Bayesian inference - Case studies in ecotoxicology for practical exercices

Sandrine CHARLES
Université de Lyon; Université Lyon 1
Laboratory of Biometry and Evolutionnary Biology
CNRS; UMR 5558
43 boulevard du 11 novembre 1918
F-69622 Villeurbanne Cedex, FRANCE

sandrine.charles@univ-lyon1.fr

May 12, 2013

Definition of bioassays

- Bioassay is a word commonly used instead of **biological assay**. It's a type of scientific experiment.
- Bioassays are typically conducted to measure the effects of potentially toxic substances on living organisms.
- Bioassays can be
  - **Qualitative**: to assess physical effects of a substance that cannot be quantified (*e.g.*, abnormality or deformity)
  - **Quantitative**: to estimate the potency of a substance by measurement of the biological response/effect it produces; quantitative bioassays are typically analyzed using **statistical methods**.

Quantitative bioassays

Several kind of substances can be studied, for example:

- Pesticides, chemical, pharmaceutical, cosmetic substances
- Effluents (industrial discharges, outputs from water plants)
- Polluted soils, waste, sewage sludge, sediments

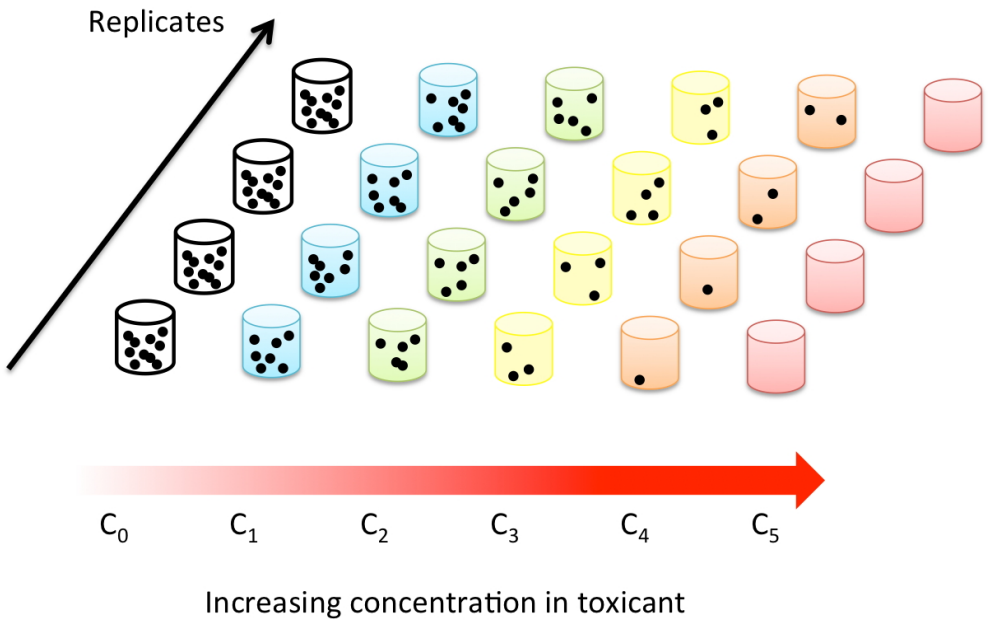
According to the substance, different kinds of experiments can be conducted:

- A control versus one treatment,
- A control and several treatments.

A treatment can be a fixed concentration of a substance (C_1, C_2...) or a time-variable concentration $C(t)$.

Classical scheme of bioassays

Under **standard protocols**, after time exposure d, individuals are counted in each flask.



Observed responses or effects can be survival, growth, reproduction for example.

Who define standard protocols?

Organisation for Economic Cooperation and Development

Free of charges

International Organization for Standardization

Not free

Other institutions:

- **US EPA** (US Environmental Protection Agency)
- **ASTM** (American Society for Testing and Materials)
- **DIN** (Deutsches Institut für Normung)
- **MITI** (Ministry of International Trade and Industry in Japan)

Acute vs. chronic bioassays

- Acute toxicity: from some hours to some days (*e.g.*, survival or mobility inhibition)
  $\rightarrow$ short-time exposure at high concentrations
  $\rightarrow$ rapid impact on organisms
- Chronic toxicity: from some days to some weeks (*e.g.*, growth or reproduction inhibition, sub-individual biomarkers)
  $\rightarrow$ long-time exposure at low concentrations

Example of a bioassay

Daphnia sp., Acute Immobilisation Test (OECD 202, 1984) and Chronic Reproduction Test (OECD 211, 2012)

Daphnia magna

Acute test: the number of immobile daphnids is determined for each concentration at 24 and 48 hours.

Chronic test: offsprings are daily counted during 21 days.

Critical Effect Concentrations

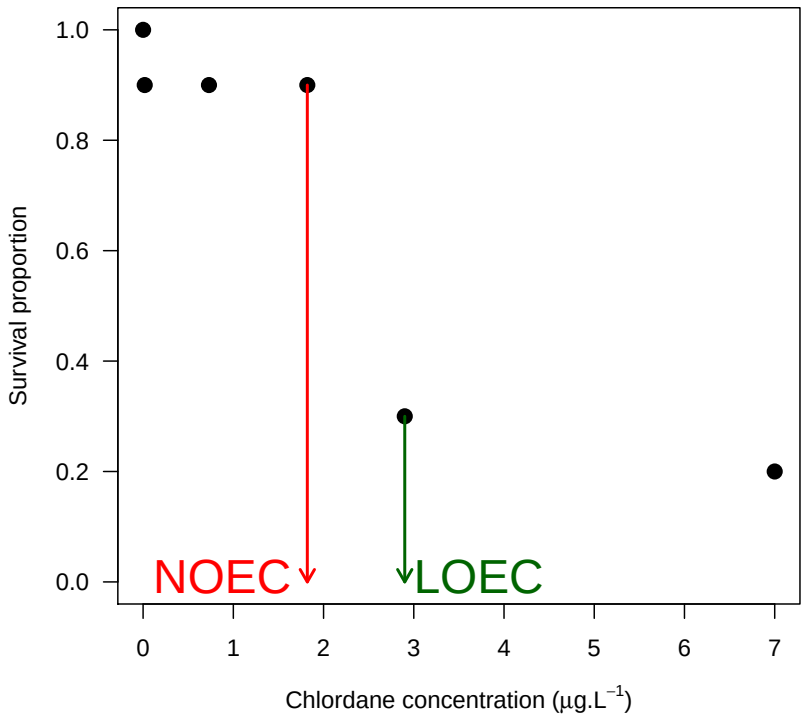
... or "summary Statistics" or "thresholds"

$\rightarrow$ Most common indicators to quantitatively assess risks for single species exposed to contaminants

$\rightarrow$ Estimation of the exposure level (*e.g.*, concentration) above which adverse effects can occur on organisms, and below which adverse effects are unlikely, *i.e.*, which can not be distinguished from background noise (OECD 2006).

OECD (2006) *Current approaches in the statistical analysis of ecotoxicity data: a guidance to application*. Technical report.

