MAXIMUM LIKELIHOOD & PARAMETER ESTIMATION

BAYESIAN STATISTICS FOR ECOLOGISTS

IGB 18. TO 26. NOVEMBER 2019

THE LIKELIHOOD PRINCIPLE

- The likelihood of a statistical model is the probability of observing your data given that model: pr($\mathbf{X}$ | $\boldsymbol{\theta}$)
- This is "backwards" to normal inference – we want to know about the model. But the likelihood is easy to evaluate
- The **likelihood principle** states that all of the relevant information about the parameters of $\boldsymbol{\theta}$ in the dataset $\mathbf{X}$ is contained within the **likelihood function**

LIKELIHOOD FUNCTIONS

- So what is the likelihood function?
- Depends on the data
- You flipped a coin 100 times, observed 47 heads
- We want to know if the coin is fair

LIKELIHOOD FUNCTION PRACTICE

- k=47, n=100, estimate θ, the probability of observing heads (a “success”)
- Evaluate pr(n=47, k=100 | θ = 50)
- What is the likelihood function?

LIKELIHOOD FUNCTION PRACTICE

- k=47, n=100, estimate θ, the probability of observing heads (a “success”)
- Evaluate pr(n=47, k=100 | θ = 50)
- What is the likelihood function?

$$\mathcal{L}(n, k|\theta) \propto \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

LIKELIHOOD FUNCTION PRACTICE

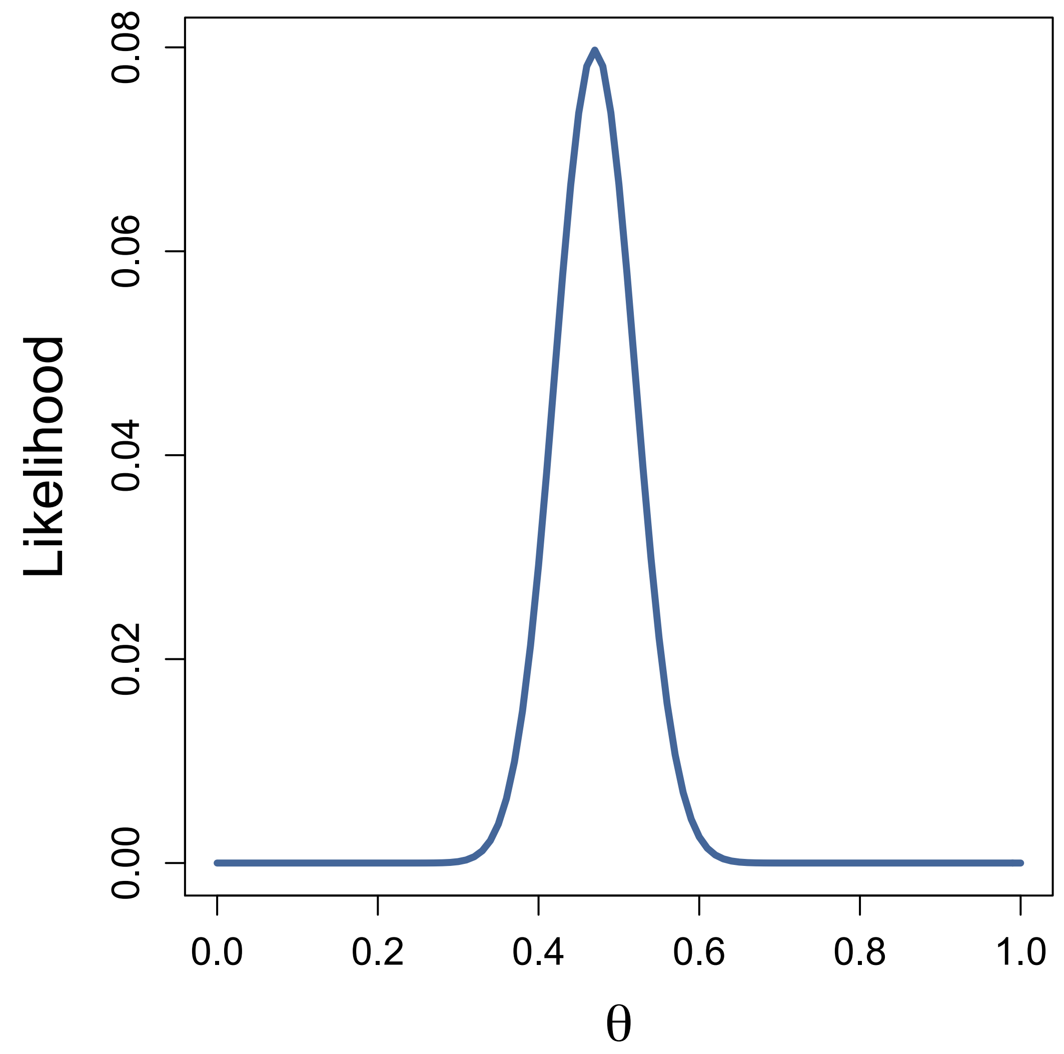
$$\mathcal{L}(n, k|\theta) \propto \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

