Hillen's class, Oct. 15, 2024

3 Parameter Estimation

AIC Akaike Information Criterion

[Chapter 7, de Vries et. al. A Course in Math Bio, SIAM 2006.]

Data, Observations

model. ODE e.g. Oncolytic virus model

parameters $p = (A, B, C, D)$

$p \in \mathbb{R}^N$

We assume that the data are typical!

we need a quality measure

$$\mathcal{L}(p \mid data) \quad \text{likelihood}$$

assume $\mathcal{L}(p \mid data) = Pr(data \mid p)$

Definition. $p \in \mathbb{R}^n$ $\mathcal{L}: \mathbb{R}^n \longrightarrow \mathbb{R}_+$

which maps the parameters $\pi \in \mathbb{R}^n$ to the probability to find given data $\mathcal{L}(\pi) = P_r(\text{data}|\pi)$

$\mathcal{L}$ is called the likelihood.

The log likelihood $\ln \mathcal{L}(\pi) = \mathcal{LL}(\pi)$

A parameter $\hat{\pi}$ that maximizes $\mathcal{L}$ (a $\mathcal{LL}$) is a maximum likelihood estimator.

Example 1. (time to cell death)

$$\dot{p} = -\mu p$$

p pop

μ parameter const.

$$p(t) = p(0)\, e^{-\mu t}$$

death probability $d(t) = 1 - p(t)$

$$= 1 - p(0) e^{-\mu t}$$

probability density of death $f(t) = \frac{d}{dt} d(t)$

$$\dot{d}(t) = p(0) \mu e^{-\mu t}$$

assume $p(0) = 1$ $\quad f(t) = \mu e^{-\mu t}$ $\Leftarrow$

i.e. death is exponentially distributed and $\frac{1}{\mu}$ is the expected survival time.

Suppose n independent cells die at ages

$$a_1, a_2, \dots a_n$$

$$\underline{\frac{1}{\mu} \sim a_i} \qquad \underline{\mu \sim \frac{1}{a_i}}$$

How to estimate μ?

1) average $\frac{1}{a_1}, \frac{1}{a_2}, \dots, \frac{1}{a_n}$: $\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a_i}$

2) average $a_1, a_2, \dots a_n$: $\hat{\mu}_2 = \underline{\frac{1}{\bar{a}}}$ $\quad \underline{\bar{a} = \frac{1}{n} \sum_{i=1}^{n} a_i}$

$\hat{\mu}_1 \neq \hat{\mu}_2$ in general.

Likelihood $\underline{\mathcal{L}(\mu) = f(a_1) f(a_2) f(a_3) \dots f(a_n)}$

$$= \mu e^{-\mu a_1} \mu e^{-\mu a_2} \dots \mu e^{-\mu a_n}$$

$$= \mu^n \exp\left(-\mu \sum_{i=1}^{n} a_i\right)$$

$$\boxed{\mathcal{L}\ell(\mu) = n \ln \mu - \mu \sum_{i=1}^{n} a_i} \leftarrow$$

$$\frac{d}{d\mu} \mathcal{LL}(\mu) = \frac{n}{\mu} - \sum_{i=1}^{n} a_i = 0$$

$$\Rightarrow \frac{1}{\mu} = \frac{1}{n} \sum a_i$$

$$\mu = \left(\frac{1}{n} \sum_{i=1}^{n} a_i \right)^{-1} = \hat{\mu}_2$$

$$\frac{d^2}{d\mu^2} \mathcal{LL}(\mu) = -\frac{n}{\mu^2} < 0 \Rightarrow \hat{\mu}_2 \text{ maximum likelihood estimator !}$$

