Chapter 3. Some Special Distributions

Section 3.1. The Binomial and Related Distributions

Note. In this section, we consider Bernoulli trials, binomial distributions, multinomial distributions, geometric distributions, and hypergeometric distributions.

Definition. A *Bernoulli experiment* is a random experiment the outcome of which can be classified in exactly one of two mutually exclusive exhaustive ways (usually termed a "success" and a "failure"). A sequence of *Bernoulli trials* occurs when a Bernoulli experiment is performed several independent times so that the probability of success, say p, remains the same from trial to trial.

Note. The Bernoulli family of the late 17th and early 18th century included a number of influential mathematicians (including his brother Johann and his nephew Daniel). Influenced by Leibniz's work, Jacob Bernoulli studies series and differential equations (the "Bernoulli [differential] equation" describes the isochrone problem; a solution to the differential equation gives the curve along which a particle will descend under gravity from any point to the bottom in the same amount of time, no matter the starting point). He is therefore recognized as contributing to mechanics and the calculus of variations. His contributions to probability include his Law of Large Numbers published in 1689, which implies that as an experiment is repeated a large number of times, then the relative frequency with which an event occurs approaches the probability of the event. In 1713, Jacob Bernoulli's *Ars Conjectandi* ("The Art of Conjecturing") was published posthumously by his nephew, Niklaus. This work included information on combinatorics related to counting and applications to games of chance concerning cards or dice. This is where the idea of a Bernoulli trail first appears. This historical information is based on the Wikipedia page on Bernoulli, the Wikipedia page on *Ars Conjectandi*, and the Saint Andrews MacTutor page on Jacob Bernoulli.

Jacob (Jacques) Bernoulli, January 6, 1655–August 16, 1705

Image from the Wikipedia page on Bernoulli

Note. Define random variable X associated with a Bernoullis trial as:

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

The probability mass function of X is

$$p(x) = p^x(1-p)^{1-x} \text{ where } x \in \{0, 1\}$$

and p is the probability of success. The expected value of X is

$$\mu = E[X] = \sum_x xp(x) = (0)(1-p) + (1)(p) = p,$$

and the variance is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 p(x) = (0-p)^2(1-p) + (1-p)^2(p)$$
$$= (p^2 + (1-p)p)(1-p) = p(1-p).$$

So the standard deviation is $\sigma = \sqrt{p(1-p)}$.

Definition. With sample space $\mathcal{C} = \{\text{failure}, \text{success}\}$ and random variable X satisfying $X(\text{success}) = 1$ and $X(\text{failure}) = 0$, the probability mass function of X, $p(x) = p^x(1-p)^{1-x}$ where $x \in \{0,1\}$ is the *Bernoulli distribution.*

Note. Consider a sequence of n independent Bernoulli trials where the probability of success remains p for each trial. Let X be the random variable equal to the number of observed successes in the n Bernoulli trials. If we perform the trials successively and x successes occur (so $x \in \{0, 1, \ldots, n\}$), then the number of ways of selecting the x positions of the successes is $\binom{n}{x} = \dfrac{n!}{x!(n-x)!}$. The probability of x successes and $n-x$ failures (in some particular order) by the multiplicative rule is $p^x(1-p)^{n-x}$. Now x successes and $n-x$ failures can happen in $\binom{n}{x}$ different ways, so the probability of x successes and $n-x$ failures in not particular order by finite additivity is $\binom{n}{x}p^x(1-p)^{n-x}$. So the probability mass function of X is

$$p(x) = \begin{cases} \binom{n}{x}p^x(1-p)^{n-x} & \text{for } x \in \{0,1,\ldots,n\} \\ 0 & \text{otherwise.} \end{cases}$$

By the Binomial Theorem $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$ and so

$$\sum_x p(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = ((p)+(1-p))^n = 1,$$

as needed.

Definition. A random variable X that has a probability mass function of the form

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x \in \{0, 1, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}$$

is said to have a *binomial distribution* and any such $p(x)$ is a *binomial probability mass function*. A binomial distribution is denoted $b(n,p)$. The constants n and p are the *parameters* of the binomial distribution.

