Statistical Inference I	**Prof. Chin-Tsang Chiang**
Lecture Notes 12	
November 1, 2015	**Scribe: Wei-Chang Lee, Chi-Ning Chou**

Today we talk about the discrete distributions related to Bernoulli distribution.

0.1 Big picture

Bernoulli distribution is a single event with two possible outcome: yes/no. The probability is p for the yes result and $(1-p)$ for the no. Intuitively, we can view a Bernoulli distribution as an indicator to identify whether an event has happened.

What if we want to consider more than one event?

Imagine the following scenario, there is a large population containing N elements and M of them are label as *type-I* and the rest $N-M$ are labeled as *type-II*. Now, as a statistician, we want to draw some inference about the population, but we have only limited access to the population, say k samples. What can we know from the experiment?

Basically, we can categorize the above scenario with two different properties:

- Draw *with replacement* ot *without replacement*.
- Draw *fix number of samples*, or keep drawing *until a certain event happens*?

With these two factors, we can extend Bernoulli distribution into the following three discrete distribution:

	Replacement	Draw	Goal
Hypergeometric	Without	k times	Number of yes
Binomial	With	k times	Number of yes
Negative Binomial	With	Wait until r yes	Number of no

0.2 Hypergeometric distribution

0.2.1 Definition

Hypergeometric distribution describes the probability of the number of *yes* result under k samples **without replacement**. The density function consists of three parameters: (N, M, k) and the pdf is

$$f(x|N,M,k) = \frac{\binom{M}{x}\binom{N-M}{k-x}}{\binom{N}{k}}\mathbf{1}_{(\max(0,k-(N-M)),min(M,k))}(x)$$

Here, we discuss the meaning of each term:

- $\binom{N}{k}$ in the denominator is the number of possible k samples outcome.
- $\binom{M}{x}$ in the numerator is the number of possible combinations of k yes instances.
- $\binom{N-M}{M-x}$ in the numerator is the number of possible combinations of $x-k$ no instances.

0.2.2 Basic properties

Here, we list the mean and variance of hypergeometric distribution and discuss the idea of reparametrize techniques.

- $\mathbb{E}[X|N,M,k] = k\frac{M}{N}$
- $var[X|N,M,k] = k\frac{M}{N}\frac{N-M}{N}\frac{N-k}{N-1}$

In the following, we are going to prove the above results via reparametrize techniques and factorial moment. **Proof**:

- The mean of $X \sim \mathit{Hypergeometric}(N,M,k)$

$$\begin{aligned}
\mathbb{E}[X|N,M,k] &= \sum_{x=\max(0,k-(N-M))}^{\min(M,k)} x\frac{\binom{M}{x}\binom{N-M}{k-x}}{\binom{N}{k}} = \sum_{x=\max(0,k-(N-M))}^{\min(M,k)} M\frac{\binom{M-1}{x-1}\binom{N-M}{k-x}}{\binom{N}{k}} \\
&= \sum_{x=\max(0,k-(N-M))}^{\min(M,k)} M\frac{\binom{M-1}{x-1}\binom{N-(M-1)}{(k-1)-(x-1)}}{\binom{N-1}{k-1}\times\frac{N}{k}} \\
&= k\frac{M}{N}\sum_x \frac{\binom{M-1}{x-1}\binom{N-(M-1)}{(k-1)-(x-1)}}{\binom{N-1}{k-1}} = k\frac{M}{N}
\end{aligned}$$

- The variance of $X \sim \mathit{Hypergeometric}(N,M,k)$

$$var[X|N,M,k] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2$$

  As we know $\mathbb{E}[X]$, it suffices to find $\mathbb{E}[X(X-1)]$. The trick that computing the expectation of $X(X-1)$ instead of that of X^2 is called *factorial moment*, which is computation-friendly when having lots of binomial terms. As a result,

$$\mathbb{E}[X(X-1)] =$$

0.3 Binomial distribution

0.3.1 Definition

The binomial distribution describe the probability of the number of *yes* results with a fixed number of i.i.d. drawing **with replacement**. The density function consists of two parameters: (N,p) and the pdf is

$$f(x|N,p) = \binom{N}{x}p^x(1-p)^{N-x}\mathbf{1}_{0,1,\ldots,N}(x)$$

Note that, the difference between the definitions of hypergeometric distribution and binomial distribution is not only with/without replacement, the underlying mechanism of binomial distribution is not a fix finite sample space as hypergeometric. For example, the number of drawing can be unbounded, or the *yes* probability should not be necessarily a rational number.

0.3.2 Basic properties

Suppose $X \sim \textit{Binomial}(N, p)$, the following is the mean and variance of X:

- $\mathbb{E}[X|N, p] = Np$
- $var[X|N, p] = Np(1 - p)$

0.4 Negative binomial distribution

0.4.1 Definition

The negative binomial distribution describes the probability of the number of *no* instances before certain number of *yes* results in a sequence of i.i.d. drawing. Formally speaking, for a negative binomial distribution with parameters: (p, r) where p is the probability of *yes* and r is the number of *yes* instances we are waiting for, the pdf is

$$f(x|p, r) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x \mathbf{1}_{0,1,\ldots}(x)$$

0.4.2 Basic properties

Suppose $X \sim \textit{Negative Binomial}(p, r)$

- $\mathbb{E}[X|p, r] = \frac{pr}{1-p}$
- $var[X|p, r] = \frac{pr}{(1-p)^2}$
- When $r = 1$, it is called *geometric* distribution.
- The drawing process is memoryless. For example, the distribution of number of *no* will remain the same as we conditioned on the number of *no* instances before.
- As we let $p \to 1$, the *yes* result will tend to happen and some how the distribution will converge to Poisson distribution similarly to binomial distribution. (Detail discussion next time)