Day 1 – MTH 490.433

Basic Probability Problem Set

The five problems in this section can all be done by hand.

1) A roulette wheel has 37 positions numbered 0,1,…,36. Assume that the ball comes to rest at each position with equal probabilities. What is the probability

a) of an even number,

b) of a number greater than 30,

c) of a number which is at most 10?

2) In a population of plants, individuals can have white flowers instead of the normal purple flowers with a probability of 0.02. Independently, plants in this population can have short anthers with a probability of 0.013. What is the probability of finding one of these rare plants, in other words a plant having white flowers and/or short anthers? Note, this problem can be solved by two subtly different ways of thinking about the problem.

3) Derive the formulas for the mean and variance of a uniform distribution from the probability density function and the definition of mean and variance. The probability density function of a uniform distribution is $f(z)=1/(b-a)$ where where $a \le z \le b$.

Day 1 – MTH 490.433

4) The body mass of a population of adult beetles is normally distributed with a mean mass of 18.7 and a standard deviation of 1.3. You want to select some large beetles for an experiment. Identify the size which defines the largest 7% of the beetles. Show how you would find this size, but you don't necessarily have to solve the equation. (Hint: a graph of the probability distribution function may help you think through this problem)

5) Now that you've achieved statistical mastery, it is time to create your very own probability distribution function! Create a probability density function defined for *z* between 2 and 4. Make the function linear and proportional to the line $(1 + 3x)$. Don't forget about the rule that must be true for all pdf's.

Sketch a graph of this new distribution and write the formula for the pdf. You should be able to do this all by hand (Optional bonus: do this all symbolically, in other words proportional to the line $m\,x + b$ defined on the interval [p,q]).

Day 1 – MTH 490.433

Exercises in Simulating Data

Example 1

Imagine a plant with a single female flower. The plant will either be visited by a pollen bearing pollinator or not, with a probability of success *p*. Examine across a number of patches. In this example, the probability of being pollinated is 0.85. We have 20 flowers in each patch, and 10 patches.

R code

```
visits <- rbinom(10,20,0.85)
plot(visits,ylim = c(0,20))
```

Mathematica code

```
<<Statistics`DiscreteDistributions`
visits=Table[Random[BinomialDistribution[20,0.85]],{10}]
ListPlot[visits,PlotStyle→PointSize[0.02],PlotRange→{0,20}];
```

Example 2

a) Imagine you now have a camera set-up to record the number of pollinator visits to each individual flower. Now a flower can be visited multiple times. The rate of pollinator visits is a continuous value, but the number of individual visits on any particular plant is a discrete value. The rate of pollinator visits times the period of time over which the flower is observed is 2.26. The outcome is the discrete number of visits.

b) Now imagine the rate of visitation depends on the size of the flower. Bigger flowers get more visits,

$$E[v] = 2.00 + 0.2\, w,$$

where $E[v]$ is the expected number of visits and w is the width of the flower corolla.

Example 3

Imagine you sample the temperature in 40 experimental ponds. The average temperature is 25.0.

Example 4

You are trying to estimate the rate of parasitism by Chewing Bird Lice in two different species. In a preliminary study, birds had been parasitized with an average of 12.3 lice on each bird. You have captured 35 Kirtland Warblers and 83 Bell's Vireos to look at differences in lice parasitism between the species. Simulate your results, assuming the Warblers are parasitized at a higher rate than the Vireos. Repeat the simulation twice using a different distribution form each time (hint: one of the distributions being the assumed distribution in classical statistics).