Example: death times 1, 15, 3, 16, 18, 11, 20, 14, 13 $= a_i$

$$\bar{a} = \frac{1}{n} \sum a_i = 12.\overline{3}$$

$$\hat{\mu}_2 = \frac{1}{\bar{a}} = 0.081$$

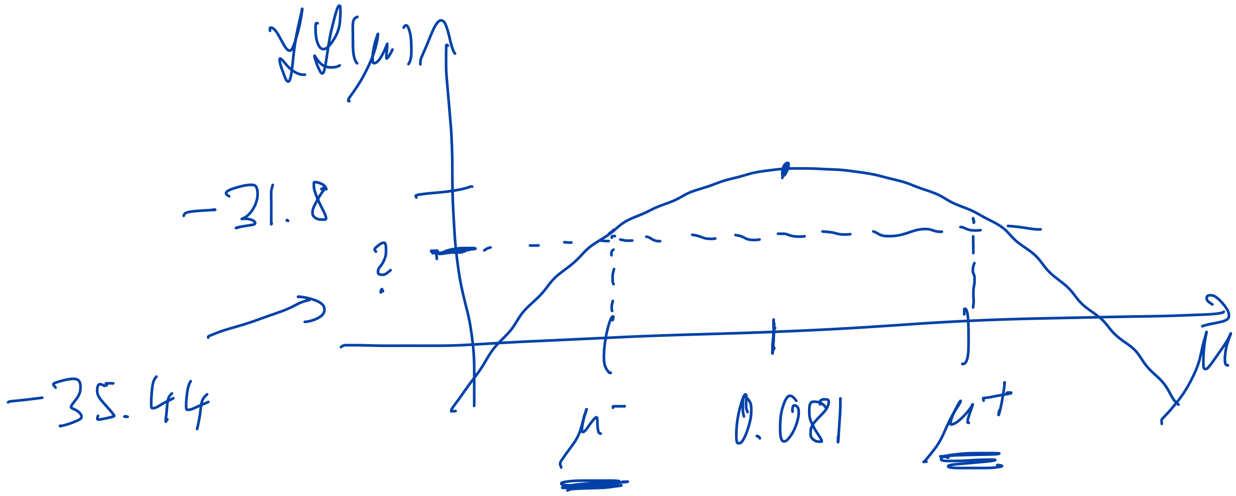
$$\max \mathcal{LL}(\hat{\mu}) = -31.6$$

Confidence Intervals

<u>Def</u>: Given a confidence level γ (0.95 or 0.99) an interval $[p^-, p^+]$ the covers the true parameter p with probability γ is called

a confidence interval.

<u>Example:</u> Cell death $\quad \max \mathcal{LL} = -31.6$



In most cases we assume $\mathcal{LL}(\mu)$ is χ^2_1 distributed. (one degree of freedom).

For example $\gamma = 0.95$ $\quad$ tolerance $\alpha = 1 - \gamma$
$= 0.05$

χ^2 table: $\quad \chi^2_{1,\gamma}(0.05) = \underline{3.84}$

Now subtract 3.84 from $\max \mathcal{LL}(\mu)$

$$-31.6 - 3.84 = -35.44$$

Find μ^-, μ^+ from solving

$$\mathcal{LL}(\mu^-) = \mathcal{LL}(\mu^+) = -35.44$$
$$= \max \mathcal{LL}(\mu) - \chi^2_1(\alpha)$$

in our example

$$n \ln \mu^+ - \mu^+ n \bar{a} = n \ln \mu^- - \mu^- n \bar{a}$$

$$= -35.44$$

$$\Rightarrow \mu^- = 0.027, \qquad \mu^+ = 0.18$$

The 95% confidence interval is

$$[0.027, 0.18]$$

and $\hat{\mu}_2 = 0.081$

Least squares

at time points $t_1, t_2, \dots, t_n$

measurements $x_1, x_2, \dots x_n$

which are fitted by a model $f(t, p)$ with parameters p.

$$x_i \simeq \underline{f(t_i, p)} \qquad i = 1, \dots, n$$

We assume

- x_i independent
- normally distributed ←
- with expectation $\mu = f(t_i, p)$
- and variance σ is unknown but const.

$$x_i \sim \underbrace{\mathcal{N}}_{\text{statistical model}}\left(f(t_i, p),\ \sigma^2 \right)$$

ODE model

$$\mathcal{L}((\mu, \sigma^2)) = P(X_i = x_i,\ i = 1, \dots, n \mid (\mu, \sigma))$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i - f(t_i, p))^2}{2\sigma^2}} \quad \leftarrow$$

$$\mathcal{L}\mathcal{L}((\mu, \sigma^2)) = -\sum_{i=1}^{n} \ln(\sqrt{2\pi}\,\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - f(t_i, p))^2$$

↯ $p \neq (\mu, \sigma)$ ↯

ODE stat.

minus

To maximize $\mathcal{L}\mathcal{L}((\mu, \sigma^2))$ we minimize

$\sum_{i=1}^{n} (x_i - f(t_i, p))^2$: <u>least square error</u> ☺

If variance varies

$$X_i \sim \mathcal{N}(f(t_i; p); \sigma_i^2)$$

Weighted least squares error

$$\min\left(\sum_{i=1}^{n} \frac{(X_i - f(t_i; p))^2}{\sigma_i^2}\right)$$

growth curves