Three data types: continuous, counts and coin flips

Cóilín Minto, Olga Lyashevska

Marine and Freshwater Research Centre
Atlantic Technological University
Galway, Ireland

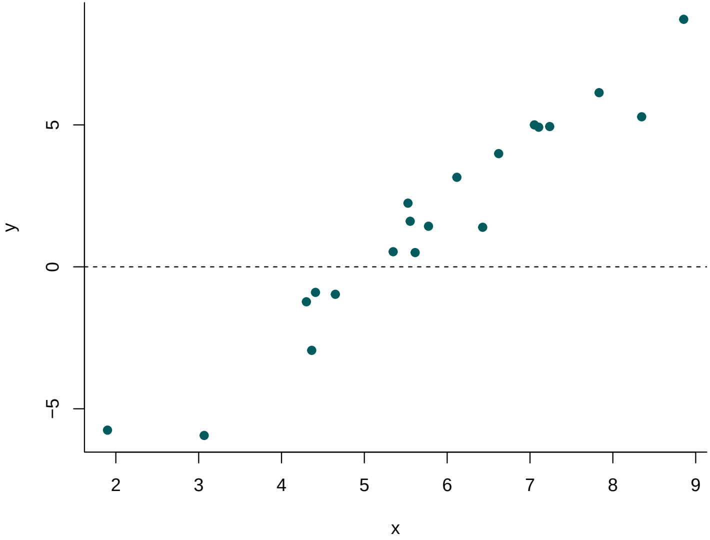
July 15th 2022

Ollscoil
Teicneolaíochta
an Atlantaigh

Atlantic
Technological
University

Outline

Describe some features of the response data y

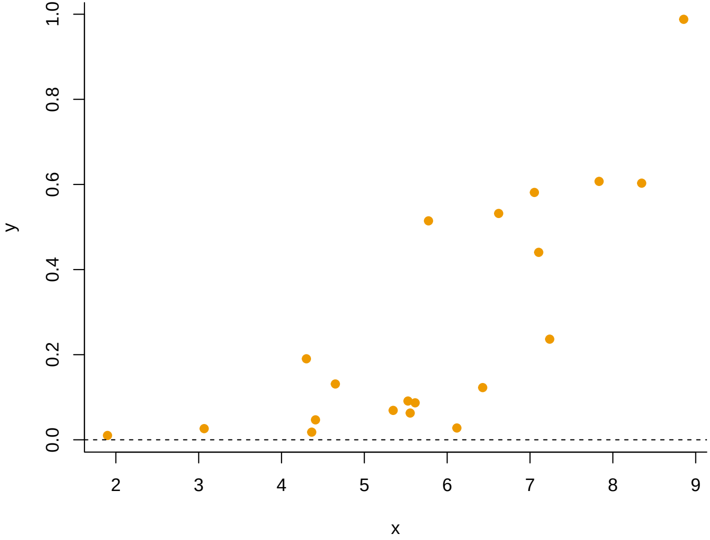


Describe some features of the response data y

Continuous data

- Response y is continuous, e.g., $y = 1.25$ possible
- Response can be positive or negative (on the real line)
- Apparent positive linear relationship with continuous variable x
- **Example** y could be a change in water height

