{if} \quad \max_i |a_i - \theta_i| > \Delta \end{cases} \tag{6.50}$$

and

$$c[\boldsymbol{a}, \boldsymbol{\theta}] = \begin{cases} 0 & \text{if} \quad \|\boldsymbol{a} - \boldsymbol{\theta}\|^2 \leq \Delta \\ 1 & \text{if} \quad \|\boldsymbol{a} - \boldsymbol{\theta}\|^2 > \Delta \end{cases} \tag{6.51}$$

where Δ is a positive real number.

In both cases, if the posterior distribution $\pi(\boldsymbol{\theta} \,|\, \boldsymbol{x})$ is continuous and smooth enough, we obtain the MAP estimate.

The corresponding estimate is given by :

$$\widehat{\boldsymbol{\theta}}_{MAP} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\tau}} \{\pi(\boldsymbol{\theta} \,|\, \boldsymbol{x})\} \tag{6.52}$$

Note that the MMAP estimate in (6.49) and the estimate (6.52) may be very different.

6.4.5 Estimation of a Gaussian vector parameter from jointly Gaussian observation

The case of the estimation of a Gaussian vector parameter $\boldsymbol{\theta} \in \mathbb{R}^m$ from a jointly Gaussian observation $\boldsymbol{x} \in \mathbb{R}^n$ is a very useful example and is used in many applications.

Suppose $\boldsymbol{\Theta}$ and $\boldsymbol{X}$ have the following *a priori* distributions:

$$\begin{aligned} \boldsymbol{\Theta} &\sim \mathcal{N}(\boldsymbol{\theta}_0, \boldsymbol{R}_\Theta) \\ \boldsymbol{X} &\sim \mathcal{N}(\boldsymbol{x}_0, \boldsymbol{R}_X) \end{aligned}$$

and

$$\begin{pmatrix} \boldsymbol{\Theta} \\ \boldsymbol{X}_0 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\theta}_0 \\ \boldsymbol{x}_0 \end{pmatrix} \quad , \quad \begin{pmatrix} \boldsymbol{R}_\Theta & \boldsymbol{R}_{\Theta X} \\ \boldsymbol{R}_{X\Theta} & \boldsymbol{R}_X \end{pmatrix} \right)$$

with $\boldsymbol{R}_{\Theta X} = \boldsymbol{R}_{X\Theta}^t$.

It is easy to show that the posterio law is also Gaussian and is given by

$$\boldsymbol{\Theta}|\boldsymbol{X} \sim \mathcal{N}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{R}}) \tag{6.53}$$

with

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + \boldsymbol{R}_{\Theta X} \boldsymbol{R}_X^{-1} (\boldsymbol{x} - \boldsymbol{x}_0) \tag{6.54}$$

$$\widehat{\boldsymbol{R}} = \boldsymbol{R}_\Theta - \boldsymbol{R}_{\Theta X} \boldsymbol{R}_X^{-1} \boldsymbol{R}_{X\Theta} \tag{6.55}$$

We also have

$$\mathrm{E}\left[\boldsymbol{\Theta} \,|\, \boldsymbol{X} = \boldsymbol{x}\right] = \widehat{\boldsymbol{\theta}} \tag{6.56}$$

$$\mathrm{Cov}\left\{\boldsymbol{\Theta} \,|\, \boldsymbol{X} = \boldsymbol{x}\right\} = \widehat{\boldsymbol{R}} \tag{6.57}$$

The corresponding minimum Bayes risk is

$$r(\widehat{\boldsymbol{\theta}}) = \mathrm{tr}\left\{\boldsymbol{Q}\widehat{\boldsymbol{R}}\right\} = \mathrm{tr}\left\{\boldsymbol{Q}\boldsymbol{R}_\Theta\right\} - \mathrm{tr}\left\{\boldsymbol{Q}\boldsymbol{R}_{\Theta X}\boldsymbol{R}_X^{-1}\boldsymbol{R}_{X\Theta}\right\} \tag{6.58}$$

6.4.6 Case of linear models

When the observation vector is related to the vector parameter $\boldsymbol{\theta}$ by a linear model we have

$$X_i = \sum_{j=1}^{m} h_{i,j}\,\Theta_j + N_i, \quad i = 1,\ldots,n \tag{6.59}$$

or in a matrix form

$$\boldsymbol{X} = \boldsymbol{H}\boldsymbol{\Theta} + \boldsymbol{N} \tag{6.60}$$

with

$$\begin{aligned} \boldsymbol{\Theta} &\sim \mathcal{N}(\boldsymbol{\theta}_0, \boldsymbol{R}_\Theta) \\ \boldsymbol{N} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_N) \end{aligned}$$

Then we have

$$\boldsymbol{X}|\boldsymbol{\Theta}=\boldsymbol{\theta} \;\sim\; \mathcal{N}(\boldsymbol{H}\boldsymbol{\theta}, \boldsymbol{R}_N)$$

$$\begin{aligned} \boldsymbol{R}_X &= \boldsymbol{H}\boldsymbol{R}_\Theta\boldsymbol{H}^t + \boldsymbol{H}\boldsymbol{R}_{\Theta N} + \boldsymbol{R}_{N\Theta}\boldsymbol{H}^t + \boldsymbol{R}_N \\ \boldsymbol{R}_{X\Theta} &= \boldsymbol{H}\boldsymbol{R}_\Theta + \boldsymbol{R}_{\Theta N}, \quad \boldsymbol{R}_{\Theta X} = \boldsymbol{R}_{X\Theta}^t \end{aligned}$$

