1 Some Discrete Probability Distributions

1.1 Introduction and Motivation

- A handful of important probability distribution describe many of the discrete random variables encountered in practice.
- **Binomial distribution**: test the effectiveness of a new drug.
- **Hypergeometric distribution**: test the number of defective items from a batch of production.
- **Negative binomial distribution (Geometric distribution)**: the number of trial on which the first success occurs.
- **Poisson distribution**: the number of outcomes occurring during a given time interval or in a specified region.

1.2 Discrete Uniform Distribution

- **Discrete Uniform Distribution:** If the random variable X assumes the values $x_1, x_2, \ldots x_n$, with equal probabilities, then the discrete uniform distribution (probability mass function) is given by

$$f(x;k) = \frac{1}{k},\ x = x_1, x_2, \ldots x_k$$

- **Example 5.1**: When a light bulb is selected at random from a box that contains a 40-watt bulb, a 60-watt bulb, a 75-watt bulb, and a 100-watt bulb, each element of the sample space $S = \{40, 60, 75, 100\}$ occurs with probability 1/4.
- Therefore, we have a uniform distribution, with

$$f(x;4) = \frac{1}{4},\ x = 40, 60, 75, 100$$

- **Example 5.2**: When a die is tossed, each element of the sample space $S = \{1, 2, 3, 4, 5, 6\}$ occurs with probability 1/6.
- **Example 5.2**:Cont.
  - Therefore, we have a uniform distribution, with

$$f(x;6) = \frac{1}{6},\ x = 1, 2, 3, 4, 5, 6$$

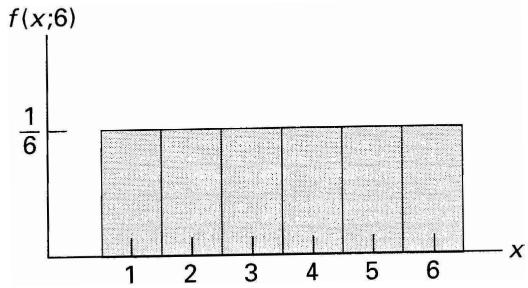

Figure 1: Histogram for the tossing of a die.

- **Theorem 5.1:**:

  The mean and variance of the discrete uniform distribution f(x; k) are

$$E(X) = \mu = \frac{1}{k}\sum_{i=1}^{k} x_i, \quad \sigma_x^2 = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2$$

- **Example 5.3**: Referring to Example 5.2 (tossing a die), we find that

$$\mu = 3.5, \quad \sigma^2 = 2.92$$

1.3 Binomial and Multinomial Distribution

- **Bernoulli Random Variable**: Suppose that we have a random variable X that has just two outcomes (e.g., success/failure) with probability p and $1 - p = q$, respectively.

- We call the random variable Bernoulli random variable. By representing a random variable X to be the number of successes,

$$X = \begin{cases} 1, & \textit{with probability } p, \\ 0, & \textit{with probability } 1 - p, \end{cases}$$

- An experiment that involve the Bernoulli random variable is called the Bernoulli experiment or Bernoulli trial.

- **The Bernoulli Probability Distribution**: The probability distribution of X is given by

$$P(X = x) = p(x) = p^x q^{1-x}, \ x = 0, 1$$

  where p is called the parameter of Bernoulli probability distribution.

- Since X can be only 0 and 1;

$$E(X) = 1 * p + 0 * q = p, \sigma_X^2 = E(X^2) - [E(X)]^2 = p - p^2 = pq.$$

- **Example 5.2**: Generalization of the model of tossing a coin: $S = \{Success, Failure\}$ with $P(Success) = p$ and $P(Failure) = 1 - p = q$.

- An experiment often consists of repeated trials, each with two possible outcomes that may be labeled success or failure.

- The properties of the Bernoulli Process
  - The experiment consists of n repeated trials.
  - Each trial is called **Bernoulli trial**. Each trial results in an outcome that may be classified as a success or a failure.
  - We may choose to define either outcome as a success.
  - The probability of success, denoted by p, remains constant from trial to trial.
  - The repeated trials are **independent**.

