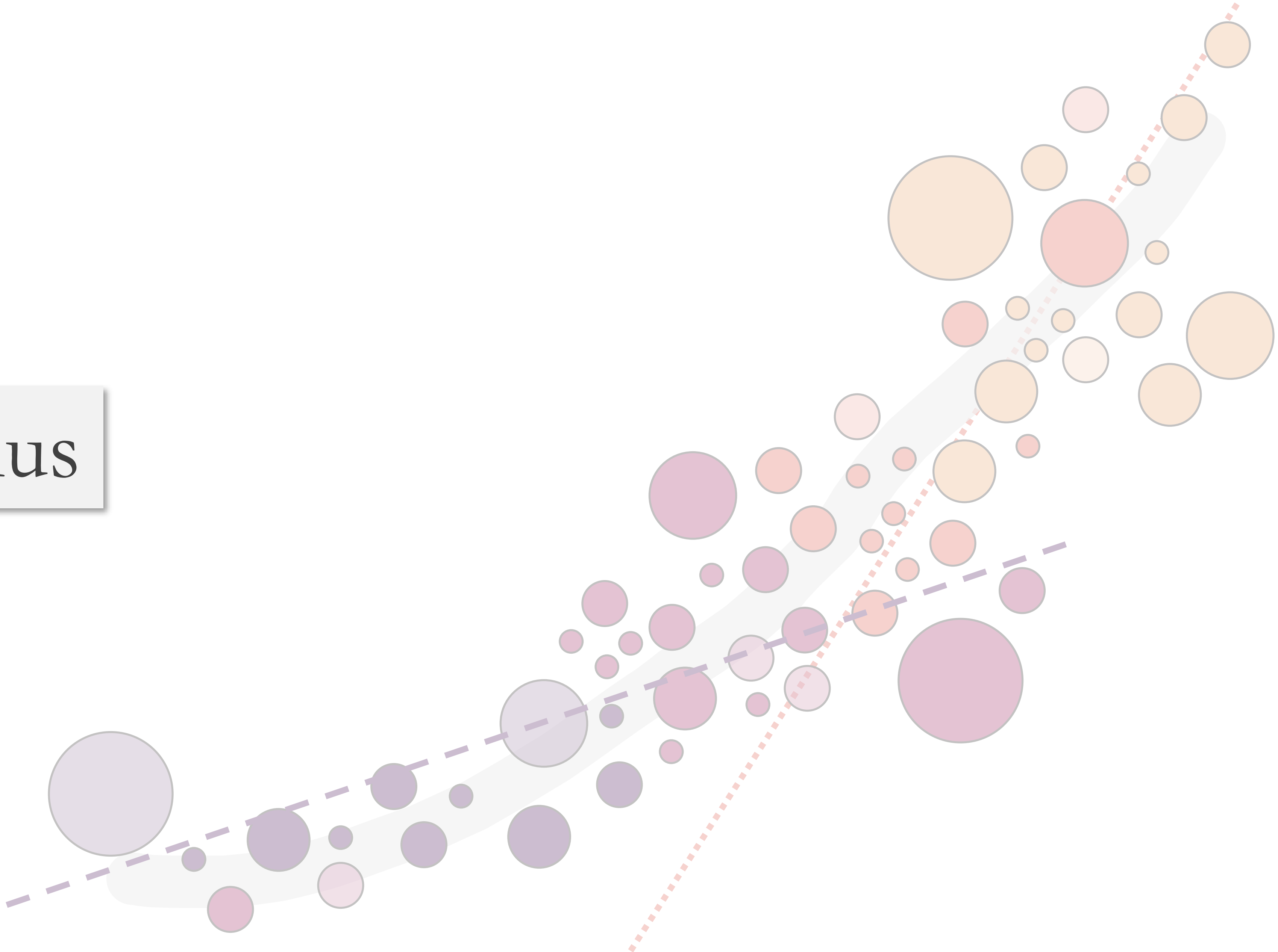
More calculus

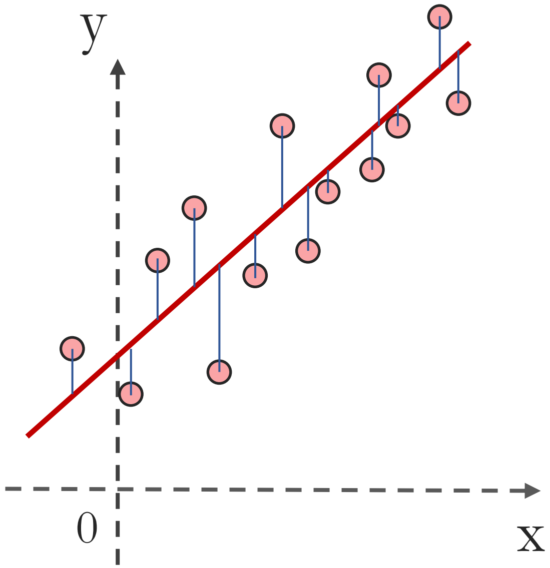


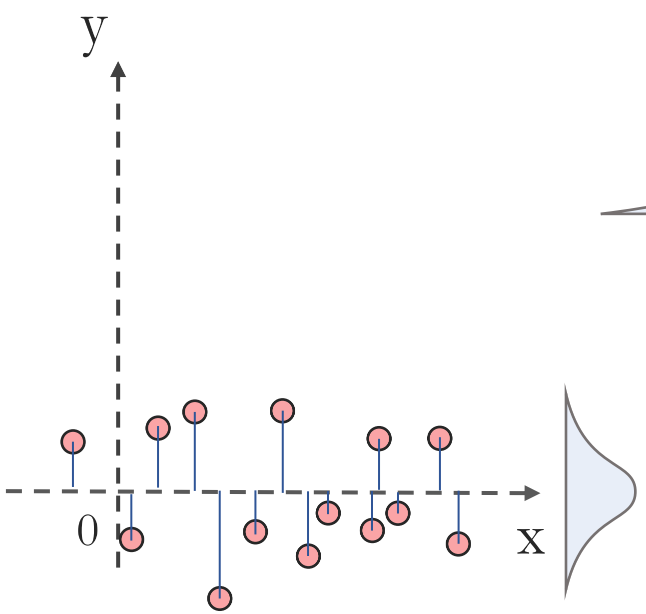
Nicolas Legrand – Researcher– ILAB (Interacting Minds Centre)

AARHUS UNIVERSITY

Maximum Likelihood Estimation

$$X\vec{\beta} = \vec{y} + \vec{\epsilon} \qquad \epsilon \sim N(0, \sigma^2)$$



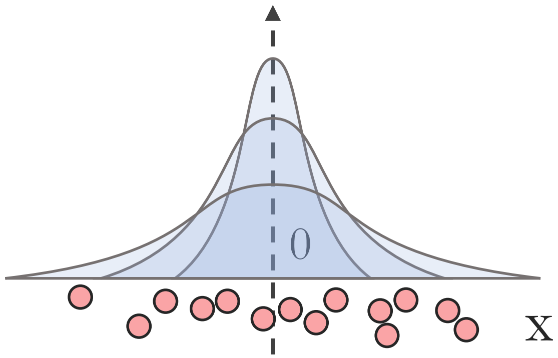


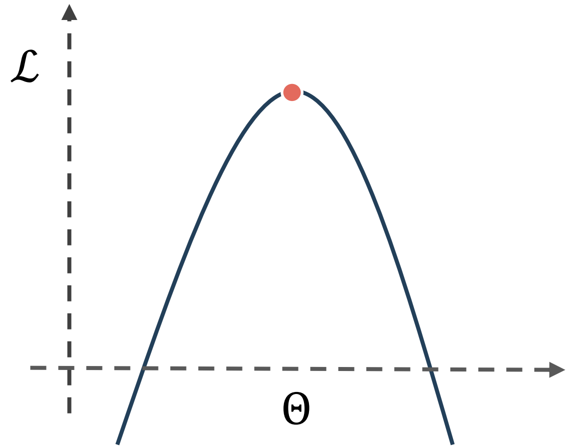
Considering the simple linear model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \epsilon \sim N(0, \sigma^2)$$

and using the probability density function of the Normal distribution, write the likelihood function $\mathcal{L}(\beta_0, \beta_1, \sigma^2)$.

Maximum likelihood estimation (**MLE**) is a method of estimating the parameters of an assumed probability distribution, given some observed data.





This is achieved by maximizing a **likelihood function** (**$\mathcal{L}$**) so that, under the assumed statistical model, the observed data is most probable.

When random variables are statistically independent, if their **joint probability** can be expressed in terms of their **marginal probabilities**.

$$\underbrace{P(X_1 = x_1, \ldots, X_n = x_n)}_{\text{Joint probability}} = \underbrace{\prod_{i=1}^{n} p(X_i = x_i)}_{\text{Marginal probability}}$$

$$
\begin{aligned}
\mathcal{L}(y|\beta_0, \beta_1, \sigma^2) &= p(y|\beta_0, \beta_1, \sigma^2) \\
&= \prod_{i=1}^{n} p(y_i \mid \beta_0, \beta_1, \sigma^2) \\
&= \prod_{i=1}^{n} N(y_i \mid \beta_0, \beta_1, \sigma^2) \\
&= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2}} \\
&= \frac{1}{\sigma\sqrt{2\pi}^{n}} \cdot exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2\right]
\end{aligned}
$$

Write the log-likelihood function.

$$\log(\mathcal{L}(y|\beta_0,\beta_1,\sigma^2)) = \log(p(y|\beta_0,\beta_1,\sigma^2))$$
$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

Exponentiation

$$\log_b(b^x) = b^{\log_b(x)} = x$$

Division

$$\log\left(\frac{x}{y}\right) = \log(x) - \log(y)$$

Square root

$$\sqrt{x} = x^{\frac{1}{2}}$$

Write the partial derivative of the log-likelihood function with respect to β_1.
Using the example from last week, we assume that $\beta_0 = 18$.

Assuming that:

$$f = \log(\mathcal{L}(y|\beta_0, \beta_1, \sigma^2))$$

$$\frac{\partial f}{\partial \beta_1} = -\frac{1}{2\sigma^2} \cdot \frac{d}{d\beta_1}\left[\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2\right]$$

$$= -\frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n} -2x_i(y_i - \beta_0 - \beta_1 x_i)$$

Power rule

$$= \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} \left(x_i y_i - x_i \beta_0 - \beta_1 x_i^2\right)$$

Using this partial derivative, find the value of β_1 that maximize the log likelihood function.

$$\frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} (x_i y_i - x_i \beta_0 - \beta_1 x_i^2) = 0$$

$$\sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i = \beta_1 \sum_{i=1}^{n} x_i^2$$

$$\frac{\sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i}{\sum_i x_i^2} = \beta_1$$