If we assume that the noise $\boldsymbol{N}$ and the vector parameter $\boldsymbol{\Theta}$ are independant, we have

$$\boldsymbol{R}_X = \boldsymbol{H}\boldsymbol{R}_\Theta\boldsymbol{H}^t + \boldsymbol{R}_N \tag{6.61}$$

$$\boldsymbol{R}_{X\Theta} = \boldsymbol{H}\boldsymbol{R}_\Theta, \quad \boldsymbol{R}_{\Theta X} = \boldsymbol{R}_\Theta^t\boldsymbol{H}^t \tag{6.62}$$

$$\boldsymbol{R}_X^{-1} = \boldsymbol{R}_N^{-1} - \boldsymbol{R}_N^{-1}\boldsymbol{H}\left(\boldsymbol{R}_\Theta^{-1} + \boldsymbol{H}^t\boldsymbol{R}_N^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^t\boldsymbol{R}_\Theta^{-1} \tag{6.63}$$

and

$$\boldsymbol{\Theta}|\boldsymbol{X} \sim \mathcal{N}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{R}}) \tag{6.64}$$

with

$$\left\{\begin{aligned} \widehat{\boldsymbol{\theta}} &= \mathrm{E}\left[\boldsymbol{\Theta}\,|\,\boldsymbol{x}\right] = \boldsymbol{\theta}_0 + \boldsymbol{R}_{\Theta X}\boldsymbol{R}_X^{-1}(\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}_0) \\ \widehat{\boldsymbol{R}} &= \mathrm{E}\left[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\,|\,\boldsymbol{x}\right] = \boldsymbol{R}_\Theta - \boldsymbol{R}_{\Theta X}\boldsymbol{R}_X^{-1}\boldsymbol{R}_{X\Theta} \end{aligned}\right.$$

$$\left\{\begin{aligned} \widehat{\boldsymbol{\theta}} &= \boldsymbol{\theta}_0 + \boldsymbol{R}_\Theta\boldsymbol{H}^t\left[\boldsymbol{H}\boldsymbol{R}_\Theta\boldsymbol{H}^t + \boldsymbol{R}_N\right]^{-1}(\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}_0) \\ &= \boldsymbol{\theta}_0 + \left[\boldsymbol{H}^t\boldsymbol{R}_N^{-1}\boldsymbol{H} + \boldsymbol{R}_\Theta^{-1}\right]^{-1}\boldsymbol{H}^t\boldsymbol{R}_N^{-1}(\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}_0) \\ \widehat{\boldsymbol{R}} &= \boldsymbol{R}_\Theta - \boldsymbol{R}_\Theta\boldsymbol{H}^t\left[\boldsymbol{H}\boldsymbol{R}_\Theta\boldsymbol{H}^t + \boldsymbol{R}_N\right]^{-1}\boldsymbol{H}\boldsymbol{R}_\Theta \\ &= \left[\boldsymbol{H}^t\boldsymbol{R}_N^{-1}\boldsymbol{H} + \boldsymbol{R}_\Theta^{-1}\right]^{-1} \end{aligned}\right. \tag{6.65}$$

Consider now the particular case of $\boldsymbol{R}_N = \sigma_b^2\boldsymbol{I}$, $\boldsymbol{R}_\Theta = \sigma_x^2(\boldsymbol{D}^t\boldsymbol{D})^{-1}$ and $\boldsymbol{\theta}_0 = \boldsymbol{0}$. We then have

$$\left\{\begin{aligned} \widehat{\boldsymbol{\theta}} &= \left(\boldsymbol{H}^t\boldsymbol{H} + \boldsymbol{\lambda}\boldsymbol{D}^t\boldsymbol{D}\right)^{-1}\boldsymbol{H}^t\boldsymbol{x}, \\ \widehat{\boldsymbol{R}} &= \sigma_b^2\left(\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right)^{-1}, \quad \text{with} \quad \lambda = \sigma_b^2/\sigma_b^2 \end{aligned}\right. \tag{6.66}$$

6.5 Examples

6.5.1 curve fitting

We consider here a classical problem of curve fitting that any engineer is almost anytime faced to. We analyse this problem as a parameter estimation : Given a set of data $\{(x_i, t_i), i = 1, \ldots, n\}$ estimate the parameters of an algebraic curve to fit the best these data. Among different curves, the polynomials are used very commonly.

A polynomial model of degree p relating $x_i = x(t_i)$ to t_i is

$$x_i = x(t_i) = \theta_0 + \theta_1 t_i + \theta_2 t_i^2 + \cdots + \theta_p t_i^p, \quad i = 1, \ldots, n \tag{6.67}$$

Noting that this relation is linear in θ_i, we can rewrite it in the following

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & \cdots & t_1^p \\ 1 & t_2 & t_2^2 & \cdots & \cdots & t_2^p \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & t_n & t_n^2 & \cdots & \cdots & t_n^p \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_p \end{pmatrix} \tag{6.68}$$

or

$$\boldsymbol{x} = \boldsymbol{H}\boldsymbol{\theta} \tag{6.69}$$

The matrix $\boldsymbol{H}$ is called the *Vandermond* matrix. It is entirely determined by the vector $\boldsymbol{t} = [t_1, t_2, \ldots, t_n]^t$.

In the case where $n = p + 1$, this matrix is invertible iff $t_i \neq t_j, \forall i \neq j$. In general, however we have more data than unknowns, *i.e.* $n > p + 1$.

Note that the matrix $\boldsymbol{H}^t\boldsymbol{H}$ is a *Hankel* matrix:

$$[\boldsymbol{H}^t\boldsymbol{H}]_{kl} = \sum_{i=1}^{n} t_i^{k-1}\, t_i^{l-1} = \sum_{i=1}^{n} t_i^{k+l-2}, \quad k, l = 1, \ldots, p+1 \tag{6.70}$$

and the vector $\boldsymbol{H}^t\boldsymbol{x}$ is such that

$$[\boldsymbol{H}^t\boldsymbol{x}]_k = \sum_{i=1}^{n} t_i^{k-1}\, x_i, \quad k = 1, \ldots, p+1 \tag{6.71}$$

Line fitting is the following particular case

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} \tag{6.72}$$

In this case we have

$$\boldsymbol{H}^t\boldsymbol{H} = \begin{pmatrix} n & \sum_{i=1}^{n} t_i \\ \sum_{i=1}^{n} t_i & \sum_{i=1}^{n} t_i^2 \end{pmatrix} \tag{6.73}$$