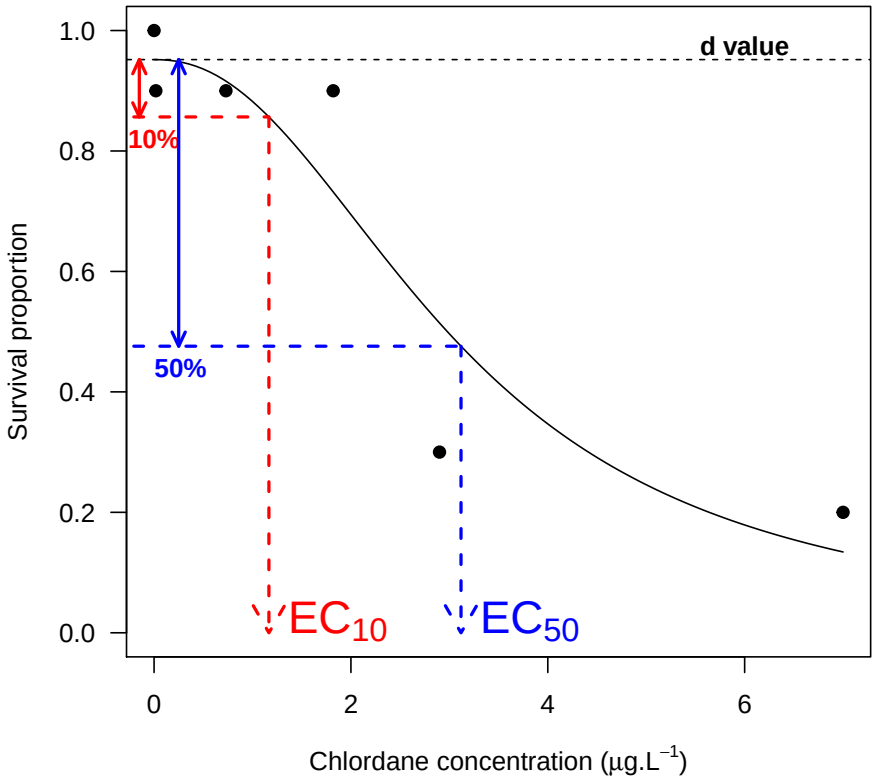
No (Lowest) Observed Effect Concentration



- NOEC*: the highest concentration **among the tested concentrations** at which no statistically significant difference from the control can be highlighted.
- Good reasons to avoid such a concept: dependance on the experimental design, an improper use of *p*-values

* The NOEC has been here estimated by a Cochran-Armitage's test [Forfait-Dubuc *et al.*, 2012]

x% Effective Concentration

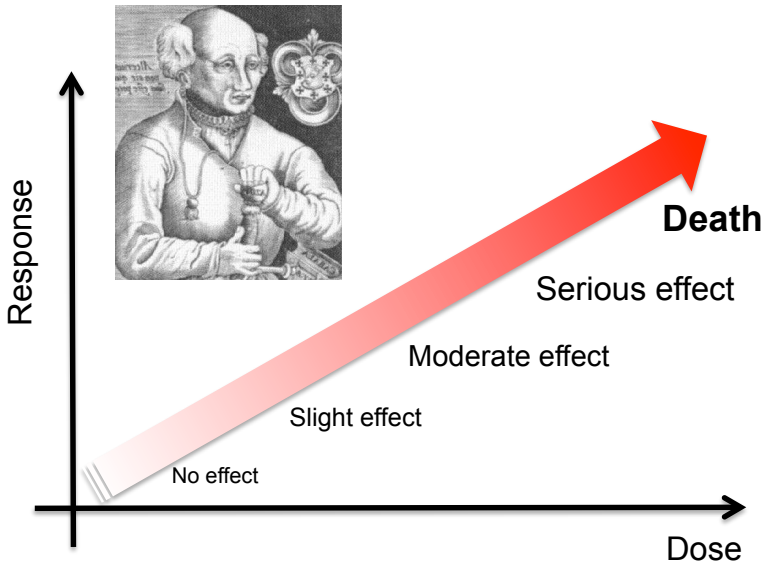


- EC_x: the **estimated** concentration that causes x% change in response.
- The plain line corresponds to a three-parameter log-logistic model fitted to the data (*see further*)

When the effect is mortality, LC_x is the abbreviation used.

Does the dose make the poison?

Generally **yes**, for mono-substance exposure.



"All substances are poisons: there is none which is not a poison. The right dose differentiates a poison and a remedy."

Paracelsus (1493-1541)

Endocrine disruptors are counter-examples (hormesis phenomenon).

Dose or Concentration-response or effect relationships?

Review of vocabulary:

- **Dose** refers to exposure, *i.e.*, the amount of toxicant actually deposited **within** the body of organisms. But in ecotoxicology, only the **exposure concentration** is know:
  $\rightarrow$ We rather speak about **concentration**-response or effect relationships.
- Concentration-**response** relationships refer to the link between the exposure concentration and the proportion of individuals responding with an all-or-none effect.
- Concentration-**effect** relationships refer to the link between the exposure concentration and the magnitude of the induced biological change, measured in appropriate units.

Definition

A **concentration-response/effect relationship** is a simple X-Y graph relating increasing levels of exposure (X) to the response/effect (Y) at a certain exposure time.

Examples of response:

- Quantal data, expressed as proportion or probability (*e.g.*, mortality or immobilization)

Examples of effects:

- Ordered descriptive categories (*e.g.*, severity of a lesion)
- Counts (*e.g.*, reproduction products like eggs or clutches)
- Continuous measurements (*e.g.*, body size)

Modelling : general considerations

From concentration-response/effect experiment, if there is a reasonable number of concentrations (> 5) of the toxicant and a reasonably well-behaved response/effect, it is straightforward to **fit a regression model**.

A regression model relating a dependent variable Y (the response or the effect) to an explanatory variable X (the concentration) is composed of two parts:

1. a **deterministic part**, which describes the mean value (or curve), *e.g.*, a log-logistic model;
2. a **stochastic part**, which represents the distribution around the mean curve (*e.g.*, a normal distribution).

Nevertheless, each part depends on the nature of data to analyze.

The deterministic part

The log-logistic model is the most commonly concentration-response/effect model used in toxicology [Ritz, 2010].

$$Y = c + \frac{d - c}{1 + exp(b \times \log(\frac{X}{e}))}$$

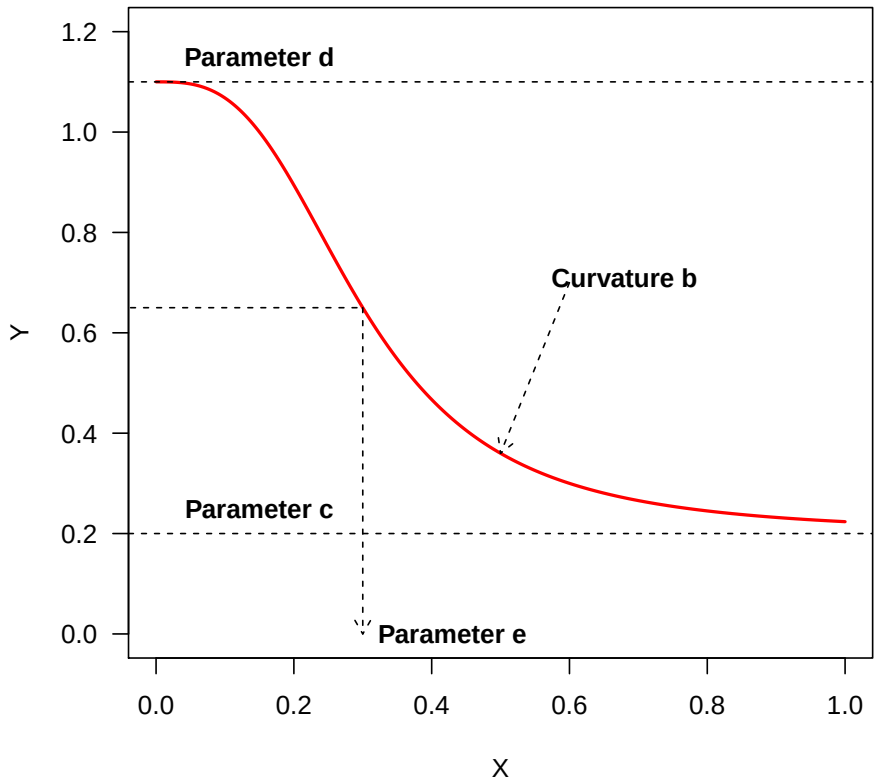
$$\Updownarrow$$

$$Y = c + \frac{d - c}{1 + (\frac{X}{e})^b}$$

b, c, d, e are positive, all with a geometric meaning.

