Course Review

SOC 512 & CSSS 505

written by Laina Mercer & Jessica Godwin

Aaron Osgood-Zimmerman

Department of Statistics
University of Washington

March 11, 2021

Review of Algebra & Functions

- Math notation
- Order of operations
- Rules of exponents, logarithms
- Equation of a line
- Functions, domain, range, examples
- Function transformations
- Continuous and piecewise functions
- Limits

Matrix Algebra

- Continuous vs. Piecewise Functions
- Limits
- Matrix Algebra
  - Definitions, notation
  - Matrix Arithmetic
  - Determinants - existence of an inverse
  - Linear equations
  - Least Squares and Regression with matrices

Matrix Arithmetic

Inverse Example

Once we know the inverse exists, we can find it.
For a 2×2 matrix,

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$A^{-1} = \frac{1}{D(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

where again $D(A) = a \cdot d - b \cdot c$. Example:

$$A = \begin{bmatrix} 4 & 12 \\ 3 & 6 \end{bmatrix}, \quad A^{-1} = \frac{1}{-12} \begin{bmatrix} 6 & -12 \\ -3 & 4 \end{bmatrix} = \begin{bmatrix} -1/2 & 1 \\ 1/4 & -1/3 \end{bmatrix}$$

Note: Finding the inverse for higher dimensions involves more comp;located formulas and is usually solved by a math software. (`solve()` in `R`)

Linear Equations

Examples

Solving our system of equations is the same as solving for z in the matrix equation:

$$A \cdot z = w$$

So how do we solve for z? First, left-multiply the equation by A^{-1}:

$$A^{-1} \cdot A \cdot z = A^{-1} \cdot w$$

By definition $A^{-1} \cdot A = I$. Thus,

$$I \cdot z = A^{-1} \cdot w \text{ or } z = A^{-1} \cdot w.$$

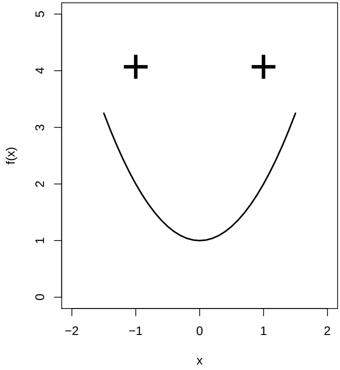
So we can find $z = (x, y)$, the solution to our system, by finding $z = A^{-1} \cdot w$.

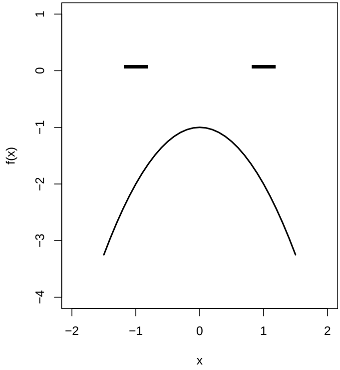
Differential Calculus

- Differentiation of functions
  - Defining the derivative
  - Basic differentiation rules
  - Second, third, etc... derivatives
- Critical points of functions
  - What is a critical point?
  - Maximum, minimum, and using the second derivative to tell the difference
- Taylor Series

Critical Values

For the max, the derivative decreases from positive to negative, so the second derivative will be negative. For the min, the derivative increases from negative to positive, so the second derivative will be positive.