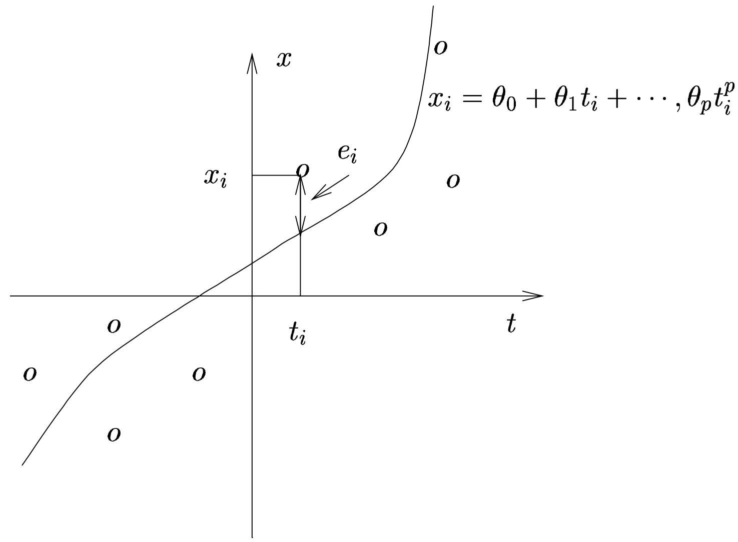


Figure 6.1: Curve fitting.

and

$$\boldsymbol{H}^t \boldsymbol{x} = \begin{pmatrix} \sum_{i=1}^n x_i \\ \sum_{i=1}^n t_i\, x_i \end{pmatrix} \tag{6.74}$$

In the following we consider the line fitting case and will see how different assumptions about the problem can give different solutions.

Model 1:
The easiest model is to assume that t_i are perfectly known, and we only have uncertainties on x_i, *i.e.*

$$x_i = x(t_i) = \theta_0 + \theta_1 t_i + e_i, \quad i = 1, \ldots, n \tag{6.75}$$

where e_i represents the error on x_i. In a geometric language, e_i is the signed distance between the point (t_i, x_i) and the point $(t_i, \theta_0 + \theta_1 t_i)$ (see figure 6.5.1).

Here, we have $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{e}$ with $\boldsymbol{\theta} = [\theta_0, \theta_1]^t$. The matrix $\boldsymbol{H}$ is perfectly known. Note that in this model, if we assume that e_i are zero mean, white and Gaussian

$$e_i = x_i - \theta_0 - \theta_1 t_i \sim \mathcal{N}\left(0, \sigma_e^2\right) \tag{6.76}$$

then the likelihood function becomes

$$f(\boldsymbol{x} \,|\, \boldsymbol{\theta}) = \mathcal{N}\left(\boldsymbol{0}, \sigma_e^2 \boldsymbol{I}\right) \propto \exp\left[-\frac{1}{2\sigma_e^2} \|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\|^2\right] \tag{6.77}$$

and the maximum likelihood estimate is

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{\|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\|^2\right\} \tag{6.78}$$

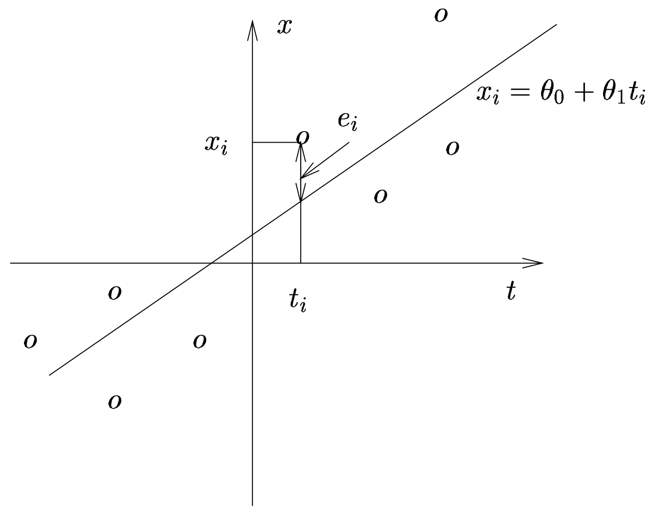


Figure 6.2: Line fitting: model 1: $e_i = x_i - (\theta_0 + \theta_1 t_i)$

If the $\boldsymbol{H}^t\boldsymbol{H}$ is invertible, $\widehat{\boldsymbol{\theta}}$ is given by

$$\widehat{\boldsymbol{\theta}} = [\boldsymbol{H}^t\boldsymbol{H}]^{-1}\boldsymbol{H}^t\boldsymbol{x} \tag{6.79}$$

To define any Bayesian estimate, we have to assign a prior probability law to $\boldsymbol{\theta}$. Let assume that θ_0 and θ_1 are independent and

$$\theta_0 \sim \mathcal{N}\left(0, \sigma_0^2\right), \qquad \theta_1 \sim \mathcal{N}\left(1, \sigma_1^2\right) \tag{6.80}$$

or

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}\right) = \mathcal{N}\left(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_\theta\right) \tag{6.81}$$

Exercise 1:

- Write the complete expressions of $f(x_i \,|\, \boldsymbol{\theta})$, $f(\boldsymbol{x} \,|\, \boldsymbol{\theta})$, $\pi(\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta} \,|\, \boldsymbol{x})$
- Show that the posterior law $\pi(\boldsymbol{\theta} \,|\, \boldsymbol{x})$ is Gaussian, *i.e.*

$$\pi(\boldsymbol{\theta} \,|\, \boldsymbol{x}) \sim \mathcal{N}\left(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Sigma}}\right) \tag{6.82}$$

  and give the expressions of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Sigma}}$.

