The Negative Binomial and the Hypergeometric distributions

Márton Balázs* and Bálint Tóth*

October 13, 2014

This little write-up is part of important foundations of probability that were left out of the unit Probability 1 due to lack of time and prerequisites. Here we define the Negative Binomial and the Hypergeometric distributions, and show some of their properties.

Definition 1 (Negative Binomal distribution) *We perform independent trials, each succeeding with probability p. Let X be the number of trials needed to see the r^{th} success. Then X has the* Negative Binomial distribution *with parameters p and r ($X \sim Neg.bin(p, r)$).*

Proposition 2 *Let $X \sim Neg.bin(p, r)$, then its mass function is given by*

$$\mathbf{P}\{X = n\} = p(n) = \binom{n-1}{r-1} \cdot p^r \cdot (1-p)^{n-r}, \qquad n = r, r+1, \dots .$$

Proof. The event that the n^{th} trial will see the r^{th} success is equivalent to the fact that out of the first $n-1$ trials exactly $r-1$ succeed (this is a Binomial($n-1$, p) probability), then the n^{th} trial succeeds as well (this has probability p). This explains the above mass function. □

Notice that the Negative Binomial variable X is the (discrete) waiting time for the r^{th} success. As such, it can be broken up into individual waiting times: let Y_1 be the waiting time for the first success, and Y_i the waiting time between the i^{th} and the $i-1^{\text{st}}$ successes, $2 \le i \le r$. Then $X = Y_1 + Y_2 + \cdots + Y_r$, and the variables Y_i are i.i.d. Geometric(p) variables. We use this fact, together with the nice properties of expectations (discussed towards the end of Probability 1) to determine the expectation and variance:

Proposition 3 *Let ($X \sim Neg.bin(p, r)$). Then*

$$\mathbf{E}X = \frac{r}{p}, \qquad \mathbf{Var}X = r \cdot \frac{1-p}{p^2}.$$

Proof.

$$\mathbf{E}X = \mathbf{E}(Y_1 + Y_2 + \cdots + Y_r) = \mathbf{E}Y_1 + \mathbf{E}Y_2 + \cdots + \mathbf{E}Y_r = \frac{1}{p} + \frac{1}{p} + \cdots + \frac{1}{p} = \frac{r}{p}.$$

$$\mathbf{Var}X = \mathbf{Var}(Y_1 + Y_2 + \cdots + Y_r) = \mathbf{Var}Y_1 + \mathbf{Var}Y_2 + \cdots + \mathbf{Var}Y_r = r \cdot \frac{1-p}{p^2}$$

(the second line needed independence of the Y_i's as well). □

Next we consider the Hypergeometric distribution. It has so many parameters that we don't even bother to formally write them out. Also we define it via an example, in fact we are talking about the intersection size of random subsets:

Definition 4 (Hypergeometric distribution) *Out of the N deer in the forest, m have been tagged. Later n of the deer are captured (we assume their numbers do not change), each with equal chance. Let X be the number of tagged deer among those captured. Then X has the* Hypergeometric distribution.

Proposition 5 *The mass function of the Hypergeometric distribution with the above parameters is*

$$\mathbf{P}\{X = i\} = p(i) = \frac{\binom{m}{i} \cdot \binom{N-m}{n-i}}{\binom{N}{n}} = \frac{\binom{n}{i} \cdot \binom{N-n}{m-i}}{\binom{N}{m}}.$$

This mass function gives zero whenever any group of deer (tagged–non-tagged, or captured–non-captured) would be negative, this is made sure by the binomial coefficients of the numerator.

*University of Bristol / Budapest University of Technology and Economics

Proof. To see the validity of this formula, just think about the number of ways we can have $X = i$ in the sample space $\{n$-combinations of N deer$\}$. The equivalence of the two formulas can be seen by expanding the binomial coefficients, or by combinatorics: X counts the elements in the intersection of a random subset of size m and one of size n of a set of cardinality N. (In fact, it is enough to have one of these subsets random.) This explains that the mass function has to be invariant to interchanging the parameters n and m. In other words, we could first capture the n deer, then letting them go we could later tag a random m of all N of them; the distribution of X would not change. □

Proposition 6

$$\mathbf{E}X = \frac{nm}{N}, \qquad \textit{and} \qquad \mathbf{Var}X = \frac{nm}{N} \cdot \Big[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\Big].$$

Proof. We only prove the expectation, the variance can be done along the same lines after a careful examination of the dependence of the following indicator variables on each other.

$$X_i = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ captured deer is tagged,} \\ 0, & \text{if the } i^{\text{th}} \text{ captured deer is not tagged,} \end{cases} \qquad i = 1, 2, \ldots, n.$$

Then $\mathbf{E}X_i = m/N$, and $X = \sum_{i=1}^{n} X_i$:

$$\mathbf{E}X = \mathbf{E}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathbf{E}X_i = \frac{nm}{N}.$$

□

Notice that the above symmetry allows a second look at the problem:

$$Y_j = \begin{cases} 1, & \text{if the } j^{\text{th}} \text{ tagged deer is captured,} \\ 0, & \text{if the } j^{\text{th}} \text{ tagged deer is not captured,} \end{cases} \qquad j = 1, 2, \ldots, m.$$

Then $\mathbf{E}Y_j = n/N$, and $X = \sum_{j=1}^{m} Y_j$:

$$\mathbf{E}X = \mathbf{E}\sum_{j=1}^{m} Y_j = \sum_{j=1}^{m} \mathbf{E}Y_j = \frac{nm}{N}.$$