Note 9

In the last class, we learned various quantities of statistical significance. Today, we finish the discussion on moment generating functions and then move on to studying two common distributions.

2.3. Special Mathematical Expectations

Review.

- Recall that the moment generating function (MGF) of a random variable X (or of its distribution) is the function $M(t)$ defined by

$$M(t) := E(e^{tX}).$$

  It meaningfully conveys information only when M is defined on an open interval (which must include the point $t = 0$ since $M(0) = 1$ always holds).

- We also learned that, if the MGF $M(t)$ of X admits an expansion of the form

$$M(t) = p_1 e^{b_1 t} + p_2 e^{b_2 t} + \cdots,$$

  then X is of the discrete type with the PDF $p(b_1) = p_1$, $p(b_2) = p_2$, etc. This is a particular case of the general statement that MGF uniquely determines the distribution, if exists.
  (*Disclaimer:* Not all MGFs take this form, especially if the distribution is not of the discrete type. We will see such examples in Section 3.)

Although any information on the distribution can be extracted from the MGF, it is particularly easier to compute the moments of X with the MGF.

Proposition

Suppose that the MGF $M(t)$ of X exists near $t = 0$. Then, for each positive integer r, we have

$$E(X^r) = M^{(r)}(0).$$

This is called the ***r*-th moment** of X.

Sketch of Proof. If X is of the discrete type with PMF $p(x)$, then

$$\frac{d^r}{dt^r} M(t) = \frac{d^r}{dt^r} \sum_{x \in S} e^{tx} p(x) = \sum_{x \in S} \frac{d^r}{dt^r} e^{tx} p(x) = \sum_{x \in S} x^r e^{tx} p(x) = E(X^r e^{tX}).$$

(In the second step, interchanging the order of summation and differentiation can be justified by the knowledge from the branch of mathematics called analysis.) Plugging $t = 0$ then proves the desired equality. $\square$

Example (Geometric distribution)

If X has a geometric distribution with parameter $p \in (0, 1)$, i.e., with the PMF

$$p(x) = \begin{cases} q^{x-1}p, & \text{for } x = 1, 2, 3, \cdots; \\ 0, & \text{otherwise,} \end{cases}$$

then the MGF is

$$M(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} q^{x-1} p = \frac{pe^t}{1 - qe^t},$$

provided $qe^t < 1$ (or equivalently, $t < -\ln q$). Then

$$E(X) = M'(X) = \left. \frac{pe^t}{(1 - qe^t)^2} \right|_{t=0} = \frac{1}{p}$$

and

$$E(X^2) = M''(0) = \left. \frac{pe^t(1 + qe^t)}{(1 - qe^t)^3} \right|_{t=0} = \frac{1 + q}{p^2}.$$

So it follows that

$$\operatorname{Var}(X) = E(X^2) - E(X)^2 = \frac{q}{p^2}.$$

2.4. The Binomial Distribution

Bernoulli distribution

- A **Bernoulli experiment** or **Bernoulli trial**: an experiment with two possible outcomes. (Such as YES/NO, SUCCESS/FAIL, etc.)

- Given a Bernoulli trial, let X be the random variable given by

$$X(\text{success}) = 1 \qquad \text{and} \qquad X(\text{failure}) = 0.$$

  Write $p = P(X = 1)$ for the probability of success and $q = 1 - p$. Then the distribution of X is called the **Bernoulli distribution** with parameter p.

