MML - Review for Exam 3

We will have our second exam next Friday, March 27. This review sheet is genuinely meant to help you succeed on that exam.

Generally, I would like to ensure that we know the formulae we need to use for various problems, how to use those formulae, and how to express their application in a mathematically coherent sentence. There are several problems along those lines on this review sheet.

For example, if I ask you to write down the formula showing that the average of 5 and 9 is 7, then you should write down

$$\frac{5+9}{2} = 7.$$

The problems

1. Suppose the discrete random variable X has the following discrete distribution:

i	$P(X=i)$
1	0.2
3	0.3
4	0.5

   a. Write down the computation that shows that the mean of X is $\mu = 3.1$.
   b. Write down the computation that shows that the variance of X is $\sigma^2 = 1.29$.

2. Suppose that X has the continuous, uniform distribution over the interval $[2, 6]$

   a. Write down the piecewise defined distribution function for X.
   b. Write down the computation that shows that the mean of X is $\mu = 4$.
   c. Write down the computation that shows that the variance of X is $4/3$.

3. Let Z denote a random variable whose distribution is the standard normal.

   a. Write down the integral that shows that
   $$P(-1 < Z < 1) \approx 0.68.$$
   b. Write down the integral that shows that
   $$\mu(Z) = 0.$$
   c. Write down the integral that shows that
   $$\sigma^2(Z) = 1.$$

4. Use u-substitution to translate the normal integral
$$\frac{1}{\sqrt{18\pi}} \int_0^5 e^{-(x-2)^2/18}\, dx$$
to a *standard* normal integral.

5. I've got a coin that comes up heads 80% of the time. Suppose I flip that coin 25 times. What's the probability that I get 20 heads?
You should express your answer in terms of the binomial distribution.

6. I've got a coin that might very well be unfair. Suppose I flip that coin 100 times and I get 25 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 25 times in 100 flips.
   c. Use calculus to find the value of p that maximizes f.

   This is essentially a simple example of maximum likelihood technique.

7. Suppose I've got a categorical variable that can take any of the three values good, bad, or ugly.

   a. Describe conceptually how one hot encoding would be set up for that variable.
   b. Given your description, what would be the encoding of the vector
   $$[3.14159, \text{ugly}]^{\mathsf{T}},$$
   where the first value is the value of some separate numeric variable.

8. Let's suppose that excessive coin flipping causes arthritis of the thumb. To study this problem, I collected data on 200 people as shown in Table 2.

Table 2: Flips per day and occurrence of arthritis

Flips per day	Arthritis Outcome
78	1
56	0
57	1
20	0
⋮	⋮

Note that a plot of this data is also shown in Figure 1.

Let's use logistic regression to model this situation.

a. What is the primary objective of logistic regression in the context of this problem?
b. 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.
c. Suppose I have the three candidate pairs of values of a and b shown in Table 3 together with their associated log-loss. Which candidate pair (a, b) should I use for my estimator?
d. What is the resulting probability estimate that an individual who flips a coin 60 times per day develops arthritis of the thumb?
e. Sketch a rough graph of your probability estimator function right on top of Figure 1.

Table 3: LR parameter candidates and their log-loss

a	b	Log-loss
0.152	7.34	0.959
0.232	8.1	1.24
0.108	5.94	0.828

9. Find the eigenvalues and corresponding eigenvectors of

$$A = \begin{bmatrix} 3 & 5 \\ 0 & -2 \end{bmatrix}.$$

10. Determine whether $\mathbf{v} = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^\mathsf{T}$ is an eigenvector of

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

If so, what is the corresponding eigenvalue?

Images

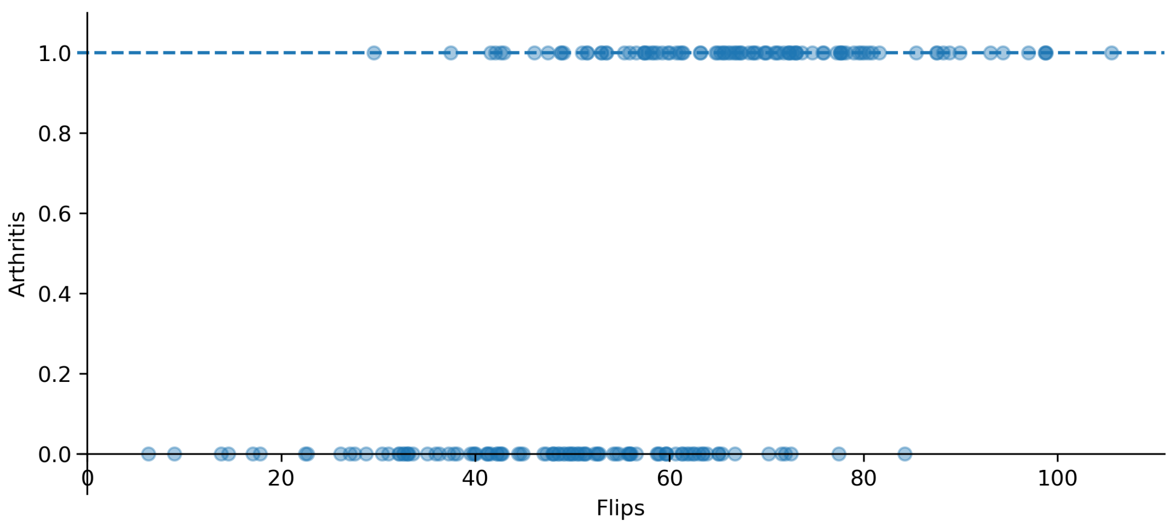

Figure 1: Data for a logistic regression