

MML - Practice for Exam 3

We will have our third exam this Friday! Here's our in class practice sheet.

The problems

1. Suppose that X has the continuous distribution

$$f(x) = \frac{1}{2}x$$

over the interval $[0, 2]$

- a. Write down the computation that shows that f is a good probability distribution.
 - b. Use an integral to compute the mean of X .
 - c. Using your computed mean from part (b), write down the integral that expresses the variance of X .
2. Use u -substitution to translate the normal integral

$$\frac{1}{\sqrt{50\pi}} \int_2^6 e^{-(x-3)^2/50} dx$$

to a *standard* normal integral.

3. I've got a coin that might very well be unfair. Suppose I flip that coin 200 times and I get 60 heads.
 - a. Based on that evidence, what's your best guess of the probability p that the coin comes up heads?
 - b. Given a value of p , use the binomial distribution to write down a function $f(p)$ that expresses the probability that the coin comes up heads 60 times in 200 flips.
 - c. Use calculus to find the value of p that maximizes f .
4. Find the eigenvalues and corresponding eigenvectors of

$$A = \begin{bmatrix} 3 & 1 \\ -2 & 0 \end{bmatrix}.$$

5. Let's suppose that excessive basketball watching causes tardiness. To study this problem, I collected data on 100 people. Below we see this data plotted and in a partial table.

Hours in March	Late at least once
55	1
46	0
52	1
24	0
\vdots	\vdots

- Suppose we model this data using logistic regression. What is the primary objective?
- Logistic regression produces an estimator function that you use to achieve your objective. When we have one input variable (as in this case), the estimator function depends upon two parameters - a and b . Write down the general formula for the estimator in terms of the parameters a and b .
- Suppose I have the three candidate pairs of values of a and b shown in Table 2 together with their associated log-loss. Which candidate pair (a, b) should I use for my estimator?
- What is the resulting probability estimate that an individual who watched 55 hours of basketball in March was late to work or school at least once during that time?
- Sketch a rough graph of your probability estimator function right on top of the plot.

Table 2: LR parameter candidates and their log-loss

a	b	Log-loss
0.152	7.34	0.959
0.23	966	0.828
0.108	5.94	1.401