- This is the binomial **PMF**, evaluate in R with `dbinom(47, 100, 0.5)`
- Plot the likelihood over various values of θ. Can you guess the value of θ that maximises the likelihood?

LIKELIHOOD FUNCTION PRACTICE

$$\mathcal{L}(n,k|\theta) \propto \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

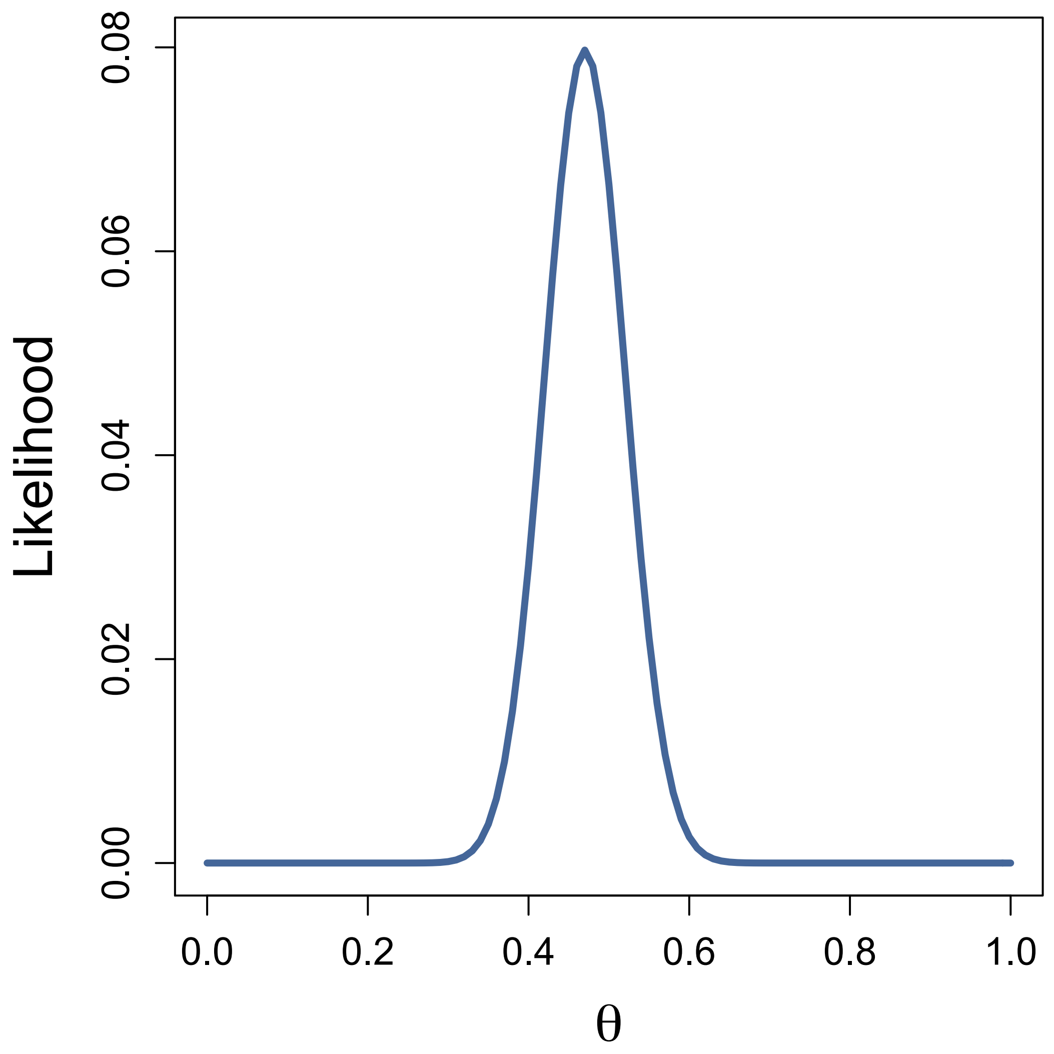
- This is the binomial **PMF**, evaluate in R with `dbinom(47, 100, 0.5)`
- Plot the likelihood over various values of θ. Can you guess the value of θ that maximises the likelihood?



