Conclusion

- Day 1: Introduce common r.v. and distributions
- Day 2: Define likelihood, likelihood function. Model, data, parameters. Maximisation by calculus and R.
- Day 3: 4 properties of MLE. Likelihood-ratio test.
- Day 4: CI calculation. Joint CI. CI by approximate normality
- Day 5: Example in population genetics

Beyond this course

- Many more R functions and packages can help you implement MLE
- `nlm()`
- `mle(), confint(),` … etc from `{stats4}` package
- Require small changes in the "grammar" of the code

Models that make MLE difficult

- No close form solution in many cases. Consider Gamma distribution MLE, only numerical solution exists (cannot be solved by pen and paper)

- Some likelihood models are not explicit. It takes time (years) for researchers to write it down.

Mixed effect model

- Sometimes when we have a slightly complex model, like a mixed effect model
- Model: $\epsilon_i = y_i - a - bx_i - random\_effect_i$
- Even if we fixed our a and b, we still cannot compute the quantity of ϵ, because we don't know exactly the value of random effect
- Say, if $y_i - a - bx_i = 0.4$, we have infinitely many choices for $\epsilon_i + random\_effect_i = 0.4$
- We don't know which part of the 0.4 belongs to the residuals, which part belongs to the random effect

Latent variables

- Say, our data y follows $f(y|\lambda)$ but λ is another random variable that follows $f(\lambda|\theta)$
- Ultimately, we would like to know $f(y|\theta)$, the density of y given θ while λ is just an intermediate step

- By the law of total probability $f(y|\theta) = \int f(y|\lambda) f(\lambda|\theta) d\lambda$
- λ is integrated out (marginalised) here

- The pdf is an integral, and usually without an analytical solution
- So imagine if you have many data points… and you need to compute the joint density i.e. the product of these many pdfs
- And this is for one value of θ only. And we need iterate through many different values of θ to reach the maxima

- Computationally intensive!

State-space model

- Underlying process $\{p\}$ (cannot be observed)
- Observed values $\{x\}$ (contains sampling error)

$$
\begin{array}{ccccccccccc}
p_0 & \rightarrow & p_1 & \rightarrow & p_2 & \rightarrow & \cdots & \rightarrow & p_{t-1} & \rightarrow & p_t \\
\downarrow & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow \\
x_0 & & x_1 & & x_2 & & \cdots & & x_{t-1} & & x_t
\end{array}
$$

- High-dimensional integration

(Possible) solutions

- Approximation to the integrals (Laplace approximation)

- Statistical sampling (Gibbs, MCMC, MH)

- We need to learn more stats and computing!

MLE is…

- Not just a method, but THE method
- A collection of methods that share the same belief
- Many canned software and functions make use of the results from MLE (with or without telling you)