The log-logistic model - Graph

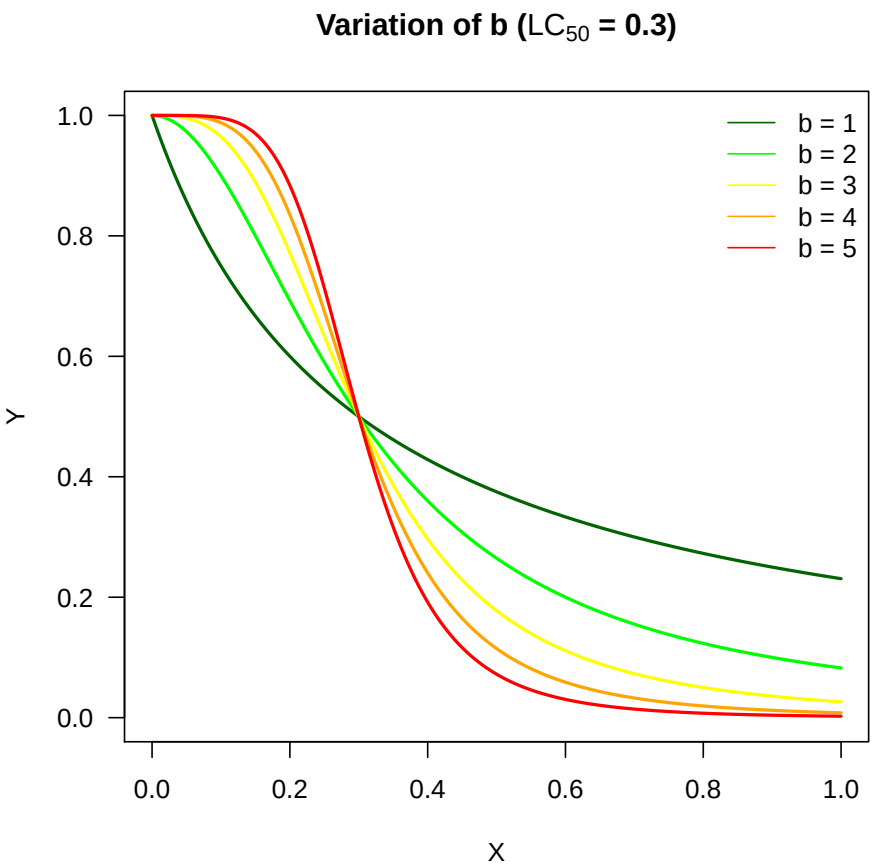
In case of survival, d corresponds to the natural mortality (usually fixed to 1) and c is fixed to 0. $b = 3, c = 0.2, d = 1.1$ and $e = 0.38$.

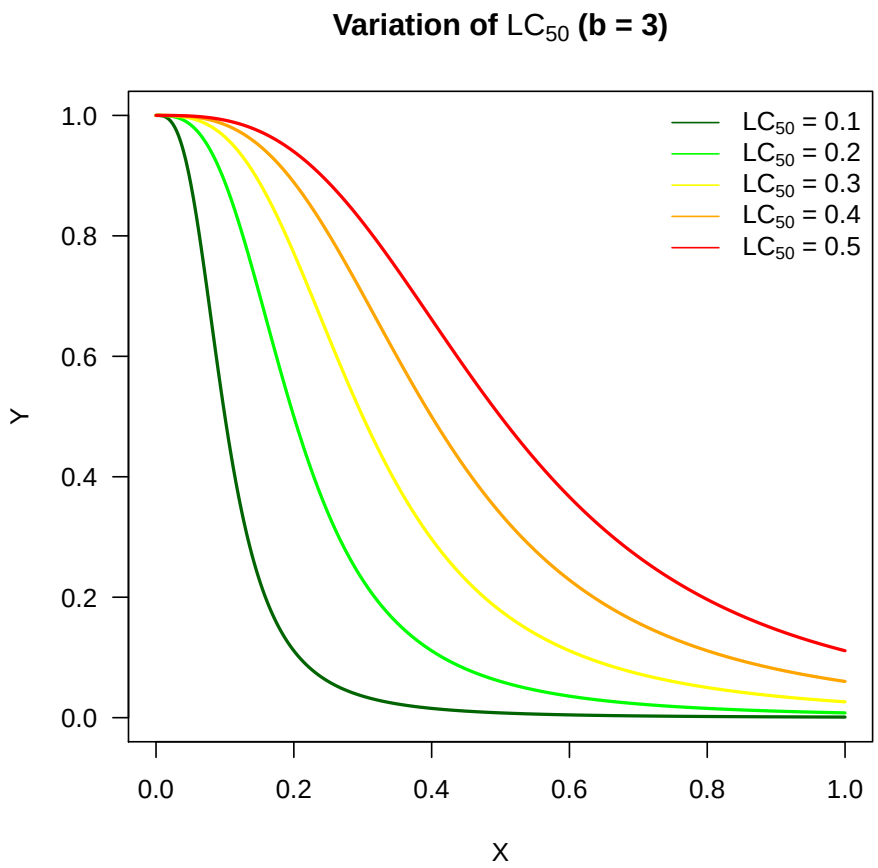


$e = 0.3$ (arbitrary unit)

The log-logistic model - Morphology

Case of survival, with $c = 0$ and $d = 1$: $e = LC_{50}$.





Another deterministic part: the Pires-Fox model

This is a **threshold** model written as follows:

$$Y = \alpha exp(-\beta(X - NEC)I(X - NEC))$$

$$I(X - NEC) = \begin{cases} 1 & \text{if} \quad X > NEC \\ 0 & \text{if} \quad X \leq NEC \end{cases}$$

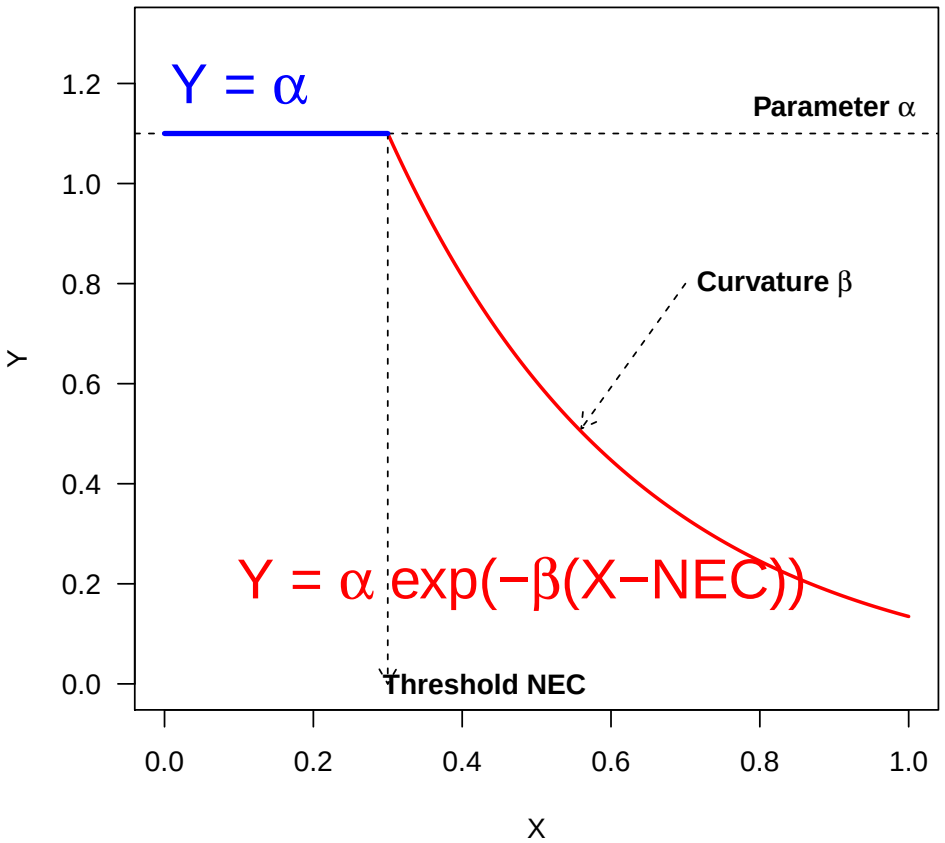
Parameter *NEC* is the **No Effect Concentration**.