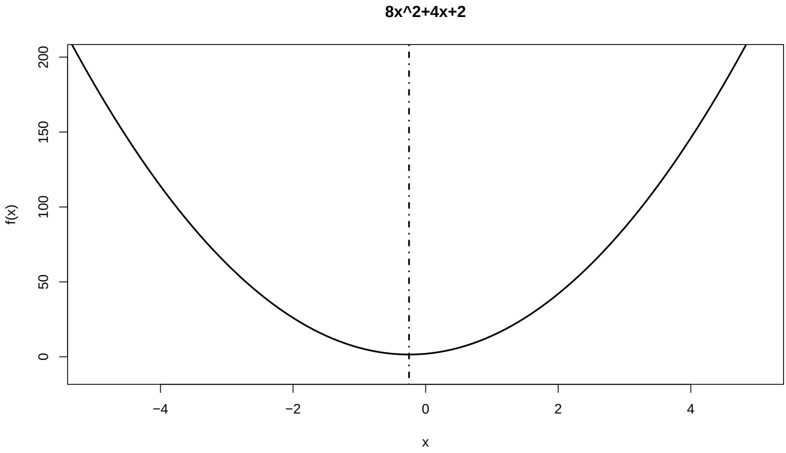


$$\begin{aligned} f(x) &= 8x^2 + 4x + 2 \\ f'(x) &= 16x + 4 \\ 0 &= 16x + 4 \Rightarrow 16x = -4 \Rightarrow x = \frac{-1}{4} \\ f''(x) &= 16 \end{aligned}$$

The cricital value is at $x = -1/4$ and the second derivative is positive, so it is a minimum.

Critical Values

Examples



Differentiation rules

distance, velocity, acceleration

Let's take d=distance, v=velocity, a=acceleration. You may remember from physics, the distance travel after time t

$$d(t) = \frac{a}{2}t^2$$

The velocity at any time t is the instantaneous rate of change of the distance, $v(t) = d'(t)$:

$$v(t) = 2 \cdot \frac{a}{2}t = at$$

The acceleration at any time t is the instantaneous rate of change of the velocity, $a(t) = v'(t) = d''(t)$:

$$a(t) = a$$

Distance

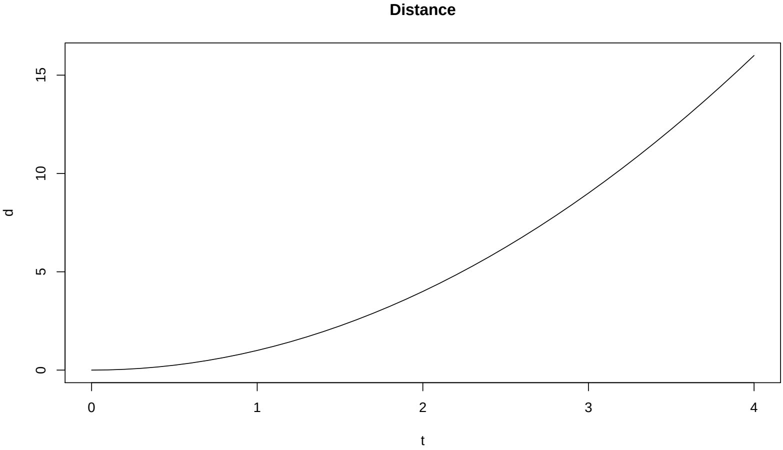


Figure: Distance over time, when $a(t) = 2$, $v(t) = 2t$, and $d(t) = t^2$.

Velocity

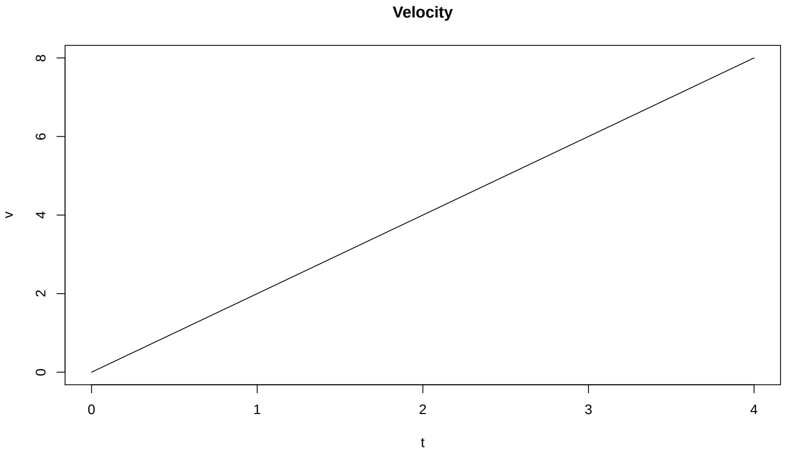


Figure: Velocity over time, when $a(t) = 2$, $v(t) = 2t$, and $d(t) = t^2$.

