Note 10

Today, we continue the study some common distributions.

2.4. The Binomial Distribution

In the last class, we verified:

Properties of the Binomial Distribution

The binomial distribution is a distribution of the discrete type satisfying:

Parameters	$n \in \{0, 1, 2, \cdots\}$: number of trials $p \in [0, 1]$: success probability of each trial
Support	$\{0, 1, \cdots, n\}$
PMF	$\binom{n}{x} p^x (1-p)^{n-x}$
Mean	np
Variance	$npq = np(1-p)$
MGF	$((1-p) + pe^t)^n, \qquad (t \in \mathbb{R})$

Before moving to the next section, we make some remarks:

- $b(1, p)$ is exactly the Bernoulli distribution with parameter p.
- Note that both the mean and variance of $b(n, p)$ is proportional to the sample size n. This is not a coincidence, as we will derive this fact using independence in the later part of this course.
- A binomial distribution is a typical example where the PMF has a single "peak". And if n is large, then a probability histogram looks like a bell-shaped curve.
- The CDF $F(x)$ of X having a $b(n, p)$ distribution can be computed as

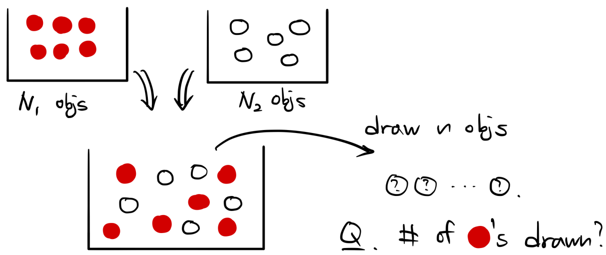
$$F(x) = P(X \leq x) = \sum_{y=0}^{\lfloor x \rfloor} p(y) = \sum_{y=0}^{\lfloor x \rfloor} \binom{n}{y} p^y (1-p)^{n-y}.$$

  (For $x < 0$, the sum is regarded as zero. For $x > n$, note that $\binom{n}{y} = 0$ for $y > n$ and hence the sum is essentially run over $y = 0, \cdots, n$.)

  Although this formula provides a way to compute the exact probabilities related to the binomial distribution, it is often impractical for numerical computation especially when n is large. We will later return to this issue when we learn the normal distribution.

The Hypergeometric Distribution

- Consider a collection of $N = N_1 + N_2$ objects with $\begin{cases} N_1 \text{ of them belonging to the } 1^{\text{st}} \text{ class;} \\ N_2 \text{ of them belonging to the } 2^{\text{nd}} \text{ class;} \end{cases}$



For $n \in \{0, \cdots, N\}$, we choose n objects from these N objects at random and without replacement and write

$$X = [\text{\# of objects selected that belong to the } 1^{\text{st}} \text{ class}].$$

Then for each non-negative integer x,

$$P(X = x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}.$$

The distribution of X is called the *hypergeometric distribution* with parameters N_1, N_2, and n, denoted $HG(N_1, N_2, n)$.

Example

Five cards are selected at random without replacement from a 52-card deck of playing cards. Let

$$X = [\text{\# of face cards (kings, queens, jacks)}].$$

Since there are total 12 face cards in a deck of playing cards, X has $HG(12, 40, 5)$ distribution. In particular, its PMF is

$$p(x) = \frac{\binom{12}{x}\binom{40}{5-x}}{\binom{52}{5}}, \qquad x = 0, 1, 2, 3, 4, 5.$$

- We summarize its properties:

Properties of the Hypergeometric Distribution

The hypergeometric distribution is a distribution of the discrete type satisfying:

Parameters	$N_1 \in \{0, 1, 2, \cdots\}$: number of objects in the 1st class $N_2 \in \{0, 1, 2, \cdots\}$: number of objects in the 2nd class $n \in \{0, \cdots, N = N_1 + N_2\}$: number of objects chosen
Support	$\{\max(0, n - N_2), \cdots, \min(n, N_1)\}$
PMF	$\binom{N_1}{x}\binom{N_2}{n-x}/\binom{N}{n}$
Mean	$n\frac{N_1}{N}$
Variance	$n\frac{N_1}{N}\frac{N_2}{N}\frac{N-n}{N-1}$