[Fox D. (2010) A Bayesian approach for determining the no effect concentration and hazardous concentration in ecotoxicology. *Ecotoxicology and Environmental Safety*, 73(2):123-31.]

The Pires-Fox model - Graph

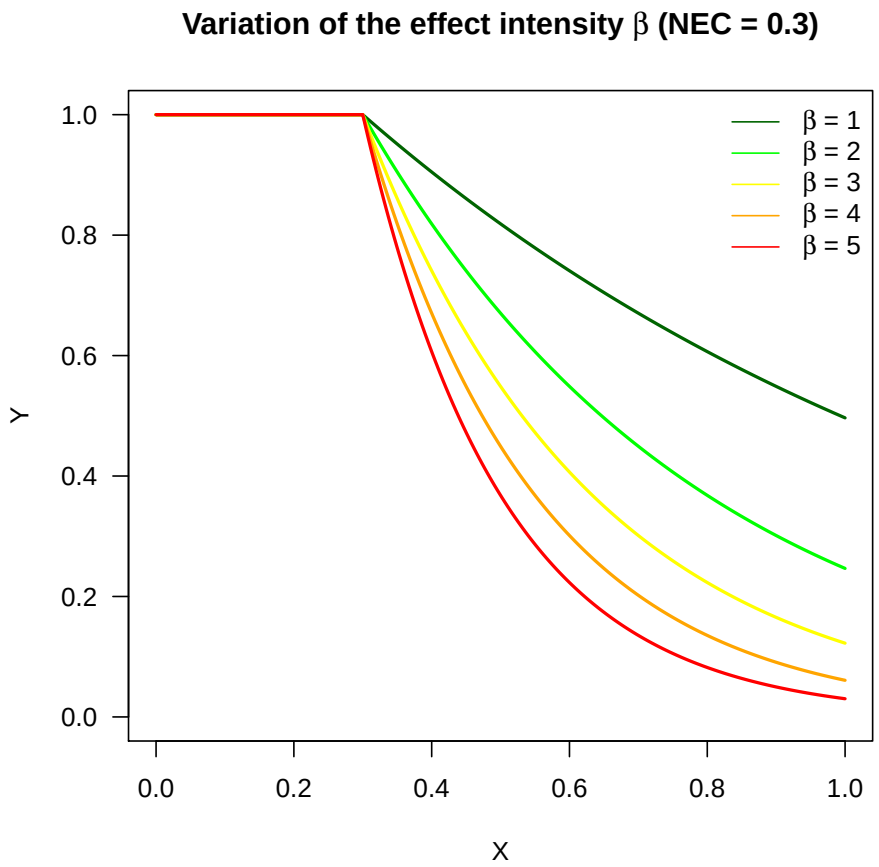
In case of survival, parameter α is close but less than 1.

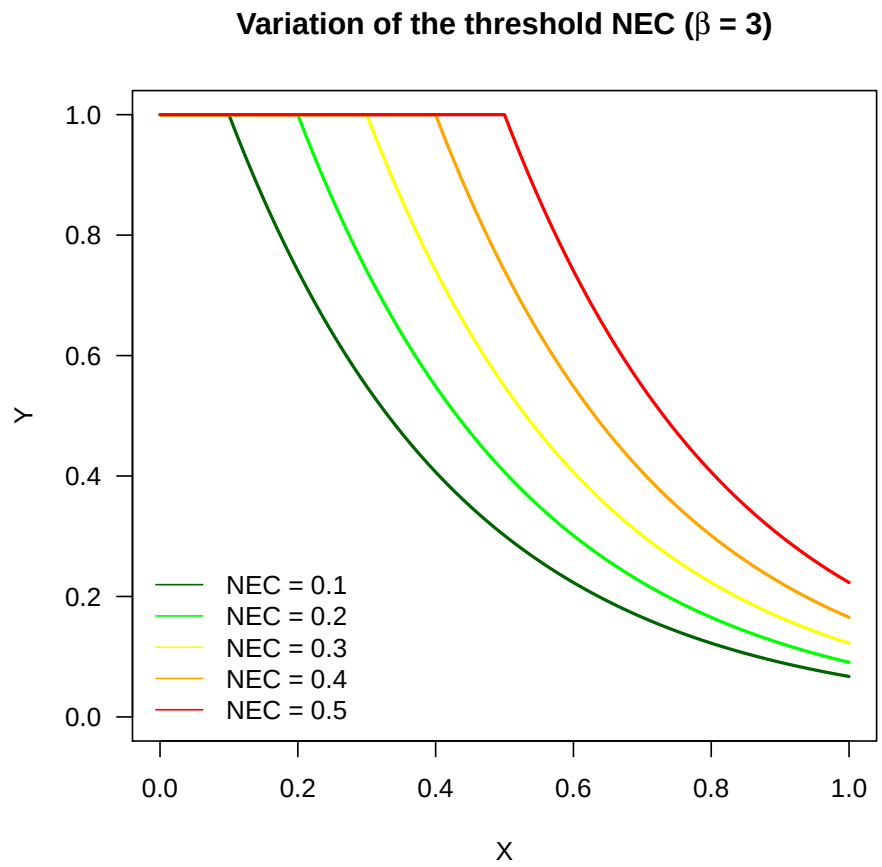
$\alpha = 1.1$, $\beta = 3$ and $NEC = 0.3$



The Pires-Fox model - Morphology

$\alpha = 1$





Quantal data - description

Quantal (or binary) data arise when a particular property is recorded to be present or absent in each individual (*e.g.*, an individual shows an effect or it does not show an effect).

Therefore, these data can exhibit **only two states**.

Typically, quantal data are presented as **the number of individuals** showing the property (*e.g.*, mortality) out of a total number of individuals observed in each experimental unit. Although this can be expressed as a fraction, **the total number of individuals cannot generally be omitted**.

Quantal data - Binomial stochastic part

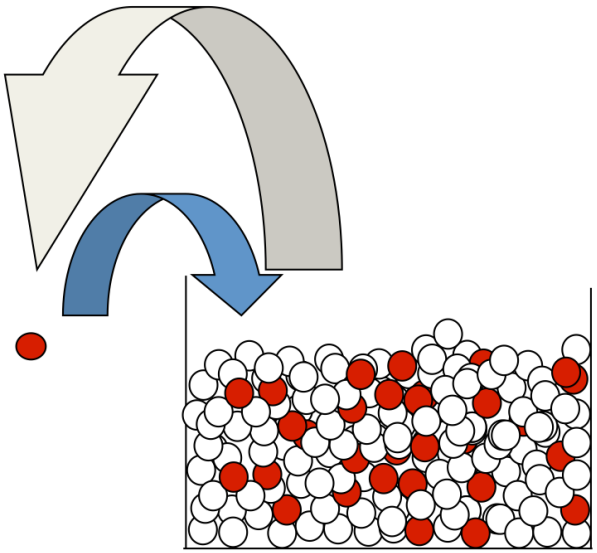
With **quantal** experimental data (*e.g.*, survival data), the stochastic part of the model is necessarily **binomial**.