Describe some features of the response data y

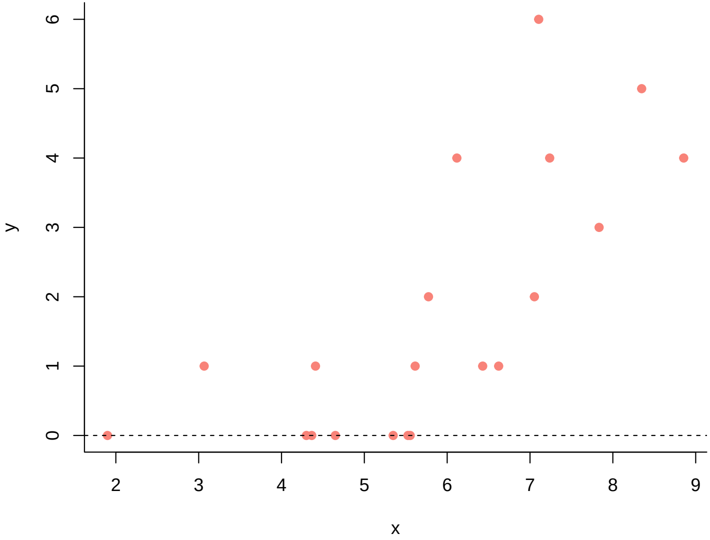


Describe some features of the response data y

Positive continuous data

- Response y is also continuous, e.g., $y = 0.25$ possible
- Response can only be positive (on the positive real line)
- Apparent positive non-linear relationship with continuous variable x
- **Example** y could be mass of individuals
  - Discuss what values mass/weight of a fish could be

Describe some features of the response data y

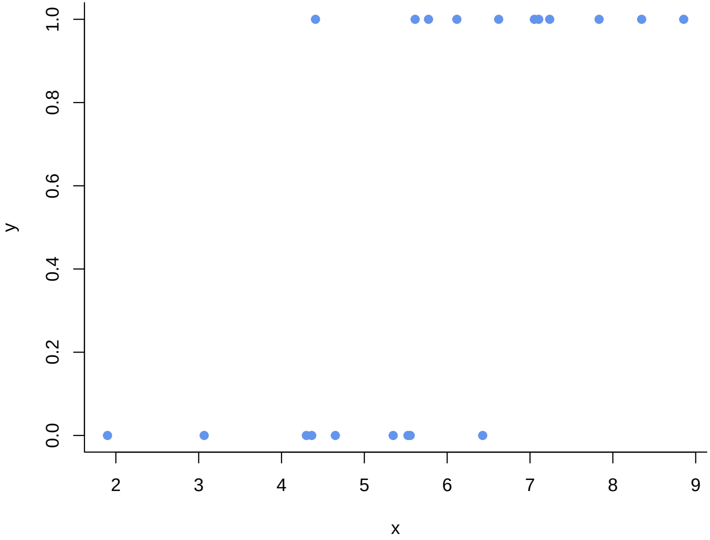


Describe some features of the response data y

Count data

- Response y is a count (discrete), e.g., $y = 1.25$ impossible
- Response can be zero or a positive integer
- Apparent positive non-linear relationship with continuous variable x
- **Example** y could be an organism count per unit area (abundance)
    - Discuss what values of abundance are possible

Describe some features of the response data y



Describe some features of the response data y

Binary data

- Response y can be either a 1 or a 0 (or other binary categories, e.g., on/off)
  - Often it is a sum of positives out of a given number of trials, e.g., total number of heads in 10 coin flips
  - Key thing is that for any one flip there can only be 2 outcomes
- Apparent positive non-linear relationship with continuous variable x
- **Example** y could be maturity status (mature/immature) for an organism
  - Discuss other binary data examples

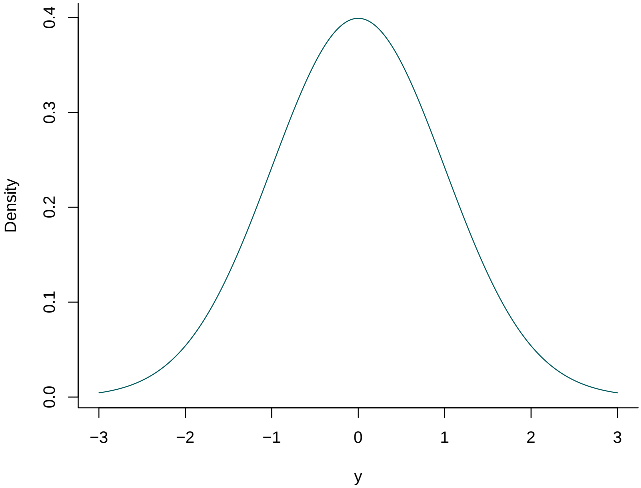
Outline

Probability distribution

A function that describes the probabilities associated with possible outcomes for an experiment (think of the response y)