LIKELIHOOD FUNCTION PRACTICE

$$\mathcal{L}(n, k|\theta) \propto \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

- This is the binomial **PMF**, evaluate in R with `dbinom(47, 100, 0.5)`
- How to precisely find the maximum likelihood estimate for θ where n=100, k=47?



MAXIMUM LIKELIHOOD

$$\mathcal{L}(n, k|\theta) \propto \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

MAXIMUM LIKELIHOOD

$$\mathcal{L}(n,k|\theta) \propto \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

- We can take the derivative of the function and set it equal to zero, then solve for θ

MAXIMUM LIKELIHOOD

$$\mathcal{L}(n,k|\theta) \propto \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

- We can take the derivative of the function and set it equal to zero, then solve for θ
- Can be difficult (or impossible) to compute analytically

MAXIMUM LIKELIHOOD

$$\mathcal{L}(n, k|\theta) \propto \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

- We can take the derivative of the function and set it equal to zero, then solve for θ
- Can be difficult (or impossible) to compute analytically
- We can also use an optimisation algorithm

MAXIMUM LIKELIHOOD

$$\mathcal{L}(n, k|\theta) \propto \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

- We can take the derivative of the function and set it equal to zero, then solve for θ
- Can be difficult (or impossible) to compute analytically
- We can also use an optimisation algorithm
- **Using `optim` in R, compute the maximum of the `dbinom` function with n = 47 and k = 100**

GENERALISING TO MORE THAN A SINGLE DATUM

$$\mathcal{L}(\mathbf{y}|\theta) = \prod_{i=1}^{n} \mathcal{L}(y_i|\theta)$$

GENERALISING TO MORE THAN A SINGLE DATUM

- For an i.i.d. data vector $\mathbf{y}_{1..n}$

$$\mathcal{L}(\mathbf{y}|\theta) = \prod_{i=1}^{n} \mathcal{L}(y_i|\theta)$$

GENERALISING TO MORE THAN A SINGLE DATUM

- For an i.i.d. data vector $\mathbf{y}_{1..n}$
- This is a direct consequence of the independence assumption

