QUEEN MARY, UNIVERSITY OF LONDON

MTH 4106 Introduction to Statistics

Practical 4 31 January 2012

Today we will use Minitab's inbuilt information about the most important discrete random variables. We will learn how to

- list the values of the probability mass function or cumulative distribution function, and plot these on a graph;
- use Minitab like a book of statistical tables to calculate quantities like $\mathbb{P}(2 \leq X \leq 6)$;
- verify particular cases of the claims that hypergeometric random variables are approximately binomial and that binomial random variables are approximately Poisson.

1 (The probability mass function of a binomial random variable) We are going to list the probability mass function for a random variable X whose distribution is $\mathrm{Bin}(9, 0.5)$. This random variable takes values 0, 1, …, 9, so we have to start by making a column called `number` with these values in it. Do this by

`Calc` → `Make Patterned Data` ▶ `Simple Set of Numbers....`

Store the patterned data in `number`, From first value 0, To last value 9, In steps of 1. List each value and the whole sequence once.

Now click on

`Calc` → `Probability Distributions` ▶ `Binomial...`,

click on `Probability`, then make the `Number of Trials` equal to 9 and make the `Event probability` equal to 0.5. Set the `Input column` to `number` and Click on `OK`. What does Minitab do?

Do the second step again, but this time choose to store the new variable in a column named `bin9pr`.

2 (The cumulative distribution function of a binomial random variable) Get the `Binomial` menu back for a third time. This time, click on `Cumulative probabality` and store the result in `bin9cdf`.

Compare the two columns `bin9pr` and `bin9cdf`. How can you calculate one from the other?

3 (Plotting the pmf and cdf) Plot `bin9pr` against `number` on a scatterplot. Then plot `bin9cdf` against `number` on another scatterplot.

If you find these graphs useful, you may want to copy them to a report pad and print them off for inserting in your lecture notes. When you have done that, remember to close both graphs, to free up working memory in Minitab.

4 (Calculating the probability of being in an interval) If $X \sim \text{Bin}(9, 0.5)$ then we may want to calculate a quantity like $\mathbb{P}(5 \leq X \leq 7)$. You can work that out from the column `bin9pr` or from the column `bin9cdf`. Now we will see how to do it without printing those two columns in full.

Get the `Binomial` menu back yet again. Click on `Cumulative probabality`. Also click on `Input constant`, and enter `7`, choosing to store the result in `upto7`.

Minitab does not seem to put the result anywhere! Use

`Data` → `Display Data...`

and choose `upto7`. The **Session Window** should now display $\mathbb{P}(X \leq 7)$. What is this probablity equal to?

In the same way, calculate a constant called `upto4` equal to $\mathbb{P}(X \leq 4)$. Then use the `Calc` menu to calculate `upto7` − `upto4`. Where has Minitab put the result? If necessary, display it in the **Session Window**, and hence complete the box below.

$\mathbb{P}(5 \leq X \leq 7) =$

5 (Calculating an interval probability for a geometric random variable) Minitab knows about many other distributions in addition to the binomial. Suppose that $Y \sim \text{Geom}(0.1)$. Use the method in the previous question to find

$\mathbb{P}(5 \leq Y \leq 10) =$

6 (Hypergeometric random variables) Suppose that we have a field of 40 sheep, of which 20 are black and 20 are white. We choose a random sample of 9 sheep without replacement. Let W be the number of white sheep in the sample. Then $W \sim \text{Hg}(9, 20, 40)$.

Use Minitab's inbuilt distributions to create a column called `hg9pr` giving the probability mass function for W.

7 (Comparing the hypergeometric and binomial distributions) If, instead, we take a random sample of 9 sheep *with* replacement, then the number of white sheep in the sample has the $\text{Bin}(9, 0.5)$ distribution, because $20/40 = 0.5$. Lecturers in both *Introduction to Probability* and *Introduction to Statistics* claim that these two distributions are almost the same if the sample size is "very small" in comparison with the numbers of black sheep and white sheep. Compare the columns `hg9pr` and `bin9pr`. Are they similar?

8 (Plotting two distributions on the same graph) It might be easier to compare these two columns if we plotted both of them on the same graph. Get the menu for `Scatterplot`, and enter

```
bin9pr  number
hg9pr   number
```

Use the button `Multiple Graphs...` to plot them on the same graph.

Where are these two probability mass functions most far apart?

9 (A Poisson random variable) The other family of discrete random variables that we have met consists of the Poisson random variables. Suppose that $V \sim \text{Poisson}(6)$. Create a column showing the probablility mass function for V. Strictly speaking, your input column will need to go all the way to ∞, so you will need to decide where is a sensible place to stop it.

10 (Poisson approximation to binomial) In *Introduction to Probability*, it was shown that $\text{Bin}(n, p)$ is approximately the same as $\text{Poisson}(np)$ if n is large. Create three new columns giving the probablity mass functions of three binomial random variables. Put $n = 10$, $n = 20$ and $n = 30$. In each case, choose the value of p to make the distribution close to the Poisson distribution that you have already tabulated.

Make a single graph plotting the probability mass functions of these three binomial distributions and the probability mass function of your Poisson distribution. You may want to maximize it to the full screen so that you can see all the points distinctly.

Append this graph to a report pad, and then edit the report pad by adding a sentence or two explaining what this graph shows.