- We collect the properties of the Bernoulli distribution:

Properties of the Bernoulli distribution

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

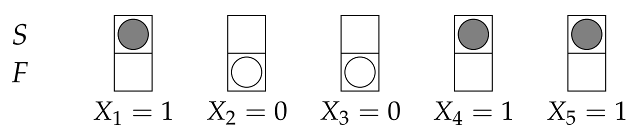
Parameter	$p \in [0,1]$: success probability
Support	$\{0,1\}$
PMF	$p^x(1-p)^{1-x}$
Mean	p
Variance	$pq = p(1-p)$
MGF	$(1-p)+pe^t, \qquad (t \in \mathbb{R})$

Indeed,

- The formula for the PMF f conveniently encodes $f(1) = p$ and $f(0) = 1 - p = q$.
- The mean is computed as $E(X) = 0 \cdot (1-p) + 1 \cdot p = p$.
- The variance is computed as $E[(X-p)^2] = (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p = p(1-p)$.

Binomial distribution

- If n independent Bernoulli trials are performed, an n-tuple of zeros and ones will be observed. This observed tuple is often called a **random sample** of size n from a Bernoulli distribution. (Later we will also consider random samples from other distributions.) Given a random sample from a Bernoulli distribution, we are often interested in the total number of successes.



$\Rightarrow \quad X = \sum_{i=1}^{5} X_i = 3$

- A **binomial experiment**:
  1. A Bernoulli experiment is performed n times.
  2. The trials are independent.
  3. Each trial has the same probability p of success.
  4. The outcome X is the number of successes in the n trials.

  The distribution of X is called the **binomial distribution** with parameters n and p and denoted by $b(n,p)$.

- Let X has a $b(n,p)$ distribution. Then the possible values of X are $0, \cdots, n$. Moreover, for each possible value $x = 0, \cdots, n$ of the number of successes,

$$[\text{\# of ways of selecting } x \text{ positions for success}] = \binom{n}{x},$$
$$[\text{prob. of each of these ways}] = p^x(1-p)^{n-x}.$$

So the PMF of X satisfies

$$P(X = x) = \binom{n}{x} \cdot p^x(1-p)^{n-x}.$$

Using this, we prove that:

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	$((1-p) + pe^t)^n, \qquad (t \in \mathbb{R})$

Proof. We first compute the MGF of $b(n, p)$ distribution:

$$M(t) = \sum_{x=0} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0} \binom{n}{x} (e^t p)^x (1-p)^{n-x} = (1 - p + pe^t)^n.$$

Then by the chain rule,

$$E(X) = M'(0) = \Big[n(1 - p + pe^t)^{n-1}(pe^t) \Big]_{t=0} = np.$$

Similarly, by the product rule and the chain rule,

$$\begin{aligned} E(X^2) = M''(0) &= \Big[n(1 - p + pe^t)^{n-1}(pe^t) + n(n-1)(1 - p + pe^t)^{n-2}(pe^t)^2 \Big]_{t=0} \\ &= np + n(n-1)p^2 = np(1-p) + (np)^2. \end{aligned}$$

So it follows that

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = np(1-p).$$

$\square$

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.
- The CDF $F(x)$ of X having a $b(n, p)$ distribution can be computed as

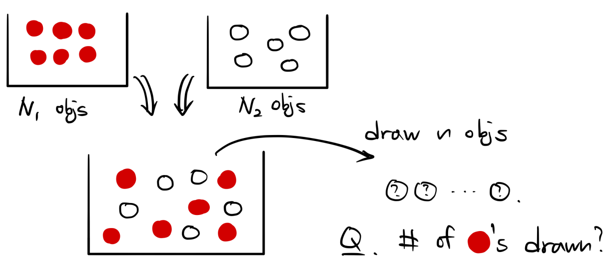
$$F(x) = P(X \le x) = \sum_{y=0}^{\lfloor x \rfloor} p(y) = \sum_{y=0}^{\lfloor x \rfloor} \binom{n}{y} p^y (1-p)^{n-y}.$$

(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 $HG(N_1, N_2, n)$.

- 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 1^{st} class $N_2 \in \{0, 1, 2, \cdots\}$: number of objects in the 2^{nd} 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

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

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. □

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 $HG(N_1, N_2, n)$ and $b(n, N_1/N)$ are close to each other in an appropriate sense, correctly reflecting our intuition that 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.