Acceleration

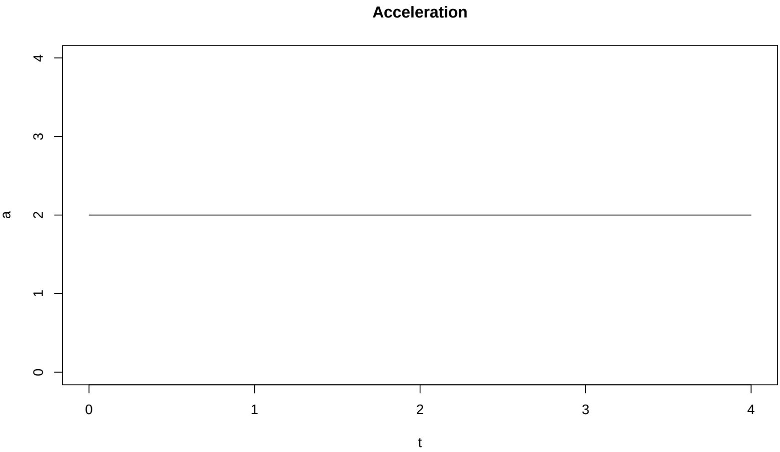


Figure: Acceleration over time, when $a(t) = 2$, $v(t) = 2t$, and $d(t) = t^2$.

Integral Calculus

- Motivation for Integrals
- Rules of Integration
- Lots of Examples

What is the velocity at t=3 when a=2?

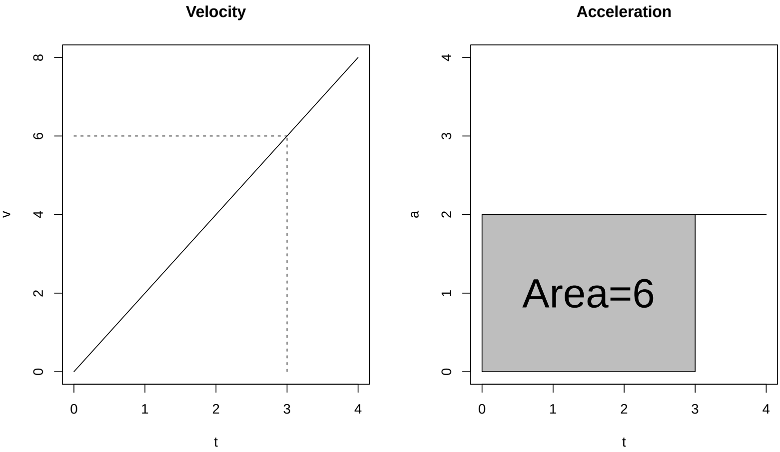
We know that $v(t) = 2t$, so clearly

$$v(3) = 2 \cdot 3 = 6.$$

However we can also find the velocity, by looking at the area under the acceleration curve from $t = 0$ to $t = 3$. This would just be the area of a rectangle (`base X height`),

$$(3 - 0) \cdot 2 = 3 \cdot 2 = 6.$$

What is the velocity at t=3 when a=2?



What is the distance at t=3 when a=2?