- Show that the MAP estimate is obtained by

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \{J(\boldsymbol{\theta})\} \tag{6.83}$$

  with

$$\begin{aligned} J(\boldsymbol{\theta}) &= \frac{1}{\sigma_e^2}\|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\|^2 + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t \boldsymbol{\Sigma}_\theta^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ &= \frac{1}{\sigma_e^2}\sum_{i=1}^{n}(x_i - \theta_0 - \theta_1 t_i)^2 + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t \boldsymbol{\Sigma}_\theta^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \end{aligned}$$

- Show that, in this case an explicite expression of $\widehat{\boldsymbol{\theta}}$ is available and is given by

$$\widehat{\boldsymbol{\theta}} = \left[\frac{1}{\sigma_e^2}\boldsymbol{H}^t\boldsymbol{H} + \boldsymbol{\Sigma}_\theta^{-1}\right]^{-1} \left[\frac{1}{\sigma_e^2}\boldsymbol{H}^t\boldsymbol{x} + \boldsymbol{\Sigma}_\theta^{-1}\boldsymbol{\theta}_0\right] \tag{6.84}$$

- Compare this solution to the ML solution (6.79) which is equal to least square (LS) solution.

Model 2:
A little more complex model is

$$x_i = x(t_i) = \theta_0 + \theta_1 t_i + e_i, \quad i = 1, \ldots, n \tag{6.85}$$

with the assumption that

$$r_i = e_i \cos\phi = \frac{e_i}{\sqrt{1+\theta_1^2}} = \frac{x_i - \theta_0 - \theta_1 t_i}{\sqrt{1+\theta_1^2}}$$

the distance of the point (t_i, x_i) to the line $x(t_i) = \theta_0 + \theta_1 t_i$ is zero mean, white and Gaussian with known variance σ_r^2 (See figure 6.5.1.)

Note that r_i is no more a linear function of θ_1.

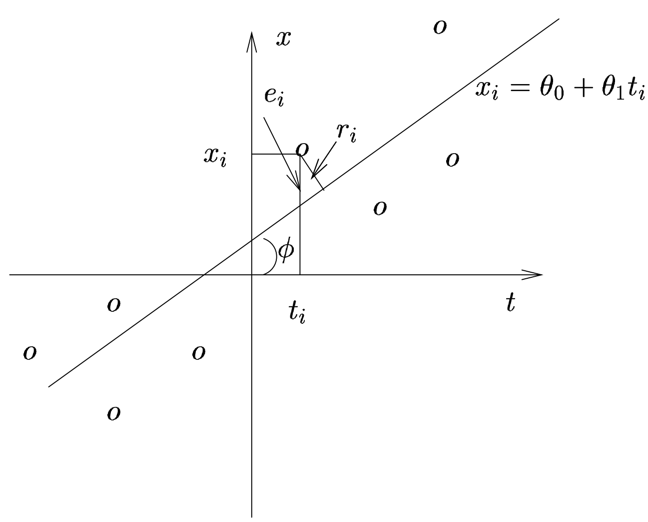


Figure 6.3: Line fitting: model 2: $r_i = e_i \cos\phi = \frac{e_i}{\sqrt{1+\theta_1^2}} = \frac{x_i-\theta_0-\theta_1 t_i}{\sqrt{1+\theta_1^2}}$

Exercise 2: With this model and assuming that r_i are zero mean, white and Gaussian with known variance $\sigma^2 = 1$:

- Write the expressions of $f(x_i \,|\, \boldsymbol{\theta})$, $f(\boldsymbol{x} \,|\, \boldsymbol{\theta})$, $\pi(\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta} \,|\, \boldsymbol{x})$
- Show that the MAP estimate is obtained by

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \{J(\boldsymbol{\theta})\} \tag{6.86}$$

with

$$J(\boldsymbol{\theta}) = n \ln\left(2\pi(1+\theta_1^2)\sigma_r^2\right) + \frac{1}{(1+\theta_1^2)\sigma_r^2} \sum_{i=1}^{n} (x_i - \theta_0 - \theta_1 t_i)^2 + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t \Sigma_\theta^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \tag{6.87}$$

where $\boldsymbol{\theta} = [\theta_0, \theta_1]^t$.

- Is it possible to obtain explicit expressions for $\widehat{\theta}_0$ and $\widehat{\theta}_1$?

Model 3:

A little different model assumes that t_i are also uncertain, *i.e.*

$$x_i = x(t_i) = \theta_0 + \theta_1(t_i + \epsilon_i) + e_i, \quad i = 1, \ldots, n \tag{6.88}$$

where ϵ_i represents the error on t_i. Here also, we have $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{e}$ with $\boldsymbol{\theta} = [\theta_0, \theta_1]^t$, but the matrix $\boldsymbol{H}$ is now uncertain.

Note that we have

$$x_i = x(t_i) = \theta_0 + \theta_1 t_i + \theta_1 \epsilon_i + e_i, \quad i = 1, \ldots, n \tag{6.89}$$

which can also be written as

$$\boldsymbol{x} = \boldsymbol{H}_0\boldsymbol{\theta} + \boldsymbol{H}_\epsilon\boldsymbol{\theta} + \boldsymbol{e}$$

with

$$\boldsymbol{H}_0 = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ & \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}, \quad \boldsymbol{H}_\epsilon = \begin{pmatrix} 0 & \epsilon_1 \\ \vdots & \vdots \\ & \\ \vdots & \vdots \\ 0 & \epsilon_n \end{pmatrix},$$

Exercise 3: With this model and assuming that ϵ_i are zero mean, white and Gaussian with known variance σ_ϵ^2 and that e_i are also zero mean, white and Gaussian with known variance σ_e^2:

- Write the expressions of $f(x_i \,|\, \boldsymbol{\theta})$, $f(\boldsymbol{x} \,|\, \boldsymbol{\theta})$, $\pi(\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta} \,|\, \boldsymbol{x})$
- Give the expressions of the ML and the MAP estimators.
- Compare them to the solutions of the previous cases.