$$\mathcal{L}(\mathbf{y}|\theta) = \prod_{i=1}^{n} \mathcal{L}(y_i|\theta)$$

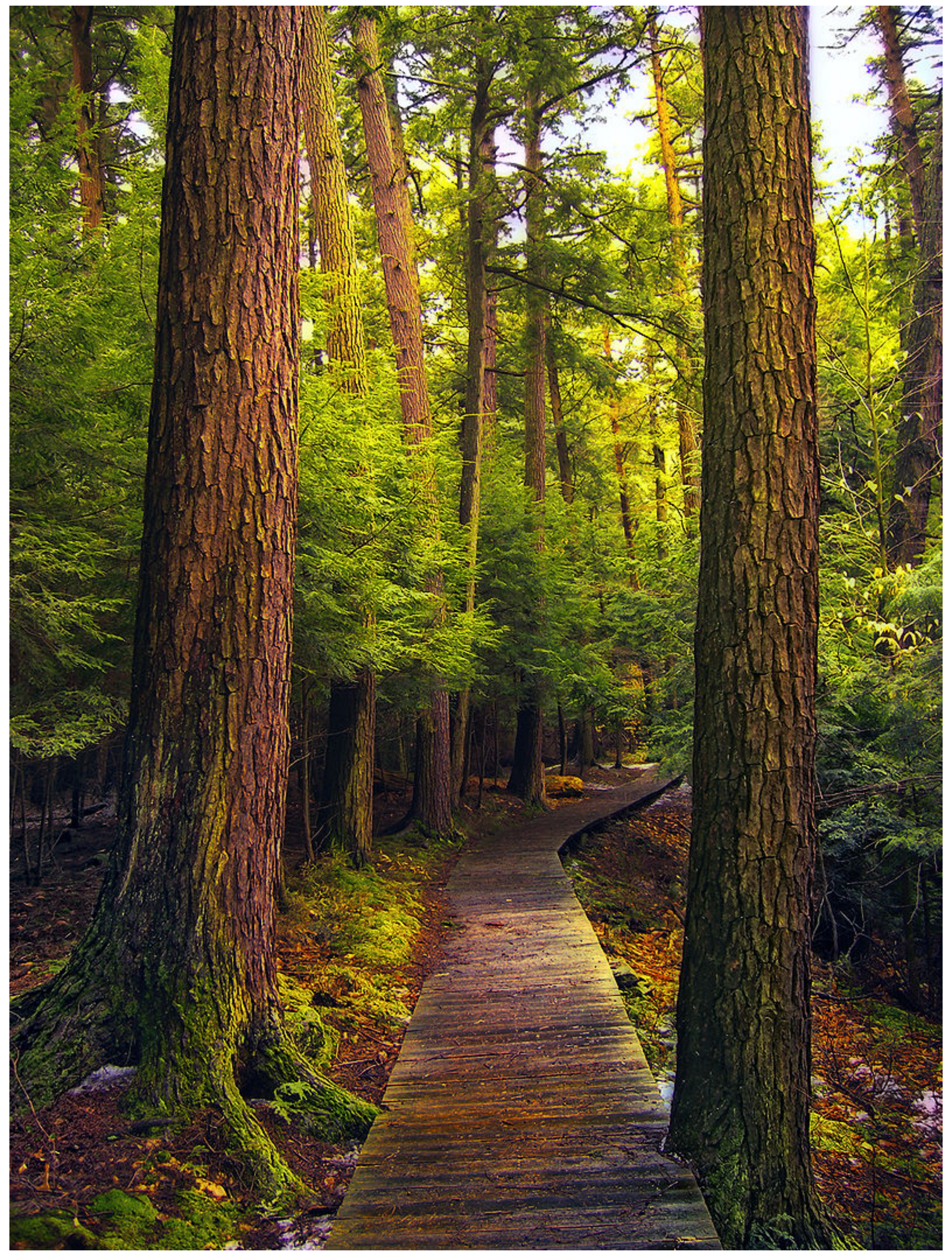
GENERALISING TO MORE THAN A SINGLE DATUM

- For an i.i.d. data vector $\mathbf{y}_{1..n}$
- This is a direct consequence of the independence assumption
- The product of many small numbers is an extremely small number, so we work with log likelihoods instead

$$\mathcal{L}(\mathbf{y}|\theta) = \prod_{i=1}^{n} \mathcal{L}(y_i|\theta)$$

$$\log \mathcal{L}(\mathbf{y}|\theta) = \sum_{i=1}^{n} \log \mathcal{L}(y_i|\theta)$$

ESTIMATING TREE MORTALITY RATE



- Use the `trees.rds` data file to estimate the mortality rate of *Tsuga canadensis* in 2005
- Write a function that takes three parameters:
    - theta (mortality rate)
    - n (vector of number of observed deaths)
    - k (vector of number of trees)
- and returns the log likelihood:
  $\log L(\mathbf{n},\mathbf{k} \mid \theta)$
- Plot the LL as a function of theta
- Find the value of theta that maximises this function
- Is the answer different from `mean(dat$died/dat$n)`, which is the average mortality rate by plot?