Note 3.1.A. The moment generating function of a binomial distribution is

$$\begin{aligned} M(t) &= E[e^{tX}] = \sum_x e^{tx} p(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} = (pe^t + (1-p))^n \text{ by the Binomial Theorem} \\ &= (1-p+pe^t)^n \end{aligned}$$

and this is valid for all t. Notice that $M'(t) = n(1-p+pe^t)^{n-1}$ and

$$M''(t) = n(n-1)(1-p+pe^t)^{n-2}(pe^t)^2 + n(1-p+pe^t)^{n-1}pe^t$$

$$= n(1-p+pe^t)n - 2pe^t((n-1)pe^t + (1-p+pe^t)) = n(1-p+pe^t)^{n-2}pe^t(npe^t+1-p).$$

In Note 1.9.B we see that $\mu = M'(0)$ and $\text{Var}(X) = \sigma^2 = M''(0) = (M'(0))^2$. Hence the mean and variance of the binomial distribution are

$$\mu = M'(0) = n(1 - p + pe^{(0)})^{n-1}pe^{(0)} = np,$$

and

$$\sigma^2 = M''(0) - (M'(0))^2 = n(1 - p + pe^{(0)})^{n-2}pe^{(0)}(npe^{(0)} + 1 - p) - (np)^2$$

$$= np(np + 1 - p) - (np)^2 = np(1 - p).$$

Example 3.1.2. If random variable X has moment generating function $M(t) = \left(\frac{2}{3} + \frac{1}{3}e^t\right)^5$, then X has a binomial distribution since

$$M(t) = \left(\frac{2}{3} + \frac{1}{3}e^t\right)^5 = (1 - p + pe^t)^5,$$

where $p = 1/3$ and $n = 5$. The mean and variance of this binomial distribution are $\mu = np = 5/3$ and $\sigma^2 = np(1 - p) = 10/9$. The probability mass function of X is (by definition):

$$p(x) = \begin{cases} \binom{5}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{5-x} & \text{for } x \in \{0, 1, 2, 3, 4, 5\} \\ 0 & \text{otherwise.} \end{cases}$$

Example 3.1.4. We now establish one form of the Weak Law of Large Numbers. Let random variable Y be the number of successes in n independent repetitions of a Bernoulli experiment with probability p of a success, so that Y has the distribution $b(n, p)$. The ratio Y/n is the *relative frequency* of success. We'll show that the relative frequency of a success, Y/n, approaches p as $n \to \infty$. By Chebyshev's Inequality (Theorem 1.10.3; see Note 1.10.A) we have for all $\varepsilon > 0$ that

$$P\left(\left|\frac{Y}{n} - p\right| \geq \varepsilon\right) \leq \frac{\text{Var}(Y/n)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2},$$

since the mean of Y is $pn/n = p$ and $\text{Var}(Y/n) = p(1-p)/(n\varepsilon^2)$ by Exercise 3.1.3. So for given $\varepsilon > 0$, $\lim_{n\to\infty} \frac{p(1-p)}{n\varepsilon^2} = 0$ and so (by the Sandwich Theorem, say)

$$\lim_{n\to\infty} P\left(\left|\frac{Y}{n} - p\right| \geq \varepsilon\right) = 0 \text{ so that } \lim_{n\to\infty} P\left(\left|\frac{Y}{n} - p\right| < \varepsilon\right) = 1.$$

Since $\varepsilon > 0$ is arbitrary, we see that $Y/n \to p$ as $n \to \infty$ (with probability 1; in some settings, such behavior is referred to as "almost surely" so that we would say $Y/n \to p$ as $n \to \infty$ almost surely—see my online notes for Graph Theory 2 [MATH 5450] on 13.3. Variance and notice that Chebyshev's Inequality in a special setting is also stated here as Theorem 13.7, and an application in graph theory is given in Note 13.3.A). This example will be generalized in Section 5.1. Convergence in Probability, where we will prove the Weak Law of Large Numbers in (see Theorem 5.1.1).

Note. We now show that a sum of independent binomial random variables (each with the same probability of success) is itself a binomial random variable.

Theorem 3.1.1. Let $X_1, X_2, \ldots, X_m$ be independent random variables such that X_i has the binomial $b(n_i, p)$ distribution for $i \in \{1, 2, \ldots, m\}$. Let $Y = \sum_{i=1}^m X_i$. Then Y has a binomial $b\left(\sum_{i=1}^m n_i, p\right)$ distribution.

Note. We again consider a sequence of independent Bernoulli trials with constant probability p of success. But now, for fixed positive integer r, we let Y be the random variable giving the total number of failures before the rth success (so random variable $Y + r$ is the number of trials that are needed to produce r successes). For the probability mass function of Y, we compute the probability of $r - 1$ successes in the first $y + r - 1$ trials (in any order), followed be a success in the $(y + r)$th trial. This probability is

$$\binom{y+r-1}{r-1} p^{r-1}(1-p)^y p = \binom{y+r-1}{r-1} p^r (1-p)^y.$$

We now name this distribution.

Definition. A distribution with probability mass function

$$p_Y(y) = \begin{cases} \binom{y+r-1}{r-1} p^r (1-p)^y & \text{for } y \in \{0, 1, 2, \ldots\} \\ 0 & \text{otherwise} \end{cases}$$

is a *negative binomial distribution* and any such $p_Y(y)$ is a negative binomial probability mass function.