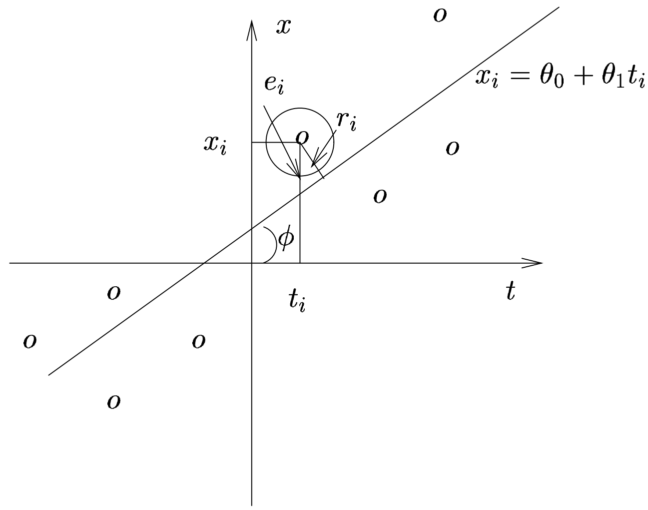


Figure 6.4: Line fitting: model 3

Model 4:

This is the combination of cases 2 and 3, *i.e.*

$$x_i = x(t_i) = \theta_0 + \theta_1 (t_i + \epsilon_i) + e_i, \quad i = 1, \ldots, n \tag{6.90}$$

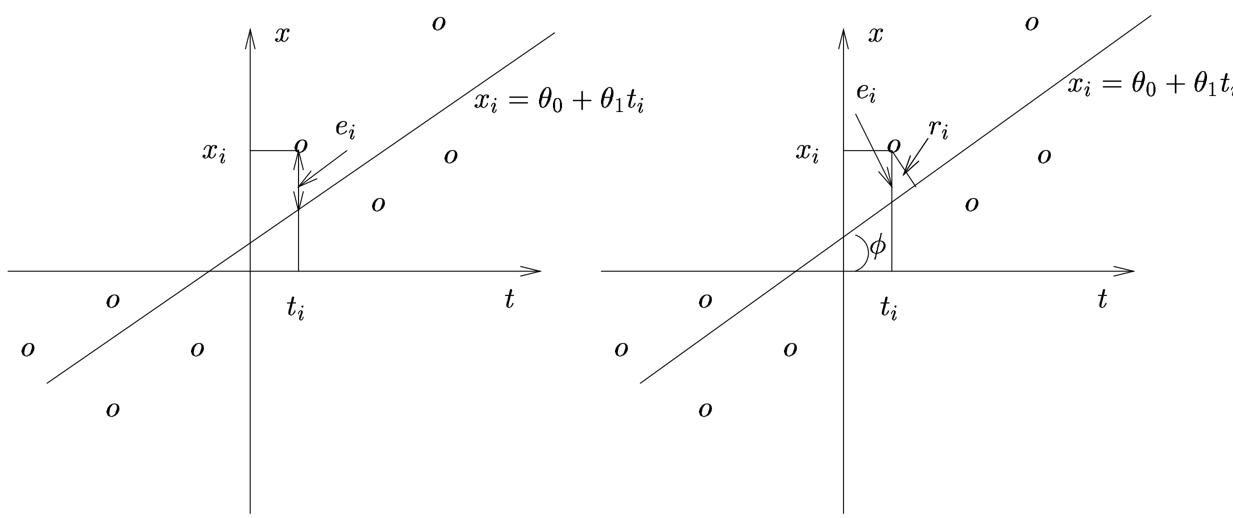
where

$$r_i = e_i \cos\phi = \frac{e_i}{\sqrt{1+\theta_1^2}} = \frac{x_i - \theta_0 - \theta_1 t_i}{\sqrt{1+\theta_1^2}}$$

the distance of the point (t_i, x_i) to the line $x(t_i) = \theta_0 + \theta_1 t_i$ are assumed zero mean, white and Gaussian.

Exercise 4: With this model and assuming that ϵ_i are zero mean, white and Gaussian with known variance σ_ϵ^2 and that r_i are also zero mean, white and Gaussian with known variance σ_r^2:

- Write the expressions of $f(x_i \mid \boldsymbol{\theta})$, $f(\boldsymbol{x} \mid \boldsymbol{\theta})$, $\pi(\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta} \mid \boldsymbol{x})$
- Give the expressions of the ML and the MAP estimators.
- Compare them to the solutions in previous examples.



Model 1: $e_i = x_i - (\theta_0 + \theta_1 t_i)$ Model 2: $r_i = [x_i - (\theta_0 + \theta_1 t_i)] \cos\phi = \frac{x_i - (\theta_0 + \theta_1 t_i)}{\sqrt{1+\theta_1^2}}$

Figure 6.5: Line fitting: models 1 and 2.

$$\begin{array}{lcl}
\boldsymbol{x}|\boldsymbol{\theta} & \sim & \mathcal{N}(\boldsymbol{H}\boldsymbol{\theta},\, \boldsymbol{R}_N) \\
\boldsymbol{\theta} & \sim & \mathcal{N}(\boldsymbol{\theta}_0,\, \boldsymbol{R}_\Theta) \\
\boldsymbol{\theta}|\boldsymbol{x} & \sim & \mathcal{N}(\widehat{\boldsymbol{\theta}},\, \widehat{\boldsymbol{R}}) \\
\widehat{\boldsymbol{\theta}} & = & \boldsymbol{\theta}_0 + \boldsymbol{R}_\Theta \boldsymbol{H}^t \left[\boldsymbol{H}\boldsymbol{R}_\Theta \boldsymbol{H}^t + \boldsymbol{R}_N\right]^{-1} (\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}_0) \\
 & = & \boldsymbol{\theta}_0 + \left[\boldsymbol{H}^t \boldsymbol{R}_N^{-1} \boldsymbol{H} + \boldsymbol{R}_\Theta^{-1}\right]^{-1} \boldsymbol{H}^t \boldsymbol{R}_N^{-1} (\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}_0) \\
\widehat{\boldsymbol{R}} & = & \boldsymbol{R}_\Theta - \boldsymbol{R}_\Theta \boldsymbol{H}^t \left[\boldsymbol{H}\boldsymbol{R}_\Theta \boldsymbol{H}^t + \boldsymbol{R}_N\right]^{-1} \boldsymbol{H}\boldsymbol{R}_\Theta \\
 & = & \left[\boldsymbol{H}^t \boldsymbol{R}_N^{-1} \boldsymbol{H} + \boldsymbol{R}_\Theta^{-1}\right]^{-1}
\end{array}$$