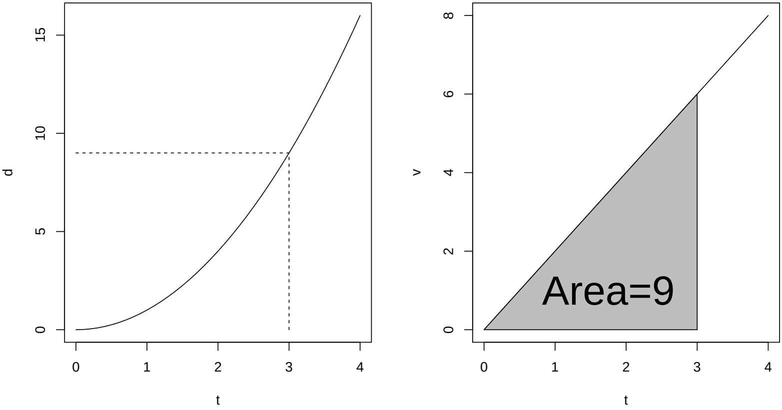
We know that $d(t) = 2/2t^2 = t^2$, so clearly

$$d(3) = 3^2 = 9.$$

However we can also find the distance, by looking at the area under the velocity curve from $t = 0$ to $t = 3$. This would just be the area of a triangle ($1/2$ `X base X height`),

$$1/2 \cdot (3 - 0) \cdot 6 = 3/2 \cdot 6 = 18/2 = 9.$$

What is the distance at t=3 when a=2?



Probability

- Notation
- Conditional Probability
- Bayes' Rule

A disease has a prevalence of 1% in the population. A blood test for the disease has high sensitivity (the probability of a positive test if someone is sick) and specificity (the probability of a negative test if someone is not sick).

- If someone has the disease, there is a 98% chance they will test positive.
- If someone does not have the disease, there is a 95% chance they will test negative

Suppose you test positive for the disease and you want to figure out the probability that you have the disease?

Probability

Motivating Example

What information do we have?

- $P(\text{diseased}) = 0.01$
- $P(+\text{ test}|\text{ diseased}) = 0.98$
- $P(-\text{ test}|\text{ healthy}) = 0.95$

What quantity do we want?

- $P(\text{diseased}|+\text{ test})$

So, what is the probability of disease given a positive test?

Bayes Rule

Testing Example

Let's use D^+ =diseased, D^- =healthy, $+$ =positive test, and $-$ =negative test.

- $P(+|D^+) = 0.98,\ P(-|D^+) = 0.02$
- $P(-|D^-) = 0.95,\ P(+|D^-) = 0.05$
- $P(D^+) = 0.01,\ P(D^-) = 0.99$

$$\begin{aligned}
P(D^+|+) = \frac{P(+|D^+)P(D^+)}{P(+)} &= \frac{P(+|D^+)P(D^+)}{P(+\cap D^+) + P(+\cap D^-)} \\
&= \frac{P(+|D^+)P(D^+)}{P(+|D^+)P(D^+) + P(+|D^-)P(D^-)} \\
&= \frac{0.98\cdot 0.01}{0.98\cdot 0.01 + 0.05\cdot 0.99} \\
&= 0.165
\end{aligned}$$

Random Variables & Probability Density Functions

- Definition of Random Variables
- Introduction to Probability Distributions
- Calculation of Means & Expectations
- Calculation of Variance
- Linearity of Means and Variances

Means & Expectations

What if all of the values were not equally likely? We find the *Expected Value* ($E[X]$) using a weighted mean:

$$E[X] = \sum_{i=1}^{n} x_i \cdot p_i$$

Like the expectation, the variance can be thought of as a weighted mean.

$$Var[X] = \sum_{i=1}^{n} (x_i - E[X])^2 \cdot p_i$$

Think of the variance as a weighted average of the difference between the x-values and the mean (squared to make all values positive).