Observations are then described by a model of the following form:

$$Y \sim \mathcal{B}(p, n)$$

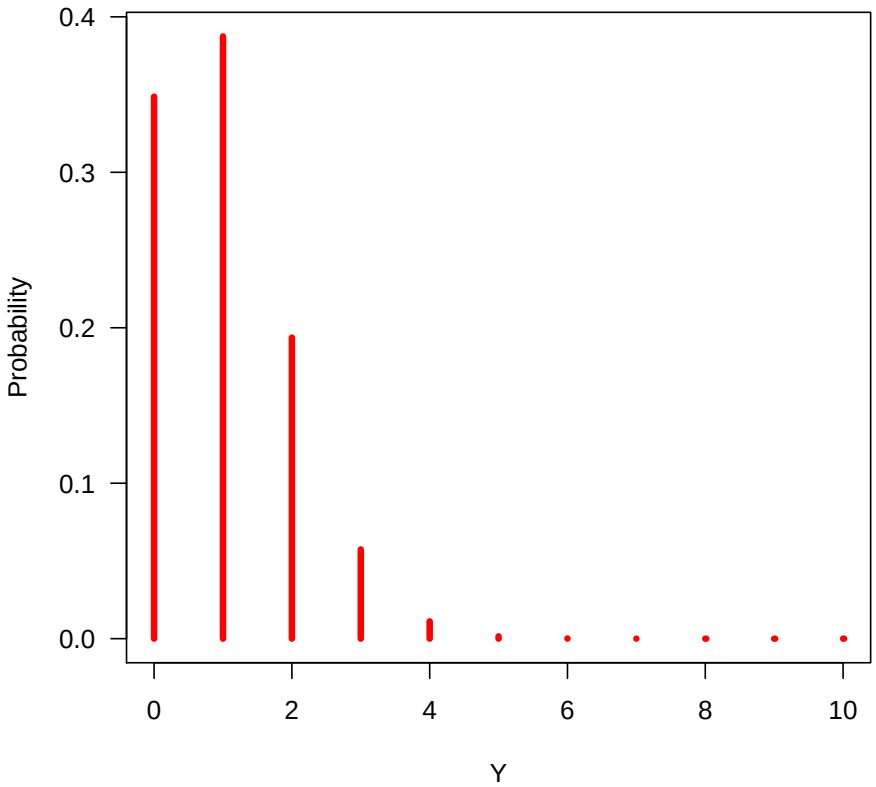
where p is the **probability of success** (*e.g.*, survival probability) as described by one of the dose-response models (log-logistic or Pires-Fox, see above), and n is the total number of individuals.

Principle of the binomial process



- A **binomial** statistical experiment consists of n repeated trials, each trial resulting in just two possible outcomes: success (of probability p) or failure.
- Trials are independant.
- Let Y be the number of successes resulting from a binomial experiment, then $Y \sim \mathcal{B}(p, n)$.
- The probability distribution of Y is a **binomial distribution** of mean $p \times n$ and variance $p \times (1 - p) \times n$.

Binomial probability distribution - Graph

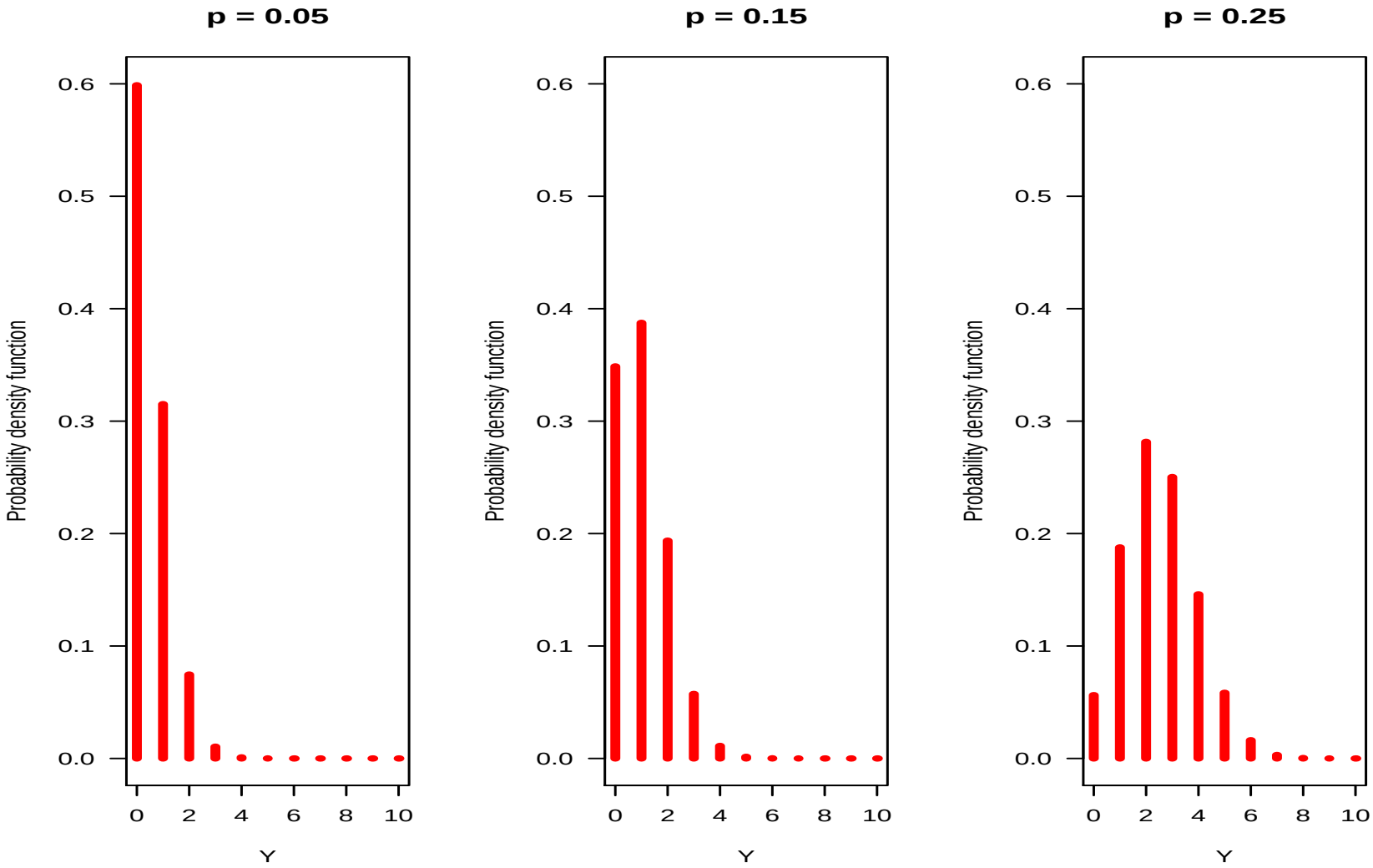


$p = 0.1$ and $n = 10$

- $P(Y = 1)$ is the highest
- $P(Y = 8)$ and $P(Y = 9)$ are very low

Binomial probability distribution - Morphology

$n = 10$



Discrete data - description

Data are **discrete** if there are only a finite number of values possible or if there is a space on the number line between each two possible values.

Discrete data are usually obtained when we are **counting something** (using whole numbers).

They are also called **count data**.

Typical discrete data are **reproduction data**.