Note. In Exercise 3.1.17 it is to be shown that the moment generating function for the negative binomial distribution is $M(t) = p^r(1-(1-p)e^t)^{-r}$ for $r < -\log(1-p)$.

Note. If we take $r = 1$ in the negative binomial distribution, then the relevant random variable Y gives the number of failures before the first success. This gives us another distribution.

Definition. A distribution with probability mass function

$$p_Y(y) = \begin{cases} p(1-p)^y & \text{for } y \in \{0, 1, 2, \ldots\} \\ 0 & \text{otherwise} \end{cases}$$

is a *geometric distribution*. The moment generating function is $M(t) = p(1-(1-p)e^t)^{-1}$.

Note. If we consider a Bernoulli trial with the probability of a success as p (and so the probability of a failure is $1-p$), then with Y as the random variable representing the number of failures before a success, then Y has the geometric distribution.

Note. To generalize the Binomial Distribution to the Multinomial Distribution, consider an experiment repeated n times where at each performance of the experiment, there is only one outcome but tat outcome falls into k categories (instead of two , as is the case in a Binomial Distribution, denoted $C_1, C_2, \ldots, C_k$. For $i \in \{1, 2, \ldots, \}$ let p_i be the probability that the outcome of the experiment is an element of C_i and assume that p_1 remains constant throughout the n independent repetitions. Define the random variable X_i as the number of outcomes that are elements of C_i for $i \in \{1, 2, \ldots k-1\}$ (notice that the number of outcomes that are elements of C_n, X_n, is determined by $X_1, X_2, \ldots, X_{n-1}$ and $X_n = n - X_1 - X_2 - \cdots - X_{n-1}$; more generally, for any $1 \leq k \leq n$ we have that X_k is determined by $X_1, X_2, \ldots X_{k-1}, X_{k+1}, \ldots, X_n$). Now the number of different arrangements of x_1 elements of C_1, x_2 elements of C_2, $\ldots$, and x_k elements of C_k

is (by the Multiplicative Rule , Rule 1 of Section 1.3):

$$\binom{n}{x_1}\binom{n-x_1}{x_2}\cdots\binom{n-x_1-x_2-\cdots-x_{k-2}}{x_{k-1}}$$

$$= \frac{n!}{(n-x_1)!x_1!}\frac{(n-x_1)!}{(n-x_1-x_2)!x_2!}\cdots\frac{(n-x_1-x_2-\cdots-x_{k-2})!}{(n-x_1-x_2-\cdots-x_{k-1})!x_{k-1}!}$$

$$= \frac{n!}{x_1!x_2!\cdots x_{k-2}!(n-x_1-x_2\cdots-x_{k-1})!x_k!}$$

$$= \frac{n!}{x_1!x_2!\cdots x_{k-2}!x_{k-1}!x_k!} \text{ since } n-x_1-x_2-\cdots-x_{k-1}=x_k,$$

and the probability of such an event is $p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}$. So the joint probability mass function of $(X_1, X_2, \ldots, X_{k-1})$ is

$$P(X_1=x_1, X_2=x_2, \ldots, X_{k-1}=x_{k-1}) = \frac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}$$

where $x_1, x_2, \ldots, x_{k-1}$ are nonnegative integers, $x_k = n - x_1 - x_2 - \cdots x_{k-1}$, and $p_k = 1 - \sum_{j=1}^{k-1} p_j$.

Definition. Random variables $X_1, X_2, \ldots, X_{k-1}$ have a *multinomial distribution* with with parameter n and $p_1, p_2, \ldots, p_{k-1}$ where the probability that X_i is an element of C_i is p_i for $i \in \{1, 2, \ldots, k-1\}$, if the joint moment generating functions is

$$M(t_1, t_2, \ldots, t_{k-1}) = \sum\sum\cdots\sum \frac{n!}{x_1!x_2!\cdots x_K!}(p_1e^{t_1})x_1(p_2e^{t_2})x_2\cdots(p_{k-1}e^{t_{k-1}})x_{k-1}p_k^{x_k},$$

where the multiple sum is taken over all nonnegative integers such that $x_1 + x_2 + \cdots + x_{k-1} \leq n$.

Note. In a multinomial distribution, the marginal distribution of X_i is the same as a binomial distribution since it has moment generating function

$$M(0,0,\ldots,0,t,0,\ldots,0) = (p_i e^{t_i} + (1-p_i))^n.$$

Note/Definition. The marginal generating function of (X_i, X_j), when $i < j$, is

$$M(0,0,\ldots,0,t_i,0,\ldots,0,t_j,0,\ldots,0) = (p_i e^{t_i} + p_j e^{t_j} + (1-p_i-p_j))^n$$

is a *trinomial distribution* with parameters n, p_i, and p_j. Think of these categories as a "success of type i," "a success of type j," and a "failure."

