Examples of the Hypergeometric Distribution

The *hypergeometric distribution*, $h(N, k, n, x)$, arises in the following way:

> Suppose we have N balls k of which are red and $N - k$ are blue. We draw a sample, without replacement, of n balls. Let X equal the number of red balls in our sample of size n.

Then

$$\Pr(X = x) = h(N, k, n, x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \quad 0 \le x \le n. \tag{1}$$

Here are two examples of the use of the hypergeometric distribution.

Example 1. (Exercise 5.1.36) A bin of 1000 turnbuckles has an unknown number D of defectives. A sample of 100 turnbuckles has 2 defectives. The *maximum likelihood estimate* for D is the number which gives the highest probability for obtaining the number of defectives observed in a sample. Find that value of D.

Solution: In this example $N = 1000$, $k = D$ (the defective turnbuckles are called red balls), and the sample size is $n = 100$. What value of D makes this the most probable event? We make a table of $h[1000, D, 100, 2]$ for varying values of D. Here's the result from Mathematica:

```
In[15]:= n1=Binomial[1000,100];

In[16]:= pr[d_]:=Binomial[d,2]*Binomial[1000-d,98]/n1;

In[17]:= Table[{d,N[pr[d]]},{d,2,40}]

Out[17]= {{2, 0.00990991}, {3, 0.0268104}, {4, 0.0483501}, {5, 0.0726546},

>    {6, 0.098248}, {7, 0.123986}, {8, 0.149}, {9, 0.172646}, {10, 0.194466},

>    {11, 0.214153}, {12, 0.231519}, {13, 0.246474}, {14, 0.259001},

>    {15, 0.269145}, {16, 0.276991}, {17, 0.282658}, {18, 0.286288},

>    {19, 0.288038}, {20, 0.28807}, {21, 0.286554}, {22, 0.283656},

>    {23, 0.27954}, {24, 0.274364}, {25, 0.268278}, {26, 0.261422},

>    {27, 0.253928}, {28, 0.245918}, {29, 0.237503}, {30, 0.228785},

>    {31, 0.219855}, {32, 0.210795}, {33, 0.201677}, {34, 0.192565},
```

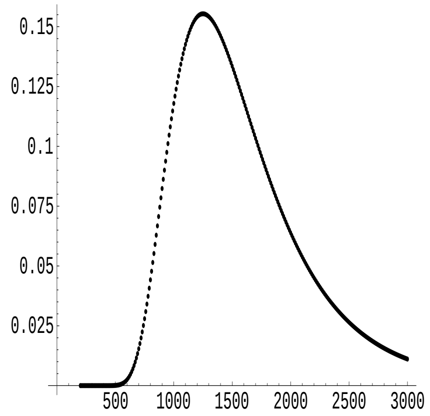

Figure 1: Moose Probabilities on Isle Royale.

```
>     {35, 0.183516}, {36, 0.174578}, {37, 0.165792}, {38, 0.157194},

>     {39, 0.148812}, {40, 0.14067}}
```

So we see that the most likely value of D is 20.

Example 2. (Exercise 5.1.37) There are an unknown number of moose on Isle Royale. To estimate the number of moose, 50 moose are captured and tagged. Six months later 200 moose are captured and it is found that 8 of these are tagged. Estimate the number of moose on Isle Royale[1] from these data.

Solution: In the hypergeometric distribution let N equal the total number of moose. We are told that k of them are tagged. (The tagged moose are the red balls.) A sample of size $k = 200$ is drawn and we are told that 8 are tagged. We ask for what choice of N maximizes the probability $h(N, 50, 200, 8]$, where $h(N, k, n, x)$ is given above in (1). If one makes a table as in the above example one finds that $N = 1250$ maximizes the hypergeometric distribution for the above values of n, k and x. Figure 1 shows a plot of $h(N, 50, 200, 8)$ as a function of N.

[1] `http://www.nps.gov/isro/wolfmoos.htm`