Discrete data - Poisson stochastic part

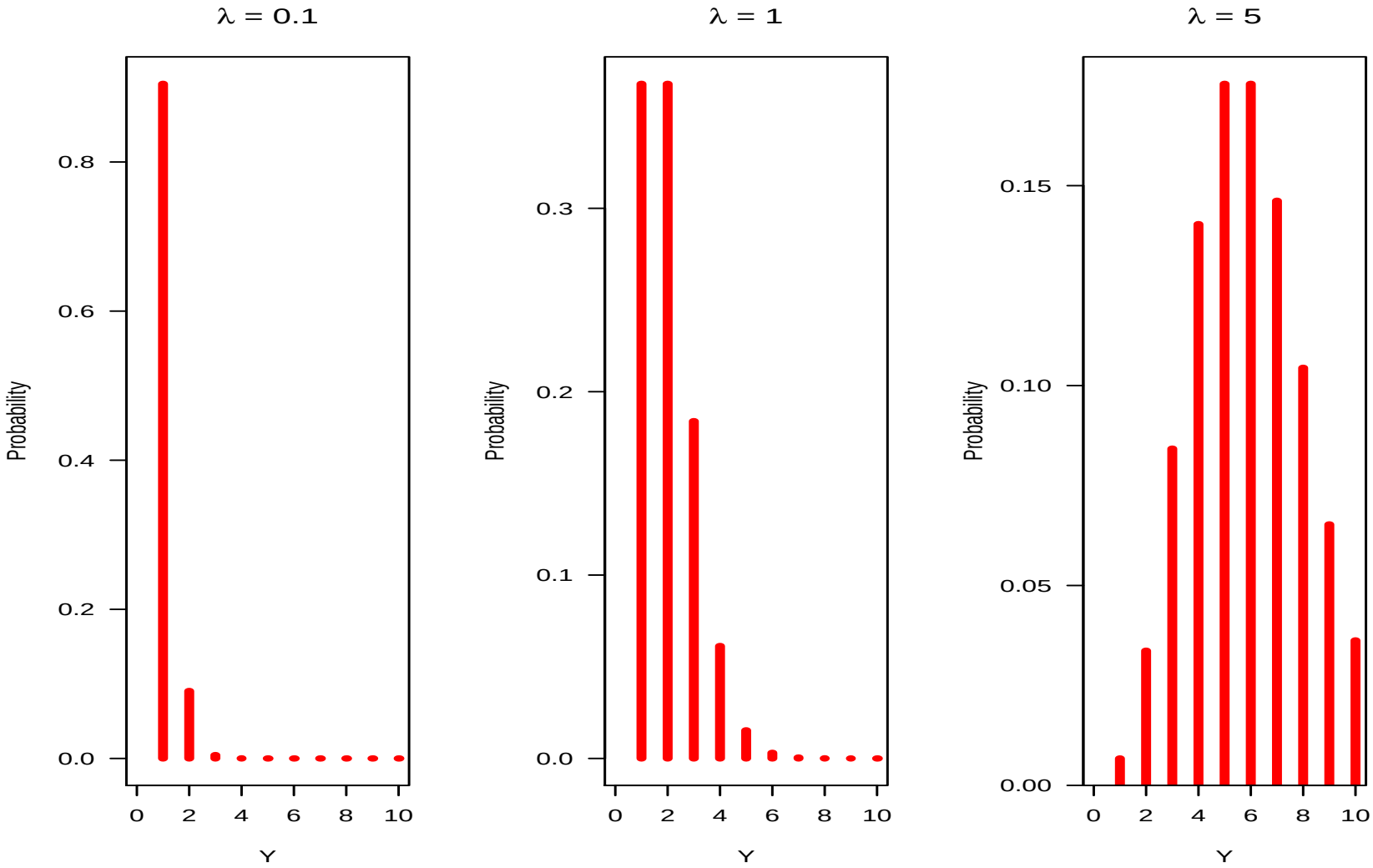
Count data are usually modelled using a **Poisson distribution**. If N is the number of reproduction outputs (*e.g.*, eggs or clutches) at concentration X, then:

$$N \sim \mathcal{P}(\lambda)$$

with $\lambda = f(X)$ the mean of the Poisson distribution, that is the predicted mean value from the log-logistic or the Pires-Fox model (see above).

Poisson probability distribution - Vizualization

λ denotes both the mean and the variance of the distribution.



Continuous data - description

Data are **continuous** when they can (theoretically) take any value in an open interval.

Examples include measurements of length, body weight, etc.

Due to practical reasons the measured resolution depends on the quality of the measurement device.

Typically, **continuous data have a dimension** (*e.g.*, g, cm, $g.L^{-1}$).

Continuous data - Gaussian stochastic part

With **continuous** experimental data, the currently encountered stochastic part is **normal** (or **Gaussian**), even if other stochastic parts may be sometimes more appropriate.

Observations are then described by a model of the following form:

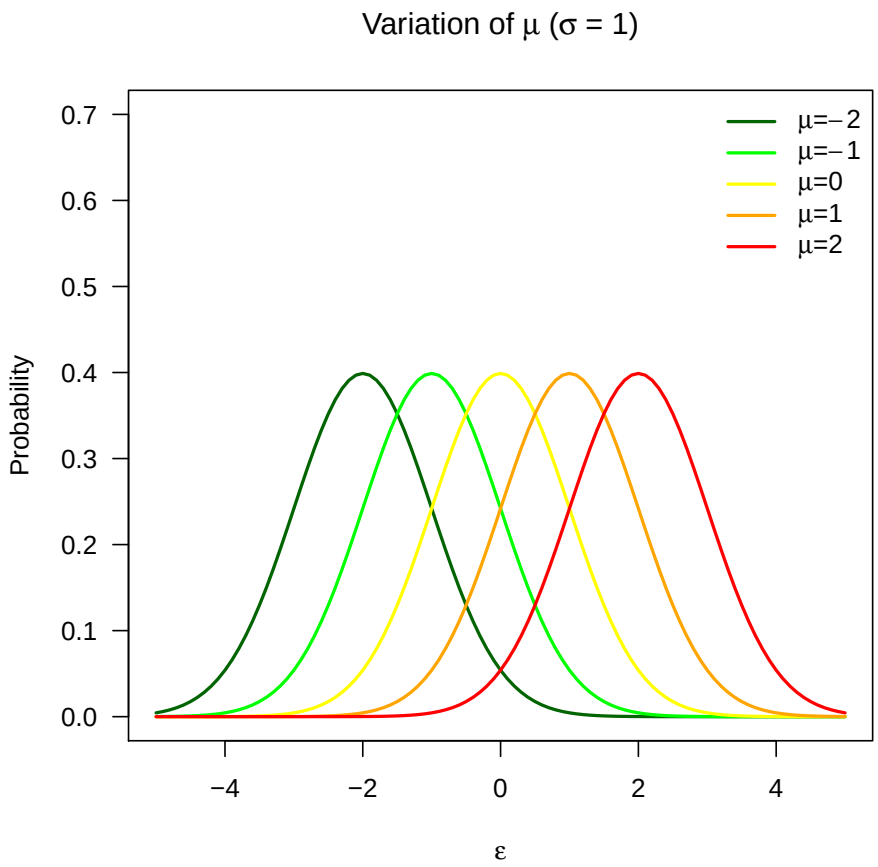
$$Y \sim \mathcal{N}(f(X, \theta), \sigma)$$