Summary of Expectations & Variances

$$E[aX + b] = aE[X] + b$$
$$Var[aX + b] = a^2 Var[X]$$

- $E[X] = 3,\ Var[X] = 1,\ Y = X + 1$
  $E[Y] = E[X] + 1 = 4,\ Var[Y] = Var[X] = 1$
- $E[X] = 4,\ Var[X] = 1/2,\ Y = 3X$
  $E[Y] = 3E[X] = 12,\ Var[Y] = 3^2 Var[X] = 9/2$
- $E[X] = 4,\ Var[X] = 1/2,\ Y = 3X + 1$
  $E[Y] = 3E[X] + 1 = 12 + 1 = 13,\ Var[Y] = 3^2 Var[X] = 9/2$

The expectation for a continuous random variable is calculated using an integral (not a sum):

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

The calculation of the variance is quite similar:

$$Var[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f(x)dx$$

Discrete Distribution

- Bernoulli
- Binomial
- Multinomial
- Geometric
- Hyper Geometric
- Poisson
- Negative Binomial

Continuous Distributions

- Uniform
- Univariate Normal
- Chi-Square
- Student's t-distribution
- Exponential

Continuous Distribution

Exponential Example for HW 8

I just bought a new light bulb and am super excited to put it in my new batman night lamp. The time until the light bulb burns out follows an exponential distribution and the average burn out time is 1000 days. What is the probability the light bulb dies within the first 60 days?

Exponential Distribution with mean 1,000 days:

$$f(x) = \frac{1}{1000} e^{-x/1000}$$

We want to find $P(X < 60)$.

Continuous Distribution

Exponential Example for HW 8

Exponential Distribution with mean 1,000 days:

$$f(x) = \frac{1}{1000} e^{-x/1000}$$

$$\begin{aligned}
P(X < 60) &= \int_0^{60} \frac{1}{1000} e^{-x/1000} dx \\
&= -e^{-x/1000}\Big|_0^{60} \\
&= -e^{-60/1000} - (-e^{0/1000}) \\
&= 1 - e^{-6/100} = 0.058
\end{aligned}$$

Introduction to Statistics & Maximum Likelihood Estimation

- Motivation
- Likelihood
- Maximum Likelihood
- Confidence intervals

Suppose you randomly sample n data points $(x_1, ..., x_n)$ from an exponential distribution. You want to estimate the expectation of the distribution. Reminder for the exponential distribution we have the following density function:

$$f(x_i) = \frac{1}{\beta} e^{-\frac{x_i}{\beta}}$$

Since the sample was random, the distributions of these points are independent. Therefore, we can compute the joint distribution by taking the product of the marginal distributions. Prove that the MLE is the sample mean.

$$f(x_1, ..., x_n) = \prod_{i=1}^{n} \frac{1}{\beta} e^{-\frac{x_i}{\beta}}$$

We want to find a the maximum likelihood estimator and create a 95% confidence interval.

Maximizing the Likelihood

Finding the Max

How do we find the maximum likelihood estimator

- Take the log of the likelihood function
- Take the derivative
- Set it equal to zero
- Solve for the parameter of interest
- Find the second derivative
- Evaluate second derivative at critical value, if negative, we have a max! (not necessary for standard distributions)

For a refresher check out the notes from Lecture 3.

Maximizing the Likelihood

Step 1: Take the log of the likelihood function

Likelihood:

$$L(\beta|x_1, ..., x_n) = \prod_{i=1}^{n} \frac{1}{\beta} e^{-\frac{x_i}{\beta}}$$

Loglikelihood:

$$\begin{aligned} l(\beta|x_1, ..., x_n) &= log\left[\prod_{i=1}^{n} \frac{1}{\beta} e^{-\frac{x_i}{\beta}}\right] = \sum_{i=1}^{n} log\left[\frac{1}{\beta} e^{-\frac{x_i}{\beta}}\right] \\ &= \sum_{i=1}^{n}\left[log\left(\frac{1}{\beta}\right) - \frac{x_i}{\beta}\right] = \sum_{i=1}^{n}\left[log\left(\beta^{-1}\right) - \frac{x_i}{\beta}\right] \\ &= \sum_{i=1}^{n}\left[-log\left(\beta\right) - \frac{x_i}{\beta}\right] \end{aligned}$$