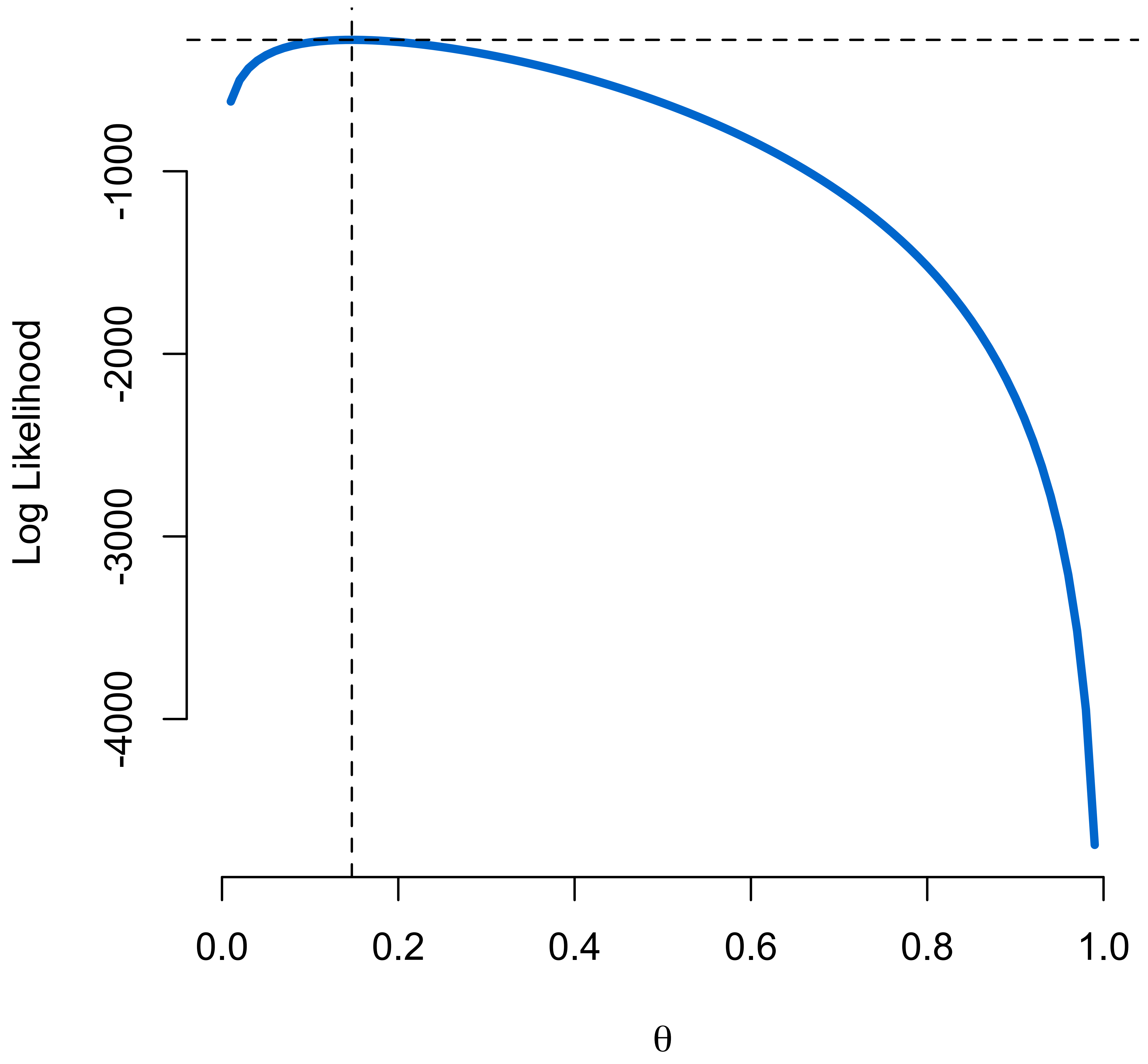
```r
library(data.table)
trees <- readRDS("data/trees.rds")
dat <- trees[grepl("TSU-CAN", species) & year == 2005 & n > 0]

lik_func <- function(theta, n, k) {

}

