Likelihood

Hobbs 2015

DRME

December 26, 2023

Probability in Biology: Hobbs 4.1

- Example: Tadpole observation in a pond.

Example: Tadpole Observation

Scenario:

- Collecting data on the number of tadpoles per volume of water in a pond.
- Observed 14 tadpoles in a 1 L sample.
- **TRUE** average number of tadpoles per liter of water in the pond is 23.

First Observation:

- It is Poisson
- Probability of observing 14 tadpoles: $P(y_1 = 14|\lambda = 23) = \text{Poisson}(y_1 = 14|\lambda = 23) = 0.0136$.

Second Observation:

- Probability of observing 34 tadpoles: $P(y_2 = 34|\lambda = 23) = \text{Poisson}(y_2 = 34|\lambda = 23) = 0.0069$.

Joint Probability:

- Assuming independence: Joint probability $= 0.0136 \times 0.0069 = 9.38 \times 10^{-5}$.

Independence of Observations

- Independence assumption: Knowledge of one observation tells us nothing about the other.
- Joint probability calculation extended to any number of independent observations.

Remarks

- Probability calculations provide insights into the likelihood of observations given a fixed average.
- Independence assumption crucial for joint probability calculations.
- The Poisson distribution to model catching probabilities.

Probability in Biology: Hobbs 4.2

- Investigating decomposition of leaf litter over time.
- Using a simple model of exponential decay: $\mu_t = e^{-kt}$.
- Data: y_t - observed proportions, **modeled with a beta distribution**.
- Parameters: k (decay rate) and σ^2 (variance).

Beta Distribution for y_t

- Model the probability density of y_t with a beta distribution:

$$y_t|\mu_t, \sigma^2 \sim \text{beta}(\alpha_t, \beta_t)$$

- Moment matching for α_t and β_t:

$$\alpha_t = \mu_t^2 - \mu_t^3 - \mu_t\sigma^2$$
$$\beta_t = \mu_t - 2\mu_t^2 + \mu_t^3 - \sigma^2 + \mu_t\sigma^2$$

Conditional on Decay Rate k and σ^2

- Conditional on known, fixed decay rate $k = 0.01\,\text{day}^{-1}$ and known, fixed $\sigma^2 = 6 \times 10^{-4}$:
- Calculate parameters for the beta distribution on day 30:

$$\alpha_{30} = 236.33$$
$$\beta_{30} = 82.68$$

Probability Density Calculation

- Given $y_{30} = 0.7$, calculate the probability density:

$$f(y_{30} = 0.7) = 4.040$$

- Interpretation: The probability that 70% of the mass remains at time $t = 30$ is 4.040.

Remarks

- The beta distribution to model decay over time.
- Moment matching provides a method for estimating distribution parameters.

Introduction to Likelihood

- Likelihood measures the support provided by the observed data for different values of the parameter in a statistical model.
- The likelihood function is the foundation of maximum likelihood estimation (MLE).

Likelihood Function

- The likelihood function, denoted as $L(\theta; \mathbf{x})$, represents the probability of observing the given data $\mathbf{x}$ for various parameter values θ in the model.
- The likelihood function is not a probability distribution but provides a basis for estimating parameters.

Likelihood Example: Coin Toss

- Consider a simple example: coin toss.
- Let θ be the probability of getting heads ($\theta \in [0, 1]$).
- If we observe k heads in n tosses, the likelihood function is given by the binomial distribution:

$$L(\theta; k, n) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}$$

Interpretation of Likelihood

- Likelihood is not a probability, but it measures the compatibility of the observed data with different parameter values.
- Larger likelihood values indicate a better fit of the model to the observed data.
- The goal is to find the parameter values that maximize the likelihood, known as maximum likelihood estimation (MLE).

Likelihood in MLE

- Maximum Likelihood Estimation (MLE) aims to find the parameter values that maximize the likelihood function.
- MLE is a common method for estimating parameters in statistical models.
- It provides point estimates that make the observed data most probable under the assumed model.

Likelihood Example: Population Growth

- Consider a simple population growth model: $N_t = N_0 \cdot e^{rt}$, where N_t is the population size at time t, N_0 is the initial population size, r is the growth rate, and e is the base of the natural logarithm.
- Likelihood function: $L(r|\mathbf{y})$, where $\mathbf{y}$ is the observed population size over time.

Likelihood Example: Phylogenetic Trees

- In evolutionary biology, likelihood is extensively used in phylogenetic analysis.
- Given a phylogenetic tree and DNA sequence data, the likelihood of observing the given sequences under different substitution models is calculated.
- MLE finds the tree and model parameters that maximize the probability of the observed data.