Note/Definition. Another distribution from the multinomial setting is the *conditional distribution* of X_2 given X_1. We have (X_1, X_2) with the multinormal distribution (namely, the trinomial distribution) and parameters n, p_1, and p_2, and we have X_1 with a binomial distribution with parameters n and p_1. By Note 2.3.A, the conditional probability mass function of $X_2 \mid X_1$ is

$$p_{X_2|X_1}(x_2 \mid x_1) = \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_1}(x_1)}.$$

Now for the trinomial distribution,

$$\begin{aligned} p_{X_1,|X_2}(x_2 \mid x_1) &= \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \\ &= \frac{n!}{x_1!x_2!(n-(x_1+x_2))!} p_1^{x_1} p_2^{x_2} (1-(p_1+p_2))^{n-(x_1+x_2)} \end{aligned}$$

since $n = x_1 + x_2 + x_3$ and $p_1 + p_2 + p_3 = 1$, and for the binomial distribution

$$p_{X_1}(x_1) = \binom{n}{x_1} p_1^{x_1}(1-p_1)^{n-x_1} = \frac{n!}{x_1!(n-x_1)!} p_1^{x_1}(1-p_1)^{n-x_1},$$

so

$$p_{X_1|X_1}(x_1,x_2) = \frac{n!p_1^{x_1}p_2^{x_2}(1-(p_1+p_2))^{n-(x_1+x_2)}}{x_1!x_2!(n-(x_1+x_2))!}\frac{x_1!(n-x_1)!}{n!p_1^{x_1}(1-p_1)^{n-x_1}}$$
$$= \frac{(n-x_1)!}{x_2!(n-(x_1+x_2))!}\frac{p_2^{x_2}}{(1-p_1)^{x_2}}\frac{((1-p_1)-p_2)^{n-x_1-x_2}}{(1-p)^{n-x_1-x_2}}$$
$$= \binom{n-x_1}{x_2}\left(\frac{p_2}{1-p_1}\right)^{x_2}\left(1-\frac{p_2}{1-p_1}\right)^{n-x_1-x_2}$$

for $0 \le x_2 \le n - x_1$. Notice that this is the same as a binomial distribution with parameters $n-x_1$ and $p_2/(1-p_1)$. Since the mean of binomial distribution $b(n,p)$ is np by Note 3.1.A, then the mean of the conditional distribution is $(n-X_1)p_2/(1-p_1)$.

Definition. Let N, D, n be positive integers. The random variable X with probability mass function

$$p(x) = \frac{\binom{N-D}{n-x}\binom{D}{x}}{\binom{N}{n}} \text{ for } x \in \{0,1,\ldots,n\}$$

has a *hypergeometric distribution* with parameters (N, D, n). We take the binomial coefficients as 0 when the top value is less than the bottom value.

Note. Consider a population of size N which contains D "defectives." Let X denote the number of defectives in a sample of size n. If a sample is taken with replacement then X has a binomial distribution with parameters n and $p = D/N$. If the sample is taken *without* replacement then there are $\binom{N}{n}$ possible samples and there are $\binom{N-D}{n-x}\binom{D}{x}$ samples that have x defectives and $n-x$ nondefectives. So this second situation is an example of a hypergeometric distribution. We introduced this idea in Example 1.6.2.

Note. The mean of random variable X with hypergeometric distribution is

$$E(X) = \sum_{x=0}^{n} xp(x) = \sum_{x=1}^{n} x \binom{N-D}{n-x}\binom{D}{x}\bigg/\binom{N}{n}$$

$$= \sum_{x=1}^{n} x \frac{\binom{N-D}{n-x} D! n! (N-n)!}{x!(D-x)!N!} = \sum_{x=1}^{n} x \frac{\binom{N-D}{n-x} D(D-1)!/(x(x-1)!(D-x)!)}{N(N-1)!/((N-n)!n!(n-1)!}$$

$$= n\frac{D}{N} \sum_{x=1}^{n} \binom{(N-1)-(D-1)}{(n-1)-(x-1)}\binom{D-1}{x-1}\binom{N-1}{n-1}^{-1} = n\frac{D}{N}.$$

We eliminate the sum in the last step because it represents the sum of probabilities over all possible values of $x-1$ in the hypergeometric distribution with parameters $(N-1, D-1, n-1)$, so that the sum is 1. In Exercise 3.1.31 it is to be shown that the variance of random variable X with hypergeometric distribution with parameters (N, D, n) is

$$\text{Var}(X) = n\frac{D}{N}\frac{N-D}{N}\frac{N-n}{N-1}.$$

Revised: 3/27/2021