```

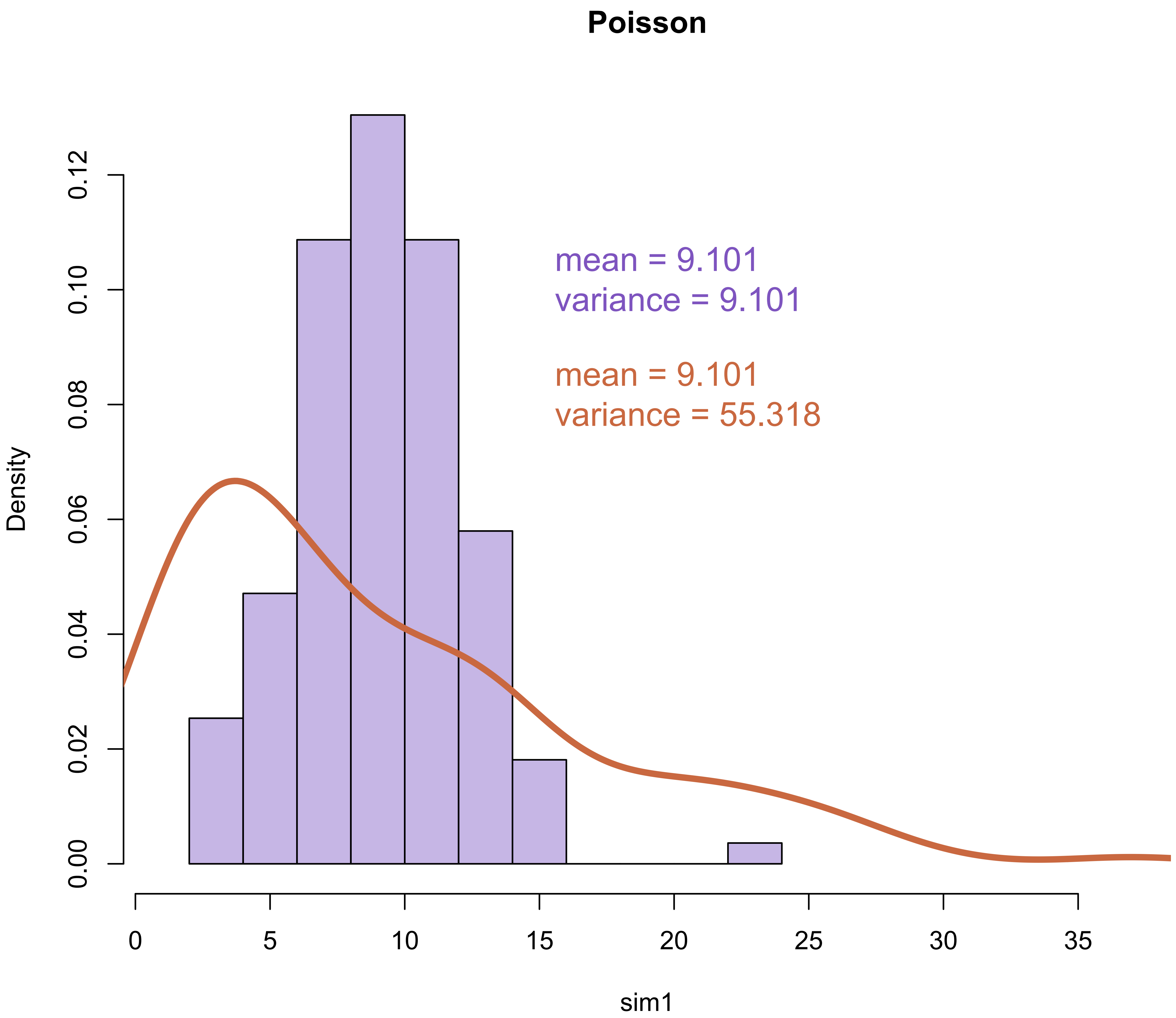
ESTIMATING TREE MORTALITY RATE



ESTIMATING TREE MORTALITY RATE

- Bonus questions:
  - use MLE to estimate the **average number of *Tsuga canadensis*** for the same species/year/plots (i.e., only plots that already had the species)
  - What likelihood function is appropriate?
  - Use your likelihood to generate a random dataset of the same size as the original.
  - Compare the histogram of tree counts of your simulation to the original. How could it be improved?

ESTIMATING TREE MORTALITY RATE



ESTIMATING TREE MORTALITY RATE