Proof. We only prove the formula for the mean, and leave that of the variance as homework. Let X has a $HG(N_1, N_2, n)$ distribution and $p(x)$ be the PMF of X. Then

$$E(X) = \sum_{x \in S} xp(x) = \sum_{x \in S} x \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} = \sum_{\substack{x \in S \\ x \neq 0}} x \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}},$$

where the last step follows by discarding the term corresponding to $x = 0$, which is possible since that term does not contribute to the value of $E(X)$. Now for $x \neq 0$, we have

$$\binom{N_1}{x} = \frac{N_1!}{x!(N_1 - x)!} = \frac{N_1}{x} \frac{(N_1 - 1)!}{(x-1)!(N_1 - x)!} = \frac{N_1}{x}\binom{N_1 - 1}{x - 1}.$$

A similar computation also shows that $\binom{N}{n} = \frac{N}{n}\binom{N-1}{n-1}$. Plugging these to the expectation of X,

$$E(X) = \sum_{\substack{x \in S \\ x \neq 0}} x \frac{\frac{N_1}{x}\binom{N_1-1}{x-1}\binom{N_2}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}} = n\frac{N_1}{N} \boxed{\sum_{\substack{x \in S \\ x \neq 0}} \frac{\binom{N_1-1}{x-1}\binom{N_2}{n-x}}{\binom{N-1}{n-1}}}.$$

But the boxed summation in the last step is exactly the total probability for $HG(N_1 - 1, N_2, n - 1)$ distribution, thus it must sum to 1. Therefore the desired formula is established. $\square$

Remarks)

- If n objects are drawn *with replacement* instead, this the experiment is a binomial experiment with the sample size n and the success probability $p = \frac{N_1}{N}$. The mean and variance of the number of objects belonging to the 1st class is then

$$np = n\frac{N_1}{N} \qquad \text{and} \qquad np(1-p) = n\frac{N_1}{N}\frac{N_2}{N}.$$

Surprisingly, the expectation is the same in both cases. The variance also looks similar, and indeed, if N is large compared to n, then $\frac{N-n}{N-1} \approx 1$ and the variance of $HG(N_1, N_2, n)$ is also close to that of $b(n, N_1/N)$. In fact, we also expect that $b(n, N_1/N)$ provides a good approximation of $HG(N_1, N_2, n)$. This intuitively makes sense, since sampling without replacement from a large pool may be approximated by independent sampling (sampling with replacement).

- The MGF of $HG(N_1, N_2, n)$ is in general not expressible in elementary functions.

2.6. The Negative Binomial Distribution

- In binomial distribution, we are interested in the number of successes from a *fixed* number of independent Bernoulli trials.

- Now we consider: r is a fixed positive integer, and

$$X = [\text{\# of independent Bernoulli trials until exactly } r \text{ successes occur}].$$

Then X is said to have a **negative binomial distribution**.

Properties of the Negative Binomial Distribution

The negative binomial distribution is a distribution of the discrete type satisfying:

Parameters	$r \in \{1, 2, \cdots\}$: number of successes $p \in [0, 1]$: success probability, $q = 1 - p$
Support	$\{r, r+1, r+2, \cdots\}$
PMF	$\binom{x-1}{r-1} p^r (1-p)^{x-r}$
Mean	$\frac{r}{p}$
Variance	$\frac{rq}{p^2}$
MGF	$\left(\frac{pe^t}{1-(1-p)e^t}\right)^r$

Proof. Since we need at least r trials to observe r successes, X can take integer values $\geq r$. So the support of X is $\{r, r+1, r+2, \cdots\}$. Then for each $x = r, r+1, \cdots$,

$$\begin{aligned} P(X = x) &= P\left(\begin{array}{c} r\text{-th success} \\ \text{at the } x\text{-th trial} \end{array} \text{ AND } \begin{array}{c} r-1 \text{ successes} \\ \text{in the first } x-1 \text{ trials} \end{array} \right) \\ &= p \cdot \binom{x-1}{r-1} p^{r-1} (1-p)^{x-r} = \binom{x-1}{r-1} p^r (1-p)^{x-r}. \end{aligned}$$