- **Binomial Distribution.** A Bernoulli trial can result in a success with probability p and a failure with probability $q = 1 - p$.

- Then the probability distribution of the binomial random variable X, the number of successes in n independent trials, is

$$\boxed{b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \ldots, n}$$

- where

$$\binom{n}{x}$$

: is the number of sample points that have x successes.

$$B(r; n, p) = \sum_{x=0}^{r} b(x; n, p)$$

: is the binomial sums.

TABLE A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x; n, p)$

n	r	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000				
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000			
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000		
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001		
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0094	0.0007	0.0000	
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0951	0.0152	0.0008	
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.000
	8		0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.000
	9		0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.002
	10		1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11			1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12				1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13					1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14						1.0000	0.9995	0.9953	0.9648	0.794
	15							1.0000	1.0000	1.0000	1.000
16	0	0.1853	0.0281	0.0100	0.0033	0.0003	0.0000				
	1	0.5147	0.1407	0.0635	0.0261	0.0033	0.0003	0.0000			
	2	0.7892	0.3518	0.1971	0.0994	0.0183	0.0021	0.0001			
	3	0.9316	0.5981	0.4050	0.2459	0.0651	0.0106	0.0009	0.0000		
	4	0.9830	0.7982	0.6302	0.4499	0.1666	0.0384	0.0049	0.0003		
	5	0.9967	0.9183	0.8103	0.6598	0.3288	0.1051	0.0191	0.0016	0.0000	
	6	0.9995	0.9733	0.9204	0.8247	0.5272	0.2272	0.0583	0.0071	0.0002	

Figure 2: Binomial Probability Sums $B(r; n, p) = \sum_{x=0}^{r} b(x; n, p)$.

- A random variable with this probability distribution is said to be binomially distributed.
- The values of the binomial sums can be found in Table A.1

- **Example**: Suppose a professional basket player tries 5 free throws. The player is known to make 80% successful rate.
- Let X be the number of free throws he will make, then $X \sim b(n = 5, p = 0.8)$. Since there are $\binom{5}{4} = 5$ cases of making 4 successes

Table 1: Possible cases of 4 successes (o) and 1 miss (x) among 5 trials with $p = 0.8, q = 0.2$.

Trial	Possible Event	$p(x)$	Probability
1	oooox	ppppq	$(0.8)^4(0.2)^1 = 0.08192$
2	oooxo	pppqp	$(0.8)^4(0.2)^1 = 0.08192$
3	ooxoo	ppqpp	$(0.8)^4(0.2)^1 = 0.08192$
4	oxooo	pqppp	$(0.8)^4(0.2)^1 = 0.08192$
5	xoooo	qpppp	$(0.8)^4(0.2)^1 = 0.08192$

among 5 trials,

$$P(X = 4) = 5(0.8)^4(0.2)^1 = 0.4096.$$

- **Example 5.4**: The probability that a certain kind of component will survive a given shock test is $\frac{3}{4}$.
- Find the probability that exactly 2 of the next 4 components tested survive.

$$b(x; n, p) =$$

$$b(2; 4, \frac{3}{4}) =$$

$$\binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 =$$

$$\frac{4!}{2!2!} * \frac{3^2}{4^4} = \frac{27}{128}$$

- Where Does the Name Binomial Come From? Binomial distribution corresponds to the binomial expansion of $(q + p)$

$$\begin{aligned}(q+p)^n &= \binom{n}{0} p^0 q^n + \binom{n}{1} p q^{n-1} + \binom{n}{2} p^2 q^{n-2} + \ldots + \binom{n}{n} p^n q^{n-n} \\ &= b(0; n, p) + b(1; n, p) + b(2; n, p) + \ldots + b(n; n, p) \\ &= 1\end{aligned}$$

