Compute the MLE for the different parameters

In this general setting no local maximum.

One assumption: $\sigma_\delta^2 = \lambda \sigma_\varepsilon^2$ for a fixed $\lambda \in \mathbb{R}$

With this assumption

$$L(\cdots | \underline{x}, \underline{y}) = \frac{1}{(2\pi)^n} \frac{\lambda^{n/2}}{\sigma_\delta^{2n}} \left( \exp - \sum_i \frac{(x_i - \xi_i)^2 + \lambda (y_i - \alpha - \beta \xi_i)^2}{2\sigma_\delta^2} \right)$$

We are going to look at local maxima for this function, as a function of the parameters (see Chapter IV on point estimators with MLE)

1) By taking derivatives with respect to $\xi_i$, and put the derivative equal to 0, one gets that a local maximum is observed for

$$\hat{\xi}_i = \frac{x_i + \lambda \beta (y_i - \alpha)}{1 + \lambda \beta^2}$$

By substituting these expressions in L, one finds

$$L(\alpha, \beta, \hat{\xi}_1, \cdots, \hat{\xi}_n, \sigma_\delta^2 | \underline{x}, \underline{y}) = \frac{1}{(2\pi)^n} \frac{\lambda^{n/2}}{\sigma_\delta^{2n}} \exp\left( - \frac{\lambda}{2\sigma_\delta^2 (1 + \lambda\beta^2)} \sum_i (y_i - \alpha - \beta x_i)^2 \right)$$

For computing the MLE for $\alpha$ and $\beta$ we set

$$y_i^* := \sqrt{\lambda}\, y_i \; ; \quad \alpha^* := \sqrt{\lambda}\, \alpha \; ; \quad \beta^* := \sqrt{\lambda}\, \beta \qquad \text{(rescaling)}$$

One gets

$$L(\alpha^*, \beta^*, \hat{\xi}_1, \cdots, \hat{\xi}_n, \sigma_\delta^2 | \underline{x}, \underline{y}) = \frac{1}{(2\pi)^n} \frac{\lambda^{n/2}}{\sigma_\delta^{2n}} \exp\left( - \frac{1}{2\sigma_\delta^2 (1 + \lambda\beta^2)} \sum_i (y_i^* - \alpha^* - \beta^* x_i)^2 \right)$$

(similar expression as at the beginning of the section for data fitting)

By the result for data fitting one gets

$$\hat{\alpha} = \frac{\overline{y^*} - \beta^* \overline{x}}{\sqrt{\lambda}} = \overline{y} - \beta \overline{x}$$

$$\hat{\beta} = \frac{\beta^*}{\sqrt{\lambda}} = \frac{-(S_{xx} - S_{y^*y^*}) + \sqrt{(S_{xx} - S_{y^*y^*})^2 + 4 S_{xy^*}^2}}{2 S_{xy^*}}$$

$$= \frac{-(S_{xx} - \lambda S_{yy}) + \sqrt{(S_{xx} - \lambda S_{yy})^2 + 4\lambda S_{xy}^2}}{2\lambda S_{xy}}$$

Remark:

For $\lambda = 1$, we get the result of the data fitting (obtained with the total least square)

This can be considered as a justification of the total least square.

Remark: From

$$L(\hat\alpha, \hat\beta, \hat\xi_1, \cdots, \hat\xi_n, \sigma_\delta^2 \mid \underline{x}, \underline{y}) = \frac{1}{(2\pi)^{n/2}} \frac{\lambda^{n/2}}{\sigma_\delta^{2n}} \exp\left(-\frac{\lambda}{2\sigma_\delta^2(1+\lambda\hat\beta^2)} \sum_i (y_i - \hat\alpha - \hat\beta x_i)^2\right)$$

We can differenciate it with respect to $\sigma_\delta^2$, and find the critical point.