Model 1:

$$\begin{array}{lcl}
e_i & \sim & \mathcal{N}(0,\, \sigma_e^2) \\
x_i|\boldsymbol{\theta} & \sim & \mathcal{N}(\theta_0 + \theta_1 t_i,\, \sigma_e^2) \\
\boldsymbol{x}|\boldsymbol{\theta} & \sim & \mathcal{N}(\boldsymbol{H}\boldsymbol{\theta},\, \sigma_e^2 \boldsymbol{I}) \\
\widehat{\boldsymbol{\theta}} & = & \boldsymbol{\theta}_0 + \left[\boldsymbol{H}^t \boldsymbol{H} + \sigma_e^2 \boldsymbol{R}_\Theta^{-1}\right]^{-1} \boldsymbol{H}^t (\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}_0) \\
\widehat{\boldsymbol{R}} & = & \sigma_e^2 \left[\boldsymbol{H}^t \boldsymbol{H} + \sigma_e^2 \boldsymbol{R}_\Theta^{-1}\right]^{-1}
\end{array}$$

Model 2:

$$\begin{array}{lcl}
r_i & = & \frac{e_i}{\sqrt{1+\theta_1^2}} \sim \mathcal{N}(0,\, \sigma_r^2) \\
e_i & \sim & \mathcal{N}\left(0,\, (1+\theta_1^2)\sigma_r^2\right) \\
x_i|\boldsymbol{\theta} & \sim & \mathcal{N}\left(\theta_0 + \theta_1 t_i,\, (1+\theta_1^2)\sigma_r^2\right) \\
\boldsymbol{x}|\boldsymbol{\theta} & \sim & \mathcal{N}\left(\boldsymbol{H}\boldsymbol{\theta},\, (1+\theta_1^2)\sigma_r^2 \boldsymbol{I}\right) \\
\pi(\theta|\boldsymbol{x}) & = & \frac{1}{m(\boldsymbol{x})} \left[2\pi(1+\theta_1^2)\sigma_e^2\right]^{-n/2} \exp\left[-\frac{1}{2(1+\theta_1^2)\sigma_e^2}\|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\|^2 - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t \boldsymbol{R}_\Theta^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right] \\
\widehat{\boldsymbol{\theta}}_{MAP} & = & \arg\min_{\boldsymbol{\theta}} \left\{J(\boldsymbol{\theta})\right\} \\
J(\boldsymbol{\theta}) & = & -\frac{n}{2}\ln[2\pi(1+\theta_1^2)\sigma_e^2] - \frac{1}{2(1+\theta_1^2)\sigma_e^2}\|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\|^2 - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t \boldsymbol{R}_\Theta^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)
\end{array}$$

Model 3:

$$\begin{array}{lcl}
\epsilon_i & \sim & \mathcal{N}(0,\, \sigma_\epsilon^2) \\
e_i & \sim & \mathcal{N}(0,\, \sigma_e^2) \\
x_i|\boldsymbol{\theta} & \sim & \mathcal{N}\left(\theta_0 + \theta_1 t_i,\, \theta_1^2\sigma_\epsilon^2 + \sigma_e^2\right) \\
\boldsymbol{x}|\boldsymbol{\theta} & \sim & \mathcal{N}\left(\boldsymbol{H}\boldsymbol{\theta},\, (\theta_1^2\sigma_\epsilon^2 + \sigma_e^2)\boldsymbol{I}\right) \\
\pi(\theta|\boldsymbol{x}) & = & \frac{1}{m(\boldsymbol{x})} \left[2\pi(\theta_1^2\sigma_\epsilon^2 + \sigma_e^2)\right]^{-n/2} \exp\left[-\frac{1}{2(\theta_1^2\sigma_\epsilon^2 + \sigma_e^2)}\|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\|^2 - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t \boldsymbol{R}_\Theta^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right] \\
\widehat{\boldsymbol{\theta}}_{MAP} & = & \arg\min_{\boldsymbol{\theta}} \left\{J(\boldsymbol{\theta})\right\} \\
J(\boldsymbol{\theta}) & = & -\frac{n}{2}\ln[2\pi(\theta_1^2\sigma_\epsilon^2 + \sigma_e^2)] - \frac{1}{2(\theta_1^2\sigma_\epsilon^2 + \sigma_e^2)}\|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\|^2 - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t \boldsymbol{R}_\Theta^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)
\end{array}$$

Model 4:

$$\begin{array}{lcl}
\epsilon_i & \sim & \mathcal{N}(0,\, \sigma_\epsilon^2) \\
e_i & \sim & \mathcal{N}(0,\, (1+\theta_1)^2\sigma_e^2) \\
x_i|\boldsymbol{\theta} & \sim & \mathcal{N}\left(\theta_0 + \theta_1 t_i,\, \theta_1^2\sigma_\epsilon^2 + (1+\theta_1)^2\sigma_e^2\right) \\
 & = & \mathcal{N}\left(\theta_0 + \theta_1 t_i,\, \theta_1^2(\sigma_\epsilon^2 + \sigma_e^2) + \sigma_e^2\right) \\
\boldsymbol{x}|\boldsymbol{\theta} & \sim & \mathcal{N}\left(\boldsymbol{H}\boldsymbol{\theta},\, (\theta_1^2(\sigma_\epsilon^2 + \sigma_e^2) + \sigma_e^2)\boldsymbol{I}\right) \\
\pi(\theta|\boldsymbol{x}) & = & \frac{1}{m(\boldsymbol{x})} \left[2\pi(\theta_1^2(\sigma_\epsilon^2 + \sigma_e^2) + \sigma_e^2)\right]^{-n/2} \exp\left[-\frac{1}{2(\theta_1^2(\sigma_\epsilon^2 + \sigma_e^2) + \sigma_e^2)}\|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\|^2 - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t \boldsymbol{R}_\Theta^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right] \\
\widehat{\boldsymbol{\theta}}_{MAP} & = & \arg\min_{\boldsymbol{\theta}} \left\{J(\boldsymbol{\theta})\right\} \\
J(\boldsymbol{\theta}) & = & -\frac{n}{2}\ln[(\theta_1^2(\sigma_\epsilon^2 + \sigma_e^2) + \sigma_e^2)] - \frac{1}{2(\theta_1^2(\sigma_\epsilon^2 + \sigma_e^2) + \sigma_e^2)}\|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}\|^2 - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t \boldsymbol{R}_\Theta^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)
\end{array}$$