since $p + q = 1$

- **Example 5.5**: The probability that a patient recovers from a rare blood disease is 0.4.
- If 15 people are known to have contracted this disease, what is the probability that

  i at least 10 survive,

$$P(X \geq 10) = 1 - P(x < 10) = 1 - \sum_{x=0}^{9} b(x; 15, 0.4) = 1 - 0.9662 = 0.0338$$

ii from 3 to 8 survive, and

$$P(3 \leq X \leq 8) = \sum_{x=3}^{8} b(x;15,0.4) = \sum_{x=0}^{8} b(x;15,0.4) - \sum_{x=0}^{2} b(x;15,0.4)$$
$$= 0.9050 - 0.0271 = 0.8779$$

iii exactly 5 survive?

$$P(X=5) = b(5;15,0.4) = \sum_{x=0}^{5} b(x;15,0.4) - \sum_{x=0}^{4} b(x;15,0.4)$$
$$= 0.4032 - 0.2173 = 0.1859$$

- **Example 5.6**: A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates that the defective rate of the device is 3%

  i The inspector of the retailer randomly picks 20 items from a shipment. What is the probability that there will be at least one defective item among these 20?
  Denote by X the number of defective devices among the 20; $b(x;20,0.03)$

  $$P(X \geq 1) = 1 - P(X=0) = 1 - b(0;20,0.03) = 1 - 0.03^0 0.97^{20-0} = 0.4562$$

  ii Suppose that the retailer receives 10 shipments in a month and the inspector randomly tests 20 devices per shipment. What is the probability that there will be 3 shipments containing at least one defective device?
  Denote by Y the number of shipments containing at least one defective item; $b(y;10,0.4562)$

  $$P(Y=3) = \binom{10}{3} 0.4562^3 (1-0.4562)^{10-3} = 0.1602$$

- **Theorem 5.2**:

  The mean and variance of the binomial distribution $b(x;n,p)$ are

  $$\mu = np \textit{ and } \sigma^2 = npq$$

- **Example 5.7**: Find the mean and variance of the binomial random variable of Example 5.5 ($n = 15, p = 0.4$), and then use Chebyshev's theorem to interpret the interval $\mu \pm 2\sigma$

- Solution: Example 5.5 was a binomial experiment with $n = 15$ and $p = 0.4$

$$\mu = np = 15 * 0.4 = 6$$

$$\sigma^2 = npq = 15 * 0.4 * 0.6 = 3.6 \Rightarrow \sigma = 1.897$$

The interval

$$\mu \pm 2\sigma = 6 \pm 2 * 1.897 \Rightarrow 2.206 \; to \; 9.794$$

has a probability of at least $\frac{3}{4}$.

- **Example 5.8**: It is conjectured that an impurity exists in 30% of all drinking wells in a certain rural community.

- In order to gain some insight on this problem, it is determined that some tests should be made. It is too expensive to test all of the many wells in the area, so 10 were randomly selected for testing.

  i Using the binomial distribution, what is the probability that exactly three wells have the impurity assuming that the conjecture is correct?

$$b(x; 10, 0.3) = P(X = 3) = B(3; 10, 0.3) - B(2; 10, 0.3)$$
$$= 0.6496 - 0.3828 = 0.2668$$

  ii What is the probability that more than three wells are impure?

$$P(X > 3) = 1 - B(3; 10, 0.3) = 1 - 0.6496 = 0.3504$$

- **Example 5.9**: Consider the situation of Example 5.8. The "30% are impure" is merely a conjecture put forth by the area water board.

- Suppose 10 wells are randomly selected and 6 are found to contain the impurity. What does this imply about the conjecture? Use a probability statement.

- Solution:

$$P(X=6) = \sum_{x=0}^{6} b(x;10,0.3) - \sum_{x=0}^{5} b(x;10,0.3)$$
$$= 0.9894 - 0.9527 = 0.0367$$

For values of $b(x;10,0.3)$, see Table A.1 from text book.

