Lecture 15

Announcement: For test: Sec 1 sit on left (when facing BB).
Sec 2 sit on Right (- - - - - -).

Last time: - Binomial, geometric, and negative binomial Random variables.

- All involve independent trials.

Today: Hypergeometric Random Variable.

Suppose you have:

- N objects
- m of those are "successes"
- $N-m$ are "failures."

If we choose a random sample of n objects (without replacement!!), then if X is

$X =$ # of successes in a sample of size n,

then X is a hypergeometric random variable

We have

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

where i satisfies
$$n-(N-m) \leq i \leq \min\{n, m\}.$$

— see the textbook for verification of the following facts. Let X be hypergeometric with parameters N, n, m. Then
$$E[X] = n\frac{m}{N}$$
$$Var(X) = n\frac{m}{N}\left(1-\frac{m}{N}\right)\left(1-\frac{n-1}{N-1}\right).$$

Setting $\frac{m}{N} = p$, we can write
$$E[X] = np$$
$$Var(X) = np(1-p)\left(1-\frac{n-1}{N-1}\right).$$

Now, if $N >> n$, $\left(1-\frac{n-1}{N-1}\right) \sim 1$

and so $Var(X) \sim np(1-p)$, $E[X] = np$.

<u>Idea:</u> If n is small compared to N, then $N-n \sim N$ and so even though the sample is without replacement, the probabilities do not change much (between trials) and so $X \sim Bin(n, p)$, where $p = \frac{m}{N}$.

Ex: Suppose an urn contains 10 balls: 5 red, and 5 blue.

Suppose we choose a sample of size 4.

What is the prob. of choosing 2 red balls?

$N=10$, $m=5$, $n=4$. $X=$ # of reds in a sample of 4.

$$P(X=2) = \frac{\binom{5}{2}\binom{10-5}{4-2}}{\binom{10}{4}}$$

$$= \frac{\binom{5}{2}^2}{\binom{10}{4}} = \frac{10}{21}.$$

so $p = \frac{m}{2m} = \frac{1}{2}$.

Suppose m is the # of red balls and $N=2m$. If $n=4$,

Then $P(X=2) = \dfrac{\binom{m}{2}\binom{2m-m}{4-2}}{\binom{2m}{4}}$

$$= \frac{3(m^2-m)}{8m^2-16m+6}$$

Letting $m\to\infty$ gives

$$\lim_{m\to\infty} \frac{3(m^2-m)}{8m^2-16m+6} = \frac{3}{8} = \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2$$

$= P(\mathrm{Bin}(4, \tfrac{1}{2}) = 2)$.

as expected

Zeta Distribution.

A Random var X is said to have the Zeta dist. if

$$P(X=k) = \frac{C}{k^{\alpha+1}} \quad k = 1, 2, 3, \ldots$$

C, α constants.. X is called "Zeta" since

$$1 = \sum_{k=1}^{\infty} \frac{C}{k^{\alpha+1}} = C \sum_{k=1}^{\infty} \frac{1}{k^{\alpha+1}}$$

$$= C \underbrace{\zeta(\alpha+1)}_{\zeta \text{ function.}}$$

If X is a Zeta RV with param α,

Then $$E[X] = \sum_{k=1}^{\infty} \frac{kC}{k^{\alpha+1}} = C \sum_{k=1}^{\infty} \frac{1}{k^{\alpha}}$$

$$= \frac{\zeta(\alpha)}{\zeta(\alpha+1)}.$$

Notice that $\zeta(1) = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$ is not convergent, so we assume $\alpha \geq 2$.

If $\alpha \geq 3$, then

$$Var(X) = E[X^2] - E[X]^2 = \frac{1}{\zeta(\alpha+1)} \sum_{k=1}^{\infty} \frac{k^2}{k^{\alpha+1}} - \frac{\zeta(\alpha)^2}{\zeta(\alpha+1)^2}$$

$$= \frac{\zeta(\alpha-1)}{\zeta(\alpha+1)} - \frac{\zeta(\alpha)^2}{\zeta(\alpha+1)^2}.$$

Expected Value of Sums of RVs

Let $X : S \to \mathbb{R}$ be a DRV.

<u>Prop:</u> $E[X] = \sum_{s \in S} X(s)p(s)$.

<u>Proof:</u> Suppose that $\text{range}(X) = \{x_1, x_2, x_3, \dots\}$.
Let $S_i = \{s \in S : X(s) = x_i\}$. Then

$$\begin{aligned}
E[X] &= \sum_{i=1}^{\infty} x_i P(X = x_i) \\
&= \sum_{i=1}^{\infty} x_i P(S_i) \\
&= \sum_{i=1}^{\infty} x_i \left(\sum_{s \in S_i} P(s) \right) \\
&= \sum_{i=1}^{\infty} \sum_{s \in S_i} x_i P(s). \\
&\left.\begin{aligned}
&= \sum_{i=1}^{\infty} \sum_{s \in S_i} X(s)P(s) \\
&= \sum_{s \in S} X(s)p(s).
\end{aligned}\right\} \text{since } \{S_i : i = 1, 2, \dots\} \text{ is exhaustive, and mutually exclusive.}
\end{aligned}$$

$\square$

<u>Corollary:</u> Let $X_1, \dots, X_n$ be discrete random variables, with $X_j : S \to \mathbb{R}$, $j = 1, 2, \dots, n$. Then $X := X_1 + X_2 + \dots + X_n$ is a DRV, $X : S \to \mathbb{R}$ and

$$E[X] = \sum_{i=1}^{n} E[X_i].$$

Pf: $E[X] = E[X_1 + \dots + X_n]$

$$= \sum_{s \in S} (X_1(s) + \dots + X_n(s))\, p(s)$$

by the previous proposition.

So
$$E[X] = \sum_{s \in S} X_1(s)p(s) + \dots + X_n(s)p(s).$$

$$= \sum_{s \in S} X_1(s)p(s) + \dots + \sum_{s \in S} X_n(s)p(s).$$

$$= E[X_1] + \dots + E[X_n].$$