This is the PMF of X. Before computing the MGF, we enjoy the consequence of this formula. Since the sum of these probabilities is 1,

$$1 = \sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r (1-p)^{x-r}, \qquad \text{which implies} \qquad \sum_{x=r}^{\infty} \binom{x-1}{r-1} q^x = \left(\frac{q}{1-q}\right)^r.$$

It can be shown that this holds not only for $q = 1-p \in [0,1]$ but also for all of $|q| < 1$. Using this, the MGF of X can be computed as:

$$\begin{aligned} M(t) = E[e^{tX}] &= \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r (1-p)^{x-r} \\ &= \left(\frac{p}{1-p}\right)^r \sum_{x=r}^{\infty} \binom{x-1}{r-1} \left((1-p)e^t\right)^x \\ &= \left(\frac{p}{1-p}\right)^r \left(\frac{(1-p)e^t}{1-(1-p)e^t}\right)^r \\ &= \left(\frac{pe^t}{1-(1-p)e^t}\right)^r \end{aligned}$$

Notice that, if $G(t) = \frac{pe^t}{1-(1-p)e^t}$ is the MGF of the geometric distribution with parameter p, then the above result simplifies to $M(t) = [G(t)]^r$. From this relation, we can check that both

$$M'(0) = rG'(0) \qquad \text{and} \qquad M''(0) - M'(0)^2 = r[G''(0) - G'(0)^2]$$

hold. So, both the mean and variance of X is exactly r times the mean and variance of the geometric distribution with parameter p, yielding

$$E(X) = \frac{r}{p}, \qquad \operatorname{Var}(X) = \frac{rq}{p^2}.$$

□

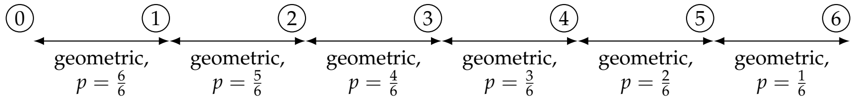
Example (Couple Collector Problem)

Each box of a brand of cereals contains a coupon, and there are 6 different types of coupons. What is the expected number of boxes to be purchased in order to collect all 6 types?

Solution. Let $i \in \{1, \cdots, 6\}$. Then, after $i-1$ different types of coupons have been collected, the probability that each box of serial contains a coupon of uncollected types is

$$p = 1 - \frac{i-1}{6} = \frac{7-i}{6}.$$

So the expected number of boxes of cereals to be purchased until a new type of coupon is collected is the mean of a geometric random variable with this $p = (7-i)/6$, which is $1/p = 6/(7-i)$.



So the expected number of boxes to be purchased in order to collect all 6 types is:

$$\frac{6}{6} + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1} = 14.7$$

□

Example ()

Let X be a random variable having the negative binomial distribution with parameters $r = 2$ and $p = \frac{1}{2}$. Find the value of $P(X \leq 10)$.

Solution. It is a bit boring to directly compute the sum

$$P(X \leq 10) = \sum_{x=2}^{10} \binom{x-1}{1} p^2 (1-p)^{x-2}.$$

Instead, we may employ the following trick: The condition $X \leq 10$ is the same as saying that there are at least 2 successes in the first 10 independent Bernoulli trials with $p = \frac{1}{2}$. So if Y denotes the number of successes in the first first 10 trials, then

$$P(X \leq 10) = P(Y \geq 2) = 1 - P(Y < 2).$$

Since

$$P(Y < 2) = \sum_{y=0}^{1} \binom{10}{y} p^y (1-p)^{10-y} = \frac{11}{1024},$$

it follows that

$$P(X \leq 10) = \frac{1013}{1024} \approx 0.989\ldots$$

□