Continuous probability distributions

Normal distribution



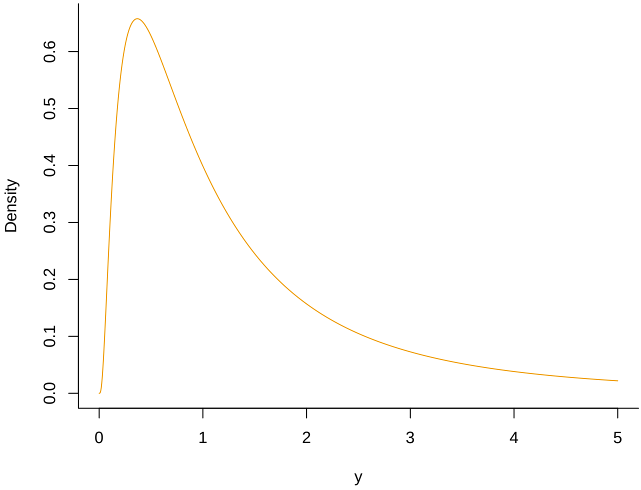
Continuous probability distributions

Normal distribution

- Distribution is continuous, e.g., $y = 1.25$ possible
- Positive or negative values possible (on the real line)
- Governed by two parameters: mean μ and variance σ^2
- Write: $y \sim \mathrm{N}(\mu, \sigma^2)$

Positive continuous probability distributions

Lognormal distribution

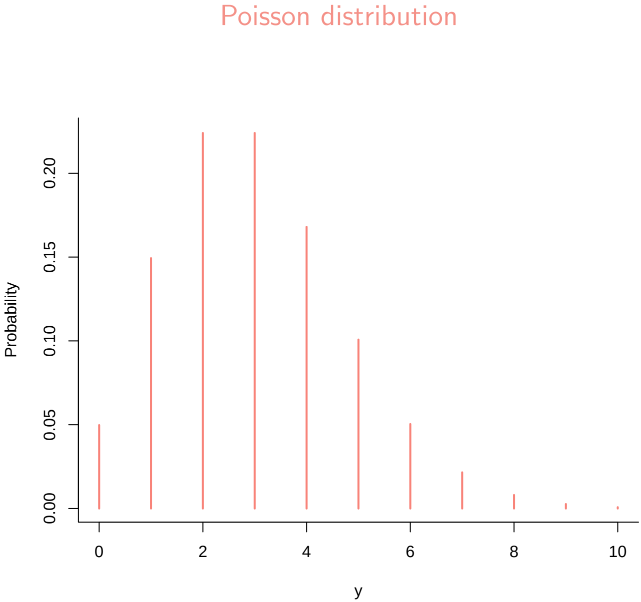


Positive continuous probability distributions

Lognormal distribution

- Distribution is continuous, e.g., $y = 1.25$ possible
- Only positive values possible (on the positive real line)
- Governed by two parameters: mean μ and standard deviation σ (both on log scale)
- Write: $y \sim \text{Lognormal}(\mu, \sigma)$