Remark: To do these calculations easily we need the following relations:

- If $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{A}+\boldsymbol{B}$ are invertible, then we have

$$\begin{aligned} \left[\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}\right]^{-1} &= \boldsymbol{A}\left[\boldsymbol{A}+\boldsymbol{B}\right]^{-1}\boldsymbol{B} = \boldsymbol{B}\left[\boldsymbol{A}+\boldsymbol{B}\right]^{-1}\boldsymbol{A} \\ \left[\boldsymbol{A}+\boldsymbol{B}\right]^{-1} &= \boldsymbol{A}^{-1}\left[\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}\right]^{-1}\boldsymbol{B}^{-1} = \boldsymbol{B}^{-1}\left[\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}\right]^{-1}\boldsymbol{A}^{-1} \end{aligned} \tag{6.91}$$

- If $\boldsymbol{A}$ and $\boldsymbol{C}$ are invertible matrices, then we have

$$\left[\boldsymbol{A}+\boldsymbol{B}\boldsymbol{C}\boldsymbol{D}\right]^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}\left[\boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B}+\boldsymbol{C}^{-1}\right]^{-1}\boldsymbol{D}\boldsymbol{A}^{-1} \tag{6.92}$$

- A special case very useful in system theory

$$\left[\boldsymbol{I}+\boldsymbol{B}(s\boldsymbol{I}-\boldsymbol{C})^{-1}\boldsymbol{D}\right]^{-1} = \boldsymbol{I}-\boldsymbol{B}\left[s\boldsymbol{I}-\boldsymbol{C}+\boldsymbol{D}\boldsymbol{B}\right]^{-1}\boldsymbol{D} \tag{6.93}$$

- If $\boldsymbol{A}$ is invertible then,

$$\left[\boldsymbol{A}+\boldsymbol{u}\boldsymbol{v}^t\right]^{-1} = \boldsymbol{A}^{-1} - \frac{(\boldsymbol{A}^{-1}\boldsymbol{u})(\boldsymbol{v}^t\boldsymbol{A}^{-1})}{1+\boldsymbol{v}^t\boldsymbol{A}^{-1}\boldsymbol{u}} \tag{6.94}$$

- If $\boldsymbol{A}$ is a bloc matrix

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}$$

then $\boldsymbol{B}=\boldsymbol{A}^{-1}$ is also a bloc matrix

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{pmatrix}$$

and

  - If $\boldsymbol{A}_{22}^{-1}$ exists, then

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ 0 & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} & 0 \\ 0 & \boldsymbol{A}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & 0 \\ \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} & \boldsymbol{I} \end{pmatrix}$$

and we have

$$\text{rank}\left\{\boldsymbol{A}\right\} = \text{rank}\left\{\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\right\} + \text{rank}\left\{\boldsymbol{A}_{22}\right\}$$

$\boldsymbol{A}^{-1}$ exists iff the matrix $\boldsymbol{T}=\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$ is invertible. Then we have

$$\begin{aligned} \boldsymbol{B}_{11}=\boldsymbol{T}^{-1} &= \left(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\right)^{-1} \\ \boldsymbol{B}_{22}= &= \boldsymbol{A}_{22}^{-1}+\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{B}_{11}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ \boldsymbol{B}_{12} &= -\boldsymbol{B}_{11}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ \boldsymbol{B}_{21} &= -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{B}_{11} \end{aligned}$$

Written differently, we have

$$\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} (\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1} & -\boldsymbol{B}_{11}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{B}_{11} & \boldsymbol{A}_{22}^{-1}+\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{B}_{11}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \end{pmatrix}$$

– If $\boldsymbol{A}_{22}^{-1}$ exists, then we have

$$\begin{aligned}
\boldsymbol{B}_{11} = &= \boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{B}_{22}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} \\
\boldsymbol{B}_{22} = \boldsymbol{D}^{-1} &= (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1} \\
\boldsymbol{B}_{12} &= -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{B}_{22} \\
\boldsymbol{B}_{21} &= -\boldsymbol{B}_{22}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}
\end{aligned}$$

The matrices $\boldsymbol{T}$ and $\boldsymbol{D}$ are called the *Shur's complement* of the matrix $\boldsymbol{A}$.

- **Particular case 1 :**
If $\boldsymbol{A}$ is a superior bloc-triangular, *i.e.* $\boldsymbol{A}_{21} = \boldsymbol{0}$, then $\boldsymbol{B}$ is also superior bloc-triangular, *i.e.*

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{A}_{11}^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ \boldsymbol{0} & \boldsymbol{A}_{22}^{-1} \end{pmatrix}$$

- **Particular case 2 :**
If $\boldsymbol{A}$ is an inferior bloc-triangular matrix, *i.e.* $\boldsymbol{A}_{12} = \boldsymbol{0}$, then $\boldsymbol{B}$ is also an inferior bloc-triangular matrix, *i.e.*

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{A}_{11}^{-1} & \boldsymbol{0} \\ -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{A}_{22}^{-1} \end{pmatrix}$$

- **Particular case 3 :**
If $\boldsymbol{A}_{22}$ is a sclar and $\boldsymbol{A}_{21}$ and $\boldsymbol{A}_{12}$ are vectors, we have