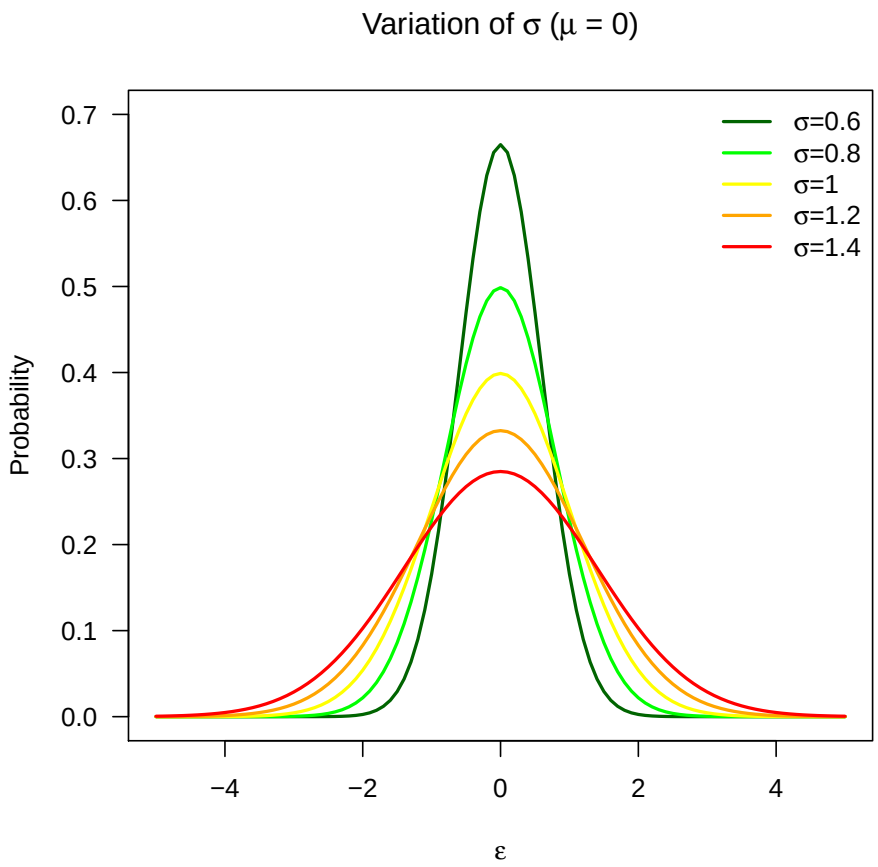
with X the independant variables (*e.g.*, the concentration), $f(X, \theta)$ the log-logistic or the Pires-Fox model and θ the vector of model parameters. This expression can equivalently be written as follows:

$$Y = f(X, \theta) + \varepsilon \quad \text{with} \quad \varepsilon \sim \mathcal{N}(0, \sigma)$$

The normal distribution is a **bell-shaped** probability density function with two parameters: mean μ and variance σ^2.

The bell-shaped Gaussian probability distribution





Illustrative data for chronic toxicity of chlordan

- **Chlordan** is an organochlorine insecticide, produced for agricultural and residential uses and for termite control.
- It is stable, highly lipophilic and persistent, with a biological half-life of several years in soils and sediments.
- It was forbidden for phytosanitary treatments in Europe in 70's, banned in US in 1978, but is still produced in China and Botswana.
- Approximately 20% of the 70,000 tons of chlordan manufactured since 1946 still exist **unaltered** in the environment.

$\rightarrow$ chlordan toxicity is still a relevant question due to its widespread occurrence and environmental persistence.

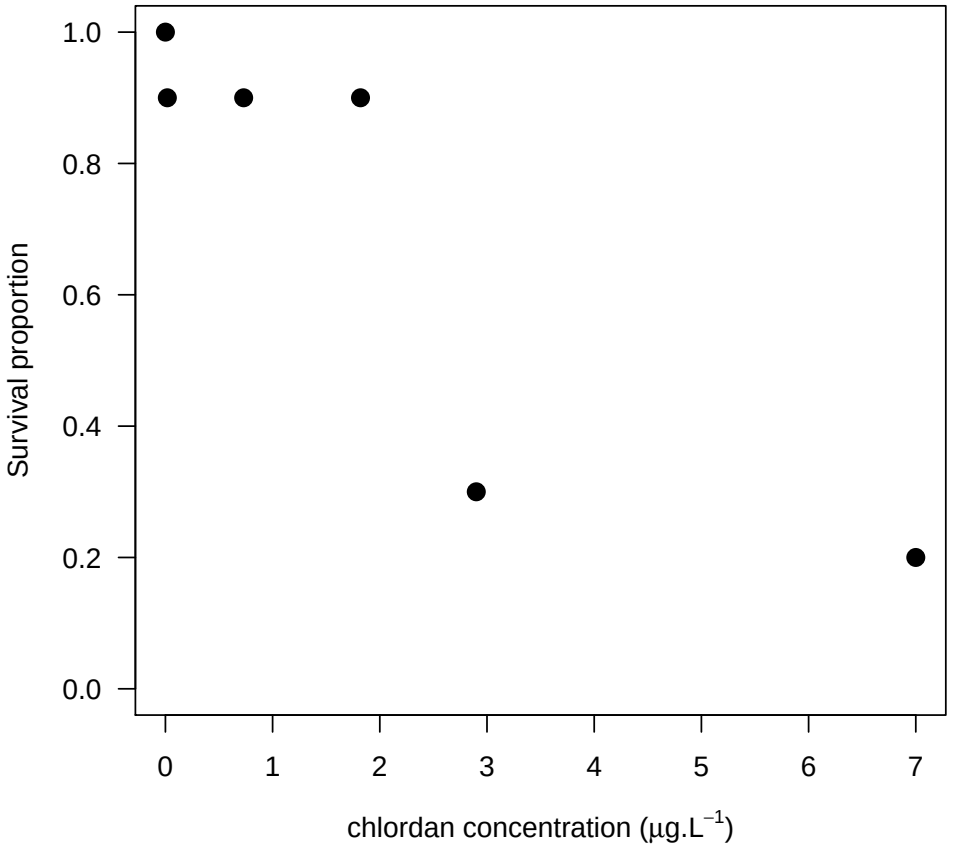
Biological model: *Daphnia magna*

- *Daphnia magna* is a cladoceran, representative of freshwater crustacean species and zooplankton.
- Daphnids are the most significant herbivores among invertebrates
- They are considered as an important source of food for fish
- The *Daphnia* model is recommended by standard methods for aquatic ecotoxicity assessment;
- It is required by many international regulations because of its sensitivity to chemical environmental stressors;
- *Daphnia* holds a central position in aquatic food webs and is an intermediate between primary producers and fish.

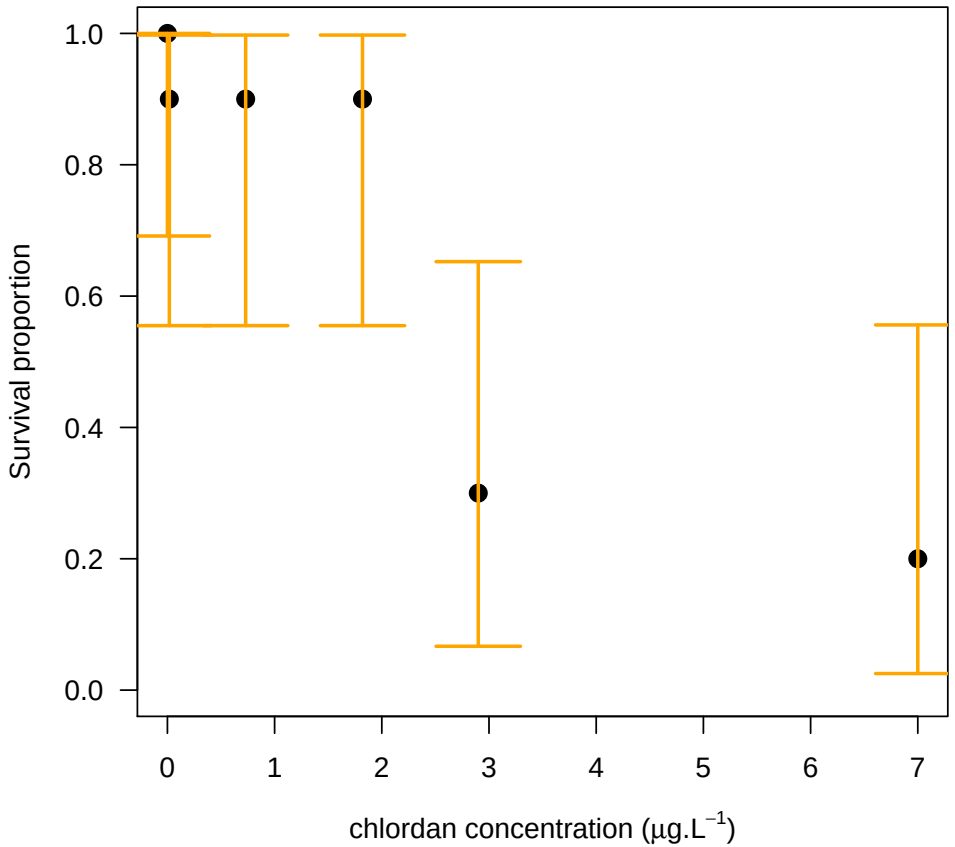
Chronic toxicity test (OECD guideline 211)