Maximizing the Likelihood

Step 2: Take the derivative of the log likelihood with respect to β

Loglikelihood:

$$l(\beta|x_1, ..., x_n) = \sum_{i=1}^{n} \left[-log\left(\beta\right) - \frac{x_i}{\beta} \right]$$

Derivative wrt β:

$$\begin{aligned}
\frac{d}{d\beta} l(\beta|x_1, ..., x_n) &= \frac{d}{d\beta} \sum_{i=1}^{n} \left[-log\left(\beta\right) - \frac{x_i}{\beta} \right] \\
&= \sum_{i=1}^{n} \frac{d}{d\beta} \left[-log\left(\beta\right) \right] - \frac{d}{d\beta} \left[\frac{x_i}{\beta} \right] \\
&= \sum_{i=1}^{n} \left[\frac{-1}{\beta} + \frac{x_i}{\beta^2} \right]
\end{aligned}$$

Maximizing the Likelihood

Step 3: Set the derivative equal to zero and solve for $\hat{\beta}$ (the estimate of β)

Derivative wrt β:

$$\frac{d}{d\beta} l(\beta | x_1, ..., x_n) = \sum_{i=1}^{n} \left[\frac{-1}{\beta} + \frac{x_i}{\beta^2} \right]$$

Set equal to 0 and solve for $\hat{\beta}$:

$$\begin{aligned}
0 &= \sum_{i=1}^{n} \left[\frac{-1}{\hat{\beta}} + \frac{x_i}{\hat{\beta}^2} \right] = -\sum_{i=1}^{n} \left[\frac{1}{\hat{\beta}} \right] + \sum_{i=1}^{n} \left[\frac{x_i}{\hat{\beta}^2} \right] \\
\sum_{i=1}^{n} \left[\frac{1}{\hat{\beta}} \right] &= \sum_{i=1}^{n} \left[\frac{x_i}{\hat{\beta}^2} \right] \quad \Rightarrow \quad \frac{n}{\hat{\beta}} = \frac{\sum_{i=1}^{n} x_i}{\hat{\beta}^2} \\
\hat{\beta} n &= \sum_{i=1}^{n} x_i \quad \Rightarrow \quad \hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}
\end{aligned}$$

Maximum Likelihood Estimator

Distribution of the MLE

If we want to know how well we are estimating our parameter (calculate the uncertainty), we need to think about the distribution of $\hat{\beta}$.

From Lecture 6 (slide 22) we know how to find the mean and variance of a random variable multiplied by a constant.

If $X \sim$Exponential(β), then $E[X] = \beta$ and $Var[X] = \beta^2$. Thus,

$$E[\hat{\beta}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n}\beta = \frac{n\beta}{n} = \beta$$

and

$$Var[\hat{\beta}] = Var\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var[x_i] = \frac{1}{n^2}\sum_{i=1}^{n}\beta^2 = \frac{n\beta^2}{n^2} = \beta^2/n$$

There are some nice results in statistics that state as $n \to \infty$ this standardization of the MLE $\sqrt{n}[\hat{\beta} - \beta] \to N(0, \beta^2)$.

In practice we are always dealing with finite samples, so we use this asymptotic result to say that $\hat{\beta}$ is approximately distributed as $N\left(\beta, \frac{\beta^2}{n}\right)$.

How well does this approximation hold? The approximation is better for larger n.

Maximum Likelihood Estimator

Uncertainty

Why do we care about the distribution of the MLE? This is how we will generate uncertainty estimates (confidence intervals).

Based on our approximate distribution $Z = \frac{\beta - \hat{\beta}}{\hat{\beta}^2/n}$ is approximately $N(0, 1)$.

The standardized version of our MLE provides a framework for generating an interval that is expected to include the true parameter value a certain percentage of the time. Generally this is set at 95%.

We can generate a confidence interval based on

$$[\hat{\beta} - 1.96 \cdot \hat{\beta}/\sqrt{n}, \hat{\beta} + 1.96 \cdot \hat{\beta}/\sqrt{n}]$$

The End

Questions?