AMES Section Notes – Week Five

Andie Creel

25th September, 2023

1   The language of limits

When we say that a sequence of numbers tends to a limit L as n tends to infinity, we mean that as n grows larger and larger, the terms of the sequence get closer and closer to the number L.

Let $f(n)$ be our sequence of numbers.

$$\lim_{n \to \infty} f(n) = L$$

Example. Let $f(n) = 1/n$. Then $L = 0$

$$\lim_{n \to \infty} \frac{1}{n} = 0$$

2   Problem set

2.1   Question 4

The key in question four is "The probability can be thought of as the fraction of acorns that geminates on average".

In question a, we're asked to find the distance from the tree with the highest *density* or *germinated* acorns. If we think of the probability as a fraction of gernminated acorns, what's our objective function? $F(X) = D(X)P(X)$

2.2   Question 8

Have we talked about taylor series approximations? A taylor series can be used to approximate any continuous function as a polynomial function. A line is a first order taylor series approximations. Typically, we never have rich enough data to approximate any function beyond a second order taylor series. In very rare cases, we may see a third order.

2.3   Question 9

Let's say we have data on y and x. We want to estimate the relationship between y and x using a first order taylor series (aka a line). Typically, applied researchers do this using linear regression. But, what is linear regression really doing??

What does the estimating equation look like?

$$y_i = a + bx_i + \epsilon_i \tag{1}$$

The goal of linear regression is to minimize the sum of squared errors (similar to finding an average, which we did in class today). If the goal is minimizing the sum of squared errors, what's our objective function? Let's first solve for our error term:

$$\epsilon_i = y_i - a - bx_i \tag{2}$$

If the objective is to minimize the sum of squared errors, our objective function would be:

$$F(a,b) = \sum_{i}^{n} (\epsilon_i)^2 \tag{3}$$

$$= \sum_{i}^{n} (y_i - a - bx_i)^2 \tag{4}$$

The solution is a solution for a as a function of the data $\{x_i, y_i\}$ and b as a function of the data $\{x_i, y_i\}$. We can define the solutions as

$$(a,b) = \arg\min_{a,b} \sum_{i}^{n} (y_i - a - bx_i)^2 \tag{5}$$

Solutions set up:

$$F_a = ... = 0 \tag{6}$$
$$F_b = ... = 0 \tag{7}$$

You have a system of two equations and two unknowns. Solve for $a(x,y)$ and $b(x,y)$.

The major trick of this problem is the trick of statistics. The *parameters* a and b are now unknown and the variables x and y are known. Realizing that we're solving for the parameter as a funciton of the data is the trick here (because we did the opposite on pset 1).