Mixture models

Mark Andrews
Psychology Department, Nottingham Trent University

mark.andrews@ntu.ac.uk

Fitting parametric models

- Assume our data is n observations $y_1, y_2 \ldots y_n$.
- If we assume that

$$y_i \sim N(\mu, \sigma^2), \quad \text{for } i \in 1 \ldots n,$$

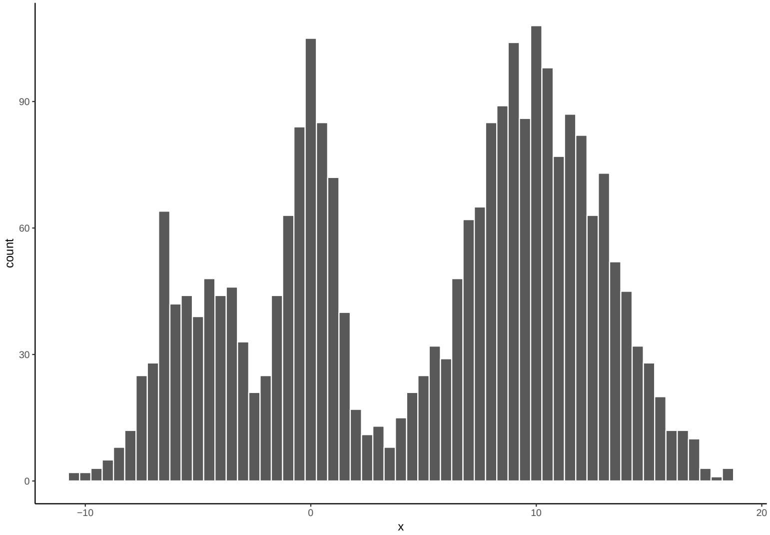
then we can calculate the likelihood function for μ and σ^2, i.e.

$$L(\mu, \sigma^2 | y_1 \ldots y_n) \propto \prod_{i=1}^{n} P(y_i | \mu, \sigma^2),$$

and maximize this function for μ and σ^2, or use it to calculate the posterior distribution.

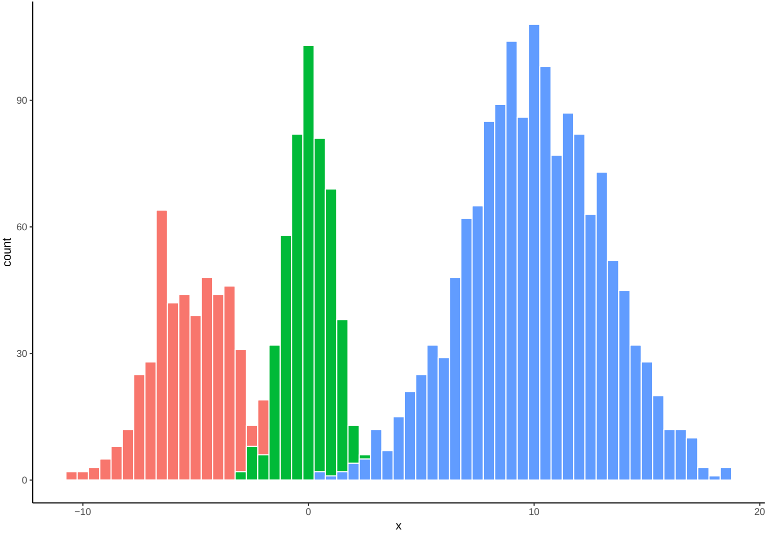
Irregular distributions

- What should we do when encounter data of the following form?



Mixture model

- A mixture model assumes that the data is sampled from independent component distributions, each of which can be modelled by parametric distributions.



Latent variables

- With irregular data, even if assume it is derived from a mixture of independent distributions, we do not know which data point came from which distributions.
- In other words, we have a set of data $y_1, y_2 \ldots y_n$, but we don't know which distribution each data point came from or even how many distributions there are.
- In this situation, we assume that for each y_i data point, there is an z_i that tells us which distribution y_i came from.
- This z_i is a *latent* variable. It has some value, but we don't or can't observe it directly.
- Another name for a model of this kind is a *latent class model*. We assume each y_i belongs to some class, but we just don't or can't observe what that class is.

Mixture models: The probabilistic generative model

- We start by assuming that there are K distinct hidden classes, e.g. $K = 3$.
- So each $z_i \in \{1, 2, 3\}$.
- Then, our model is

$$y_i \sim \begin{cases} N(\mu_1, \sigma_1^2), & \text{if } z_i = 1 \\ N(\mu_2, \sigma_2^2), & \text{if } z_i = 2 \\ N(\mu_3, \sigma_3^2), & \text{if } z_i = 3 \end{cases},$$

$$z_i \sim P(\pi),$$

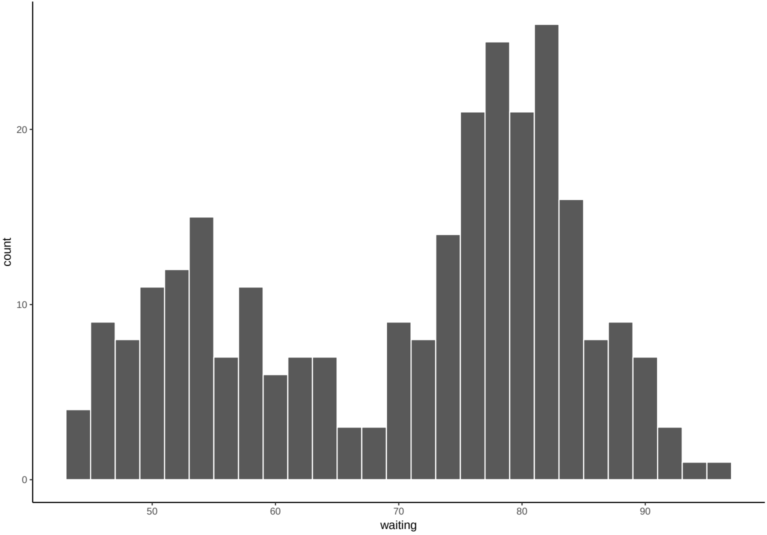
where $\pi = [\pi_1, \pi_2, \pi_3]$ is a probability distribution of $\{1, 2, 3\}$, i.e. π_1 gives the probability that the latent's class's value is class 1, π_2 gives the probability that the latent's class's value is class 2, π_3 gives the probability that the latent's class's value is class 3.

Mixture models: Inference

- In a normal mixture model with $K = 3$ components, we have 9 parameters:
  - μ_1, σ_1^2: The parameters of component distribution 1.
  - μ_2, σ_2^2: The parameters of component distribution 2.
  - μ_3, σ_3^2: The parameters of component distribution 2.
  - π_1, π_2, π_3: The relative probabilities of each component.
- In addition, we have the probability distribution over each value $x_1, x_2 \ldots x_n$.
- Inferring these values by maximum likelihood estimation is usually done by the *expectation-maximization* algorithm.

Example: Old faithful

- The distribution of waiting times.



Mixture regression models

- In a mixture of regressions, we assume that there are K regression models.
- Each data point being associated with one of them.
- Again, we don't know which component it came from. This is given by a latent variable.

$$
\begin{aligned}
y_i &\sim \begin{cases} N(\alpha_1 + \beta_1 x_i, \sigma_1^2), & \text{if } z_i = 1 \\ N(\alpha_2 + \beta_2 x_i, \sigma_2^2), & \text{if } z_i = 2 \end{cases}, \\
z_i &\sim P(\pi),
\end{aligned}
$$