1/31/2025

Friday, January 31, 2025 11:59 AM

Maximum Likelihood estimation (MLE)

- We can also consider the likelihood function for a random sample $\vec{y} = [Y_1, \dots, Y_{n_y}]$ $(\vec{\theta} \in \Gamma)$

$$L(\vec{\theta} \mid \vec{y}) = f_Y(\vec{y} \mid \theta), \quad L : \Gamma \to [0, \infty)$$

where $f_Y(\vec{y} \mid \theta)$ is the parameter-dependent joint pdf for $\vec{Y}$.

- Note that if Y_i are iid, then

$$L(\vec{\theta} \mid \vec{y}) = f_Y(\vec{y} \mid \vec{\theta}) = \prod_{i=1}^{n_y} f_{Y_i}(y_i \mid \vec{\theta})$$

- Note: since $L : \Gamma \to [0, \infty)$, we often look at the log-likelihood,

$$\ln(L(\hat{\theta} \mid \vec{y})) = LL(\hat{\theta} \mid \vec{y})$$

$\left(y_i = f(x_i; \vec{\theta}) + \epsilon_i \right)$

- If we assume $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, then $Y_i \sim N\left(f(x_i; \vec{\theta}), \sigma^2\right)$ & σ^2 is unknown

$$L(\vec{\theta}, \sigma^2 \mid \vec{y}) = \prod_{i=1}^{n_y} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2} \left(y_i - f(x_i; \vec{\theta}) \right)^2 \right)$$

$$\Rightarrow L(\vec{\theta}, \sigma^2 \mid \vec{y}) = \frac{1}{(2\pi\sigma^2)^{n_y/2}} \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^{n_y} \left(y_i - f(x_i; \vec{\theta}) \right)^2 \right)$$

- The maximum likelihood estimator, MLE, is

$$[\vec{\theta}, \sigma^2]_{MLE} = \underset{\substack{\vec{\theta} \in \Gamma \\ \sigma^2 \in [0, \infty)}}{\arg\max} \; L(\vec{\theta}, \sigma^2 \mid \vec{y})$$

- Note:

$$\ln\left(L(\vec{\theta}, \sigma^2 \mid \vec{y})\right) = -\frac{n_y}{2} \ln(2\pi) - \frac{n_y}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n_y} \left(y_i - f(x_i; \vec{\theta}) \right)^2$$

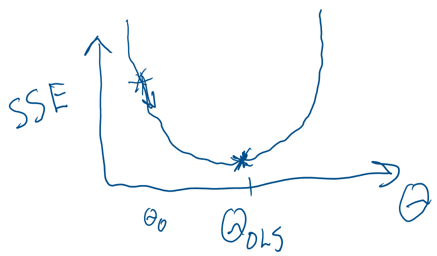
$$\ln(L(\vec{\theta},\sigma^2|\vec{y})) = -\frac{n_y}{2}\ln(2\pi) - \frac{n_y}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n_y}(y_i - f(\vec{x}_i;\vec{\theta}))^2$$

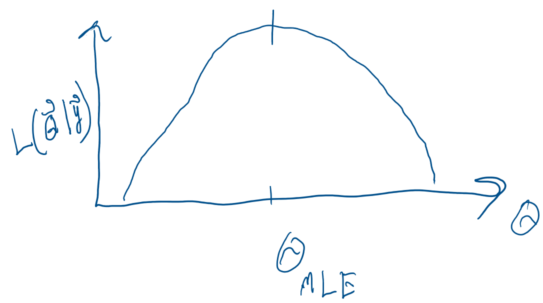
$$\propto -\frac{n_y}{2}\ln(\sigma^2) - \frac{1}{\sigma^2}\sum_{i=1}^{n_y}(y_i - f(\vec{x}_i;\vec{\theta}))^2$$

$\Rightarrow$ If σ^2 is known

$$\Rightarrow \ln(L(\vec{\theta}|\vec{y},\sigma^2)) \propto -\sum_{i=1}^{n_y}(y_i - f(\vec{x}_i;\vec{\theta}))^2$$

★ when $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$, MLE + OLS solution should overlap.





★ How do we find θ_{OLS} or θ_{MLE}?

- For θ_{OLS}, grad. descent
- For θ_{MLE}, do the same for $-LL(\vec{\theta}|\vec{y})$

$$\frac{\partial}{\partial\theta_i} L(\vec{\theta}|\vec{y},\sigma^2) = 0$$

$$\Rightarrow \sum_{i=1}^{n_y}(y_i - f(\vec{x}_i;\vec{\theta}))\nabla_\theta f(\vec{x}_i;\vec{\theta}) = 0$$

Chapter 5 | Representation of Random Inputs

- Recall that the probability triple $(\Omega, \mathcal{F}, P)$, where
  - Ω is the domain
  - $\mathcal{F}$ is the Borel σ-algebra
  - $P : \mathcal{F} \to [0,1]$

- Let $\Gamma = \Theta(\Omega) \subset \mathbb{R}^p$ is the image of $\vec{\theta}$.
  - Assume $\vec{\theta}$ has a joint pdf, $\rho(\vec{\theta})$ with support Γ.

with

· Assume θ has a joint pdf, $\rho(\theta)$

- we define $\mathcal{B}(\Gamma)$ as the σ-algebra of Γ. Then the set with $\Omega(\omega) \subseteq \Gamma$

$$\left(\Gamma,\ \mathcal{B}(\Gamma),\ \rho(\vec{\theta})\,d\vec{\theta}\right)$$

defines the probability space of Γ!

· $\rho(\vec{\theta})\,d\vec{\theta}$ is the probability measure

---

Ex)

$$\frac{\partial u}{\partial t} = \bar{\kappa}\,\frac{\partial^2 u}{\partial x^2}, \quad u \equiv \text{temperature},\ \bar{\kappa} \equiv \text{thermal diffusivity}$$

$$x \in [0, L],\quad t \in [0, T],\quad u(0,t) = u_0,\ u(L,t) = u_L$$

$$u(x,0) = u_{IC}$$

$\Rightarrow$ Given $y_i = u(\vec{x}, t_i) + \varepsilon_i$

* what are $\vec{\theta} = [\bar{\kappa}, u_0, u_L, u_{IC}]$?

Random Fields

* This will cover parameters that are spatially/temporally correlated *

- Consider a correlated, mean squared continuous random field

$$\alpha(x, \omega),\quad x \in D,\quad +\quad \omega \equiv \text{random process}$$

with the following properties:

$$E[\alpha(x,\omega)] = \bar{\alpha}(x), \quad \mathrm{Cov}[\alpha(x,\omega)] = C(x, x') \quad \left(= c(x,y)\right)$$

where $$C(x,x') = E\Big[\big(x - E[x]\big)\big(x' - E[x']\big)\Big]$$

Theorem Mercer's Theorem

- Let $C(x,x') = K(x,x')$ be a kernel s.t.

$$K : D \times D \to \mathbb{R}, \quad K(x,x') = K(x',x)$$

$\Rightarrow K$ is symmetric + pos. def.