One gets

$$\underbrace{\hat\sigma_\delta^2}_{\text{MLE for } \sigma_\delta^2} = \frac{1}{n} \frac{\lambda}{2(1+\lambda\hat\beta^2)} \sum_i (y_i - \hat\alpha - \hat\beta x_i)^2$$

What about confidence interval for $\beta$? ⤳ very complicated

One way to get an approximate solution: Consider

$$\hat\sigma_\beta^2 := \frac{(1+\lambda\hat\beta^2)(S_{xx}S_{yy} - S_{xy}^2)}{(S_{xx} - \lambda S_{yy})^2 + 4\lambda S_{xy}^2}$$

is a $\underbrace{\text{consistent}}_{\text{Chap VII}}$ estimator for $\sigma_\beta^2$ → when sample size $\to \infty$, $\hat\sigma_\beta^2 \xrightarrow{} \sigma_\beta^2$

in precise sense see Chap VII

Then by the central limit thm, one has

we don't know it

$$\frac{\hat\beta - \beta}{\sigma_\beta/\sqrt{n}} \xrightarrow{n\to\infty} \mathcal{N}(0,1)$$

from which one obtains the approximate $(1-\alpha)$ confidence interval

$$\left[\hat\beta - z_{\alpha/2}\frac{\hat\sigma_\beta}{\sqrt{n}}, \hat\beta + z_{\alpha/2}\frac{\hat\sigma_\beta}{\sqrt{n}}\right]$$

⚠ It is not a $(1-\alpha)$ confidence interval,

but for $n$ large, it converges to a $(1-\alpha)$ confidence interval.

## IX.2 Logistic regression (0,1 model)

$E(Y_i) = \pi_i = P(Y_i = 1)$

Model: $\{Y_i\}$ independent variables with $Y_i \sim \text{Bernoulli}(\pi_i)$ with

$$\pi_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} \in [0,1] \iff \ln \underbrace{\frac{\pi_i}{1-\pi_i}}_{= \frac{\text{prob of success}}{\text{prob of failure}} =: \text{odds}} = \alpha + \beta x_i$$

Remark: We cannot draw a graph $(x_i, y_i)$ and use the least squares

but we can use the MLE.

One has $L(\alpha, \beta \mid \underline{x}, \underline{y}) = \prod_i \underbrace{\pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}}_{\text{pmf for Bernoulli}}$ under the assumption of independence of measurements

with $\pi_i = \pi(x_i)$

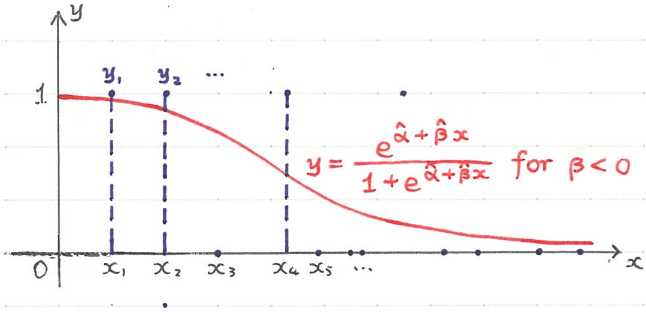
and $\pi$ the function $x \longmapsto \pi(x) := \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$

We can then compute

$$\frac{\partial}{\partial\alpha} L(\alpha, \beta \mid \underline{x}, \underline{y}) = 0 \text{ and } \frac{\partial}{\partial\beta} L(\alpha, \beta \mid \underline{x}, \underline{y}) = 0$$

and solve the system.

We can not solve this system explicitly, but a computer can do it easily.
Suppose we have obtained $\hat{\alpha}$ and $\hat{\beta}$ numerically, then we can plot the result




Remember $E(Y) = \pi(x)$

We can get some confidence intervals.

If we want to consider several measures for a given $x$,
one has to use a binomial distribution.

More precisely, if $n_i$ independent, Bernoulli observations are measured at $x_i$,
then Bernoulli $(\pi_i)$ has to be replaced by binomial $(n_i, \pi_i)$.
$\hookrightarrow = \pi(x_i)$

Then $L(\alpha, \beta | \underline{x}, \underline{y}) = \prod_i \binom{n_i}{y_i} \pi(x_i)^{y_i} (1-\pi(x_i))^{n_i - y_i}$

with $y_i$ the number of success at $x_i$,
and we can compute the MLE for $\alpha$ and for $\beta$ (with a computer)

Application : see (3.2) of Appendix 13

## Conclusion for the course

We have only touched the surface of statistics, but we have opened many doors, and you can continue in these directions.