- Daphnids ($<$ 24 h old) were exposed for 21 d to **measured** concentrations of chlordan (mean $\pm$ sd) of: 0, 0.18 $\pm$ 0.05, 0.73 $\pm$ 0.15, 1.82 $\pm$ 0.16, 2.9 $\pm$ 0.5 and 7.0 $\pm$ 3.5 $\mu g.L^{-1}$.
- Daphnids were raised individually in 50-ml glass beakers containing 40 ml of test solution
- A total of **10 replicates** for each treatment was performed.
- The incubation temperature was controlled at 20 $\pm$ 1°C
- A 16:8 h light:dark photoperiod was maintained.
- The test solution was renewed every 2 d.

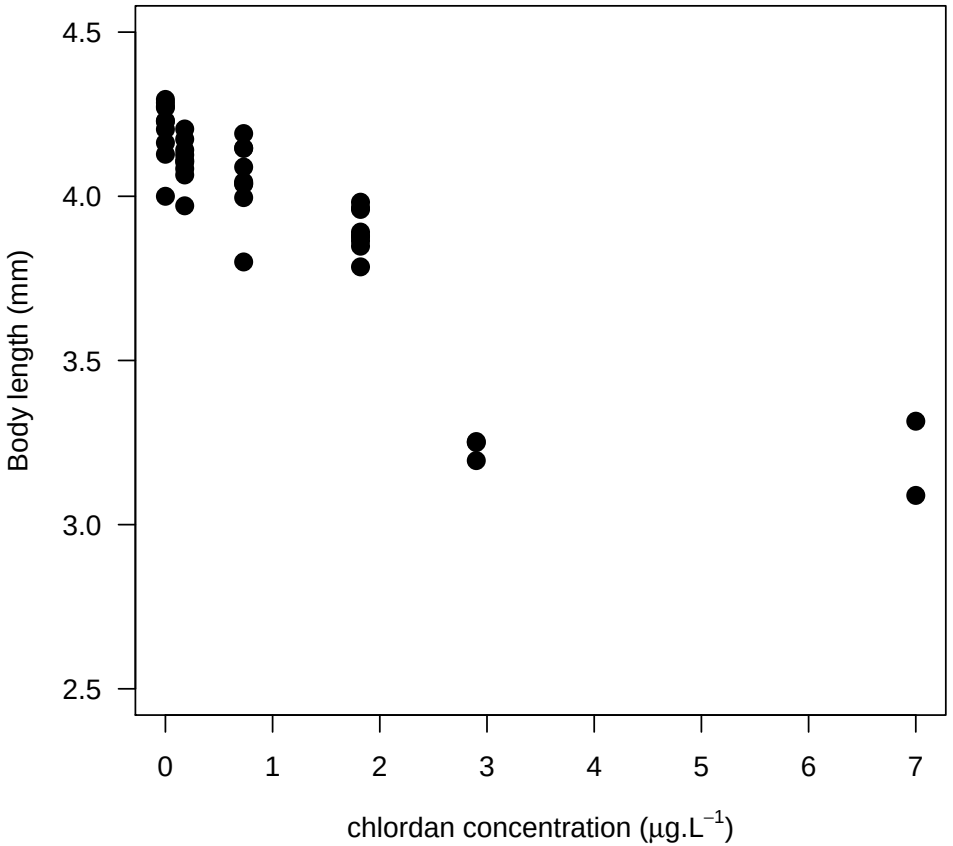
Survival data at day 21



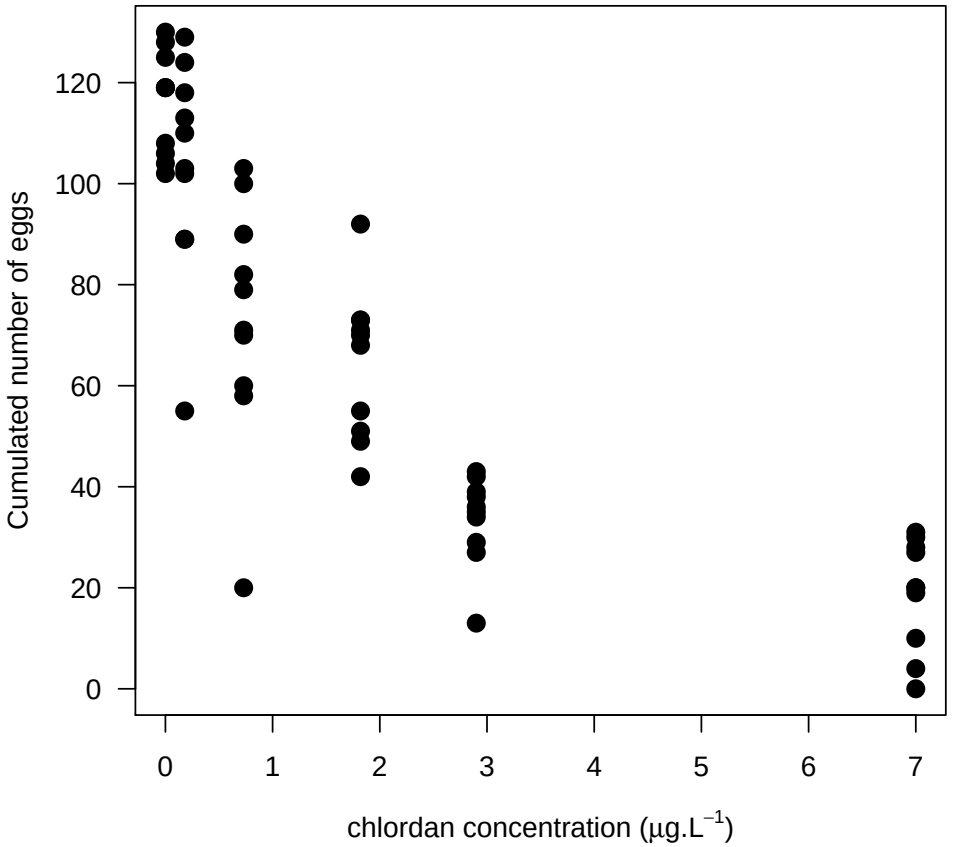
Survival data at day 21 → associated with their binomial 95% confidence interval



Growth data at day 21: body length (in mm)



Reproduction data at day 21: cumulated number of eggs



Conclusion: towards practical exercices

- *D. magna* data (**survival** and **growth**) will be used in the first part of practical exercises.
- In the second part of practical exercises, we will study **reproduction** of snails exposed to some toxicant.
- In the last part of practical exercices, we will analyze two **survival** datasets:
  1. *D. magna* exposed to cadmium,
  2. Snails exposed to some toxicant.