Count probability distributions



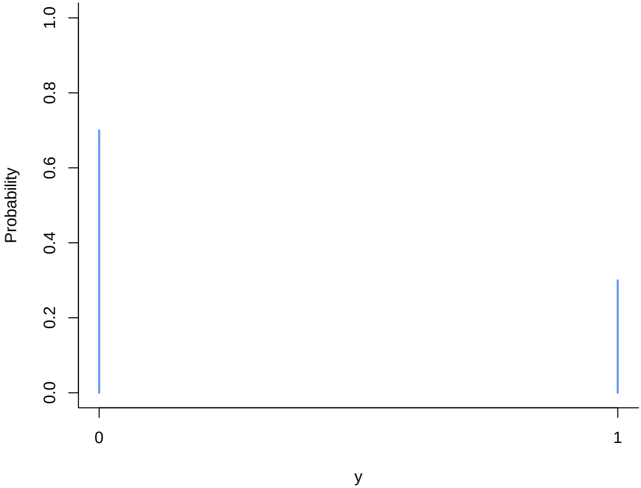
Count probability distributions

Poisson distribution

- Distribution is discrete, e.g., $y = 1.25$ impossible
- Distribution is only positive at zero and positive integers
- Governed by one parameter: rate λ (e.g., density)
  - Discuss rates in relation to counts
- Write: $y \sim \text{Pois}(\lambda)$

Binary probability distribution

Binary (Bernoulli) distribution

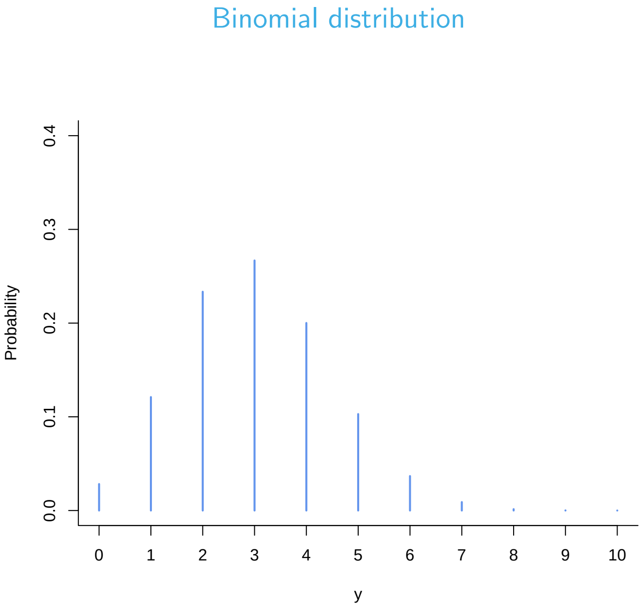


Binary probability distribution

Binary (Bernoulli) distribution

- Distribution over 0 or 1 (or other binary categories) only
- Governed by parameter: probability of success p (e.g., probability of being mature)
- Think: coin flip but coin not necessarily fair
- Write $y \sim \text{Bernoulli}(p)$

Binomial probability distribution



Binomial probability distribution

Binomial distribution

- Distribution over $\{0, 1, \ldots, n\}$ only
- Governed by 2 parameters: number of trials n (think: number coin flips) and probability of success p on any trial
- Write $y \sim \text{Bin}(n, p)$

Note: Binomial is the sum of Bernoulli trials

Outline

Explanatory variables[1]

Often a goal of an experiment or observational study is to relate observed response values to explanatory variables, e.g.,

[1] Also called "predictors"

Explanatory variables[1]

Often a goal of an experiment or observational study is to relate observed response values to explanatory variables, e.g.,

- Does mass depend on length and is this the same across areas?

[1] Also called "predictors"

Explanatory variables[1]

Often a goal of an experiment or observational study is to relate observed response values to explanatory variables, e.g.,

- Does mass depend on length and is this the same across areas?
- Does abundance relate to environment, e.g., temperature?

[1] Also called "predictors"

Explanatory variables[1]

Often a goal of an experiment or observational study is to relate observed response values to explanatory variables, e.g.,

- Does mass depend on length and is this the same across areas?
- Does abundance relate to environment, e.g., temperature?
- Does behavioural response depend on time-of-day?

[1] Also called "predictors"

Explanatory variables[1]

Often a goal of an experiment or observational study is to relate observed response values to explanatory variables, e.g.,

- Does mass depend on length and is this the same across areas?
- Does abundance relate to environment, e.g., temperature?
- Does behavioural response depend on time-of-day?
- ...

We would like to explore/model the relationships between the response and explanatory variables

[1] Also called "predictors"