Day 1 – MTH 490.433

Useful Code for Chapter 4

Return to the pollination example. We watch 100 flowers and count the number of times each flower is visited by a pollinator. 31 flowers are never visited. 41 flowers are visited once, etc.

visits	flowers
0	31
1	41
2	15
3	12
4	1

Mathematica Code

```
<<Statistics`DataManipulation`;
exampleData={{0,31},{1,41},{2,15},{3,12},{4,1}};
visits=Column[exampleData,1];
flowers=Column[exampleData,2];
```

R Code

```
exampleData = data.frame(list(visits = c(0,1,2,3,4), flowers =
c(31,41,15,12,1)))
```

Find the average number of pollinators that visited a flower

Mathmatica Code

```
N[Total[visits*flowers]/Total[flowers]]
```

R Code

```
with(exampleData,sum(flowers*visits)/sum(flowers))
```

If we think this is Poisson distributed, we can simulate a Poisson data set.

Mathematica Code

```
<<Statistics`DiscreteDistributions`
simData=RandomArray[PoissonDistribution[1.11],{100}]
```

R Code

```
simData <- rpois(100,1.11)
table(simData)
```

We can compare the observed data to the expected distribution one would get from a Poisson distribution with 100 samples and a mean of 1.11. Use the Poisson pdf and generate the probability of getting 0, 1, 2, 3, and 4 visits if the rate parameter is 1.11.

Mathmatica Code

```
Table[PDF[PoissonDistribution[1.11],i],{i,0,4}]
```

R Code

```
dpois(0:4,1.11)
```

Day 1 – MTH 490.433

Multiply each fraction by the total number of flowers to get the expected. We will extend the range of expected out to 10 in order to capture the entire distribution.

Mathematica Code

```
expected=100 Table[PDF[PoissonDistribution[1.11],i],{i,0,10}]
observed=flowers
```

R Code

```
expected <- 100*dpois(0:10,1.11)
observed <- exampleData$flowers
```

Now I can use the χ^2 test to compare the observed and expected distributions. This will give me absolute measure of how well the Poisson distribution fits these data.

$$X^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

This statistic can then be compared to a χ^2distribution to yield the probability that the observed and the expected come from the same distribution. As a rule of thumb, the expected value for each class should be greater than 5. If the expected value is less than 5 for any category, it should be combined with an adjacent class. Combine all the small expected classes and any corresponding observed classes.

Mathematica Code

```
expected=Append[Take[expected,3],Total[Drop[expected,3]]]
observed=Append[Take[observed,3],observed[[4]]+observed[[5]]]

testStat=Total[(observed-expected)^2/expected]
<<Statistics`HypothesisTests`
ChiSquarePValue[testStat,3]
```

R Code

```
expected = c(expected[1:3],sum(expected[4:10]))
observed = c(observed[1:3],observed[4]+observed[5])
chisq.test(observed,p=expected,rescale.p=TRUE)
```

Because we are comparing 4 classes, we have 3 degrees of freedom. We cannot reject the null hypothesis at the 0.05 level, so we consider the Poisson distribution to be an adequate fit to the data.

Finally, this last bit of code can be used to generate the original data from a contingency table.

Mathematica Code

```
Flatten[Apply[Table[#1,{#2}]&,exampleData,1]]
```

R Code

```
with(exampleData,rep(visits,flowers))
```

Day 1 – MTH 490.433

Example from Chapter 4

Try to work through this exercise without reading Chapter 4. It will be more informative for you to think through the distributions before you read their suggestions. After you've worked through the data here, read Chapter 4, pages 94-95, 98-103. It will give you a little more background. See if you agree with their analysis.

The data are the number of New Zealand fish tows that had an incidental catch of seabirds. For example, 807 tows had no birds caught in the gear. 37 tows had 1 bird caught in the gear, etc. Find a distribution that fit these data. You can download the data on the course website. Under Day 1 it is called bird data <http://www.unl.edu/cbrassil/ELME/2007/ch4.csv>. You can import the data using the following code.

Mathematica Code

```
mydata=Import["C:\ch4.csv","CSV"];
```

R Code

```
mydata <- read.csv("ch4.csv")
```

Birds	Tows
0	807
1	37
2	27
3	8
4	4
5	4
6	1
7	3
8	1
9	0
10	0
11	2
12	1
13	1
14	0
15	0
16	0
17	1