Math 1180 - Mathematics for Life Scientists

Computer Assignment 11 - Maximum Likelihood

Due Wednesday, April 26, 2006

Getting Started

Warm up Maple for today's problems with the commands:

```
> with(stats);
> iread(binomial);
> iread(histplot);
> iread(draw);
> iread(Poisson);
> iread(iter);
```

A short summary of instructions for the binomial and Poisson packages are given on the last page.

Problems

1. Simulate the following expertiments.

   (a) People have a 0.4 chance of having black hair. Simulate 50 groups of five people by drawing from the binomial distribution with $n = 5$ and $p = 0.4$. Generate 50 numbers from this distribution,

   ```
   > a := [seq(bindraw(5,0.4),i=1..50)]:
   > transform[tally](a);
   ```

   How many groups have exactly two out of five with black hair?

   (b) Supppose instead two out of five people are found with black hair. Simulate 50 groups of five people with different unknown probabilities p that a person has black hair. Use values of p ranging from 0 to 1.0 - try $p = 0.1, 0.2, 0.4, 0.8$. Which value produces the most groups that exactly match the data?

   (c) Find the likelihood function describing this case. Input this function into Maple and plot it. Where is the maximum? How does this compare to your previous simulation in (b)?

   (d) Explain how the simulation in (a) differs from the one done in (b). In each case, indicate what is known and what is unknown.

2. Cells are placed for one minute in an environment where they are hit by X-rays, some of which are damaging. Cells not hit by the damaging rays are healthy, those hit exactly once are damages, and those hit more than once are dead. By measuring the states of a number of cells, we wish to infer the rate at which cells are hit by damaging rays.

   Let x denote the unknown parameter of the Poisson distribution. Use the Poisson package to compute the probabilities p_0 of no hits, p_1 of one hits, and p_m of more than one hit in one minute as **functions** of x. For example, define p_0 as a **function** in Maple using

```
> p0:= x -> Poi(0,x);
```

After you define the three functions above, enter the following arrays.

```
> v:= [0,1,2];
> p:= [p0(x),p1(x),p2(x)];
```

(a) Suppose the true value of x is $x := 3.0$. Plot the histogram using

```
> histplot(v,p);
```

(b) Set a random seed **_seed** to a "random" value dredged from your unconscious. Simulate 50 cells using the draw command,

```
> a := [seq(draw(v,p),i=1..50)]:
> transform[tally](a);
```

How many cells of each type do you get from your simulation? To keep things interesting, continue sampling until you get at least one cell of each type.

(c) Compare the results for your simulation with the idealized histogram. For example, the theoretical value for type 0 is

```
> n0 := 50*p0(x);
```

(d) Now pretend that x is your unknown. Use the method of maximum likelihood to analyze your data. Find the likelihood function of the data you obtained in part (b). It is the product of the likelihoods for each of the 50 cells.

Input the likelihood function L as a **function** of the unknown parameter y into Maple.

Additionally, define S as the natural log of L. Again input this as a **function** of y in Maple. Plot $S(y)$ over a reasonable range.

The function $S(y)$ is called the *support* of the unknown parameter y. As we have seen in class and also in this problem, values of the likelihood function can be tiny. By taking the natural log, you can compare likelihoods of different values of the parameter y more easily.

(e) Find the maximum of $S(y)$ by using the following command.

```
> ymax:= fsolve(diff(S(y),y),y);
```

Mark this value on your plot.

(f) Find $S(y)$ for $y = y_{max}$ above. Additionally, find the values of $S(y)$ for $y = 2$, $y = 4$, and the "true" parameter value $y = 3$. Indicate each of them on your graph. Comment on your result

Binomial Package

The binomial package contains the following commands you may want to use today,

1. Probability Mass Function $b(k, n, p)$

   ```
   > b(k,n,p);
   ```

   This command gives the probability of obtaining *exactly* k successes in n trials each with success probability p.

2. Cumulative Distribution Function $B(k, n, p)$

   ```
   > B(k,n,p);
   ```

   This command gives the the probability of obtaining k *or fewer* successes in n trials each with success probability p.

Poisson Package

The Poisson package contains three commands. The first, typed as

```
> Poi(k,Lambda);
```

gives the value of the Poisson probability distribution denoted $p(k; \Lambda)$, the probability of exactly k events with parameter Λ. The second, typed as

```
> Poicum(k1,k2,Lambda);
```

gives the cumulative probability of between k_1 and k_2 events (inclusive). The last, typed

```
> Poiplot(Lambda,N);
```

plots out the Poisson probability distribution with parameter Λ up to a maximum of N (because the theoretical maximum is infinite).