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{x} \\ \boldsymbol{z}^t & y \end{pmatrix}, \quad \boldsymbol{B} = \boldsymbol{A}^{-1} = \frac{1}{\alpha}\begin{pmatrix} \alpha\boldsymbol{A}_{11}^{-1} + \boldsymbol{w}\boldsymbol{v}^t & \boldsymbol{w} \\ \boldsymbol{v}^t & 1 \end{pmatrix}$$

where α, $\boldsymbol{w}$, and $\boldsymbol{v}$ are given by:

$$\alpha = \frac{1}{(y - \boldsymbol{z}^t\boldsymbol{A}_{11}^{-1}\boldsymbol{x})} = \frac{|\boldsymbol{A}|}{|\boldsymbol{A}_{11}|}, \quad \boldsymbol{w} = -\boldsymbol{A}_{11}^{-1}\boldsymbol{x}, \quad \boldsymbol{v} = -\boldsymbol{A}_{11}^{-t}\boldsymbol{z}$$

- If $\boldsymbol{A}$ is a $[N, P]$ matrix

$$\boldsymbol{I}_N \pm \boldsymbol{A}\boldsymbol{A}^t = [\boldsymbol{I}_N \pm \boldsymbol{A}\boldsymbol{R}^{-1}\boldsymbol{A}^t].[\boldsymbol{I}_N \pm \boldsymbol{A}\boldsymbol{R}^{-1}\boldsymbol{A}^t]^t$$

where $\boldsymbol{R} = \boldsymbol{I}_{\boldsymbol{P}} + [\boldsymbol{I}_{\boldsymbol{P}} \pm \boldsymbol{A}^t\boldsymbol{A}]^{1/2}$.

- If $\boldsymbol{x}$ is a vector and $u(\boldsymbol{x})$ a scalar function of $\boldsymbol{x}$ and if we define the gradient vector $\nabla u = \frac{\partial u}{\partial \boldsymbol{x}} = \left[\frac{\partial u}{\partial x_i}\right]$, then we have the following relations
– If $u = \boldsymbol{\theta}^t\boldsymbol{x}$ then $\nabla u = \frac{\partial u}{\partial \boldsymbol{x}} = \boldsymbol{\theta}$
– If $u = \boldsymbol{x}^t\boldsymbol{A}\boldsymbol{x}$ then $\nabla u = \frac{\partial u}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}$

Generalized Gaussian

$$p(x) = \frac{p^{1-1/p}}{2\sigma\Gamma(1/p)} \exp\left[-\frac{1}{p}\frac{|x-x_0|^p}{\sigma^p}\right]$$

$$p = 1 \longrightarrow p(x) = \frac{1}{2\sigma} \exp\left[-\frac{|x-x_0|}{\sigma}\right]$$

$$p = 2 \longrightarrow p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{|x-x_0|^2}{\sigma^2}\right]$$

$$p = \infty \longrightarrow p(x) = \begin{cases} 1/2\sigma & \text{if} \quad |x-x_0| < \sigma \\ 0 & \text{otherwise} \end{cases}$$

Centered case $x_0 = 0$.

$$p(x) = \frac{p^{1-1/p}}{2\sigma\Gamma(1/p)} \exp\left[-\frac{1}{p}\frac{|x|^p}{\sigma^p}\right]$$

$$p = 1 \longrightarrow p(x) = \frac{1}{2\sigma} \exp\left[-\frac{|x|}{\sigma}\right]$$

$$p = 2 \longrightarrow p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{|x|^2}{\sigma^2}\right]$$

$$p = \infty \longrightarrow p(x) = \begin{cases} 1/2\sigma & \text{if} \quad |x| < \sigma \\ 0 & \text{otherwise} \end{cases}$$

Multivariable case:
Separable:

$$p(\boldsymbol{x}) = \frac{p^{n-n/p}}{2^n\sigma^n\Gamma^n(1/p)} \exp\left[-\frac{1}{p\sigma^p}\sum_{i=1}^{n}|x_i|^p\right]$$

Correlated: Markov models

$$p(\boldsymbol{x}) = Z(\alpha)\exp\left[-\alpha\sum_{i=1}^{n}\sum_{j\ i}\phi(x_i - x_j)\right]$$

$$Z(\alpha) = \int \exp\left[-\alpha\sum_{i=1}^{n}\sum_{j\ i}\phi(x_i - x_j)\right] \mathrm{d}\boldsymbol{x}$$

Example: $\phi(x) = x^2, \quad j\ i = i - 1$

$$p(\boldsymbol{x}) = Z(\alpha)\exp\left[-\alpha x_1^2 - \alpha\sum_{i=2}^{n}(x_i - x_{i-1})^2\right]$$

This can be written as

$$p(\boldsymbol{x}) = Z(\alpha)\exp\left[-\alpha\boldsymbol{x}^t\boldsymbol{D}^t\boldsymbol{D}\boldsymbol{x}\right]$$

with

$$\boldsymbol{D} = \begin{pmatrix} 1 & 0 & & & & 0 \\ 1 & -1 & & & & \vdots \\ 0 & 1 & -1 & & & \\ \vdots & & 1 & -1 & & \\ & & & & & \\ 0 & \cdots & & 0 & 1 & -1 \end{pmatrix}$$

$$Z(\alpha) = (2\pi)^{-n/2}\,(2\alpha)^{n/2}|\boldsymbol{D}^t\boldsymbol{D}|$$

Extension :

$$p(\boldsymbol{x}) = Z(\alpha)\exp\left[-\alpha\phi(x_1) - \alpha\sum_{i=2}^{n}\phi(x_i - x_{i-1})\right] = Z(\alpha)\exp\left[-\alpha\sum_{i=1}^{n}\phi([\boldsymbol{D}\boldsymbol{x}]_i)\right]$$

with $\phi(x) = |x|^p$

The questions are:

$Z(\alpha)$ exists ?

Can we obtain an analytical expression for it ?