- As a result, it is unlikely (3.6%chance) that 6 wells would be found impure if only 30% of all are impure. This casts considerable doubt on the conjecture and suggests that the impurity problem is much more severe.

- If the number of outcomes, k is more than two, it is referred to as **multinomial**. Suppose we have k possible outcomes ($k > 2$) in an experiment.

- **Multinomial Distribution**. If a given trial can result in the k outcomes $E_1, E_2, \ldots, E_k$ with probabilities $p_1, p_2, \ldots, p_k$, then the probability distribution of the random variables $X_1, X_2, \ldots, X_k$ representing the number of occurrences for $E_1, E_2, \ldots, E_k$ in n independent trials is

Let X be a random variable with probability distribution $f(x)$. The mean of the random variable $g(X)$ is

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

- $\sum_{i=1}^{k} x_i = n$ and $\sum_{i=1}^{k} p_i = 1$

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1! x_2! \ldots x_k!}$$

is the number of ways that yielding x_1 outcomes for E_1, x_2 outcomes for E_2, …, x_k outcomes for E_k.

- **Example**: A local gas station sells three types of gasoline; regular, premium and super.

- According to past sales, About 60% customers fuelregular, 30% put premium, and rest 10% buy super.

- For next 10 customers, find the probability that 5 buy regular gas, 4 fuel premium, and 1 buy super gas.
- Solution:
  Let X_i be the number of customers buy i^{th} product, $i = 1, 2, 3$. Then

$$f(5,4,1;\frac{60}{100},\frac{30}{100},\frac{10}{100},10) = \begin{pmatrix} 10 \\ 5,4,1 \end{pmatrix} \frac{60}{100}^5 \frac{30}{100}^4 \frac{10}{100}^1 = 0.07936$$

- **Example 5.10**: For a certain airport containing three runways it is known that in the ideal setting the following are the probabilities that the individual runways are accessed by a randomly arriving commercial jet:

  i Runway 1: $p_1 = 2/9$

  ii Runway 2: $p_2 = 1/6$

  iii Runway 3: $p_3 = 11/18$

- What is the probability that 6 randomly arriving airplanes are distributed in the following fashion? Runway 1: 2 airplanes, Runway 2: 1 airplanes, Runway 3: 3 airplanes.
- Solution: Using the multinomial distribution , we have

$$f(2,1,3;\frac{2}{9},\frac{1}{6},\frac{11}{18},6) = \begin{pmatrix} 6 \\ 2,1,3 \end{pmatrix} \frac{2}{9}^2 \frac{1}{6}^1 \frac{11}{18}^3$$

$$= \frac{6!}{2!1!3!} * \frac{2}{9}^2 * \frac{1}{6}^1 * \frac{11}{18}^3 = 0.1127$$

1.4 Hypergeometric Distribution

- There are two types of sampling methods from a finite population. If the population is infinite, two methods do not make any difference.
- **Binomial distribution**: the sampling **with** replacement (p is constant)
- **Hypergeometric distribution**: the sampling **without** replacement (p is not constant)
- Hypergeometric experiment:

1. A random sample of size n is selected without replacement from N items.
2. k of the N items may be classified as successes and $N - k$ as failures.

- Hypergeometric random variable: the number X of successes of a hypergeometric experiment.

- The probability distribution of the hypergeometric variable X, the number of successes in a random sample of size n selected from N items of which k are labeled success and $N - k$ labeled failure; $h(x; N, n, k)$.

$$h(x; N, n, k) =$$

$$\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

  - the number of ways of selecting x successes

$$\binom{k}{x}$$

  - the number of ways of selecting $n - x$ failures

$$\binom{N-k}{n-x}$$

  - the total number of samples of size n chosen from N items

$$\binom{N}{n}$$

- **Example 5.12**: Lots of 40 components each are called unacceptable if they contain as many as 3 defective or more.

- The procedure for sampling the lot is to select 5 components at random and to reject the lot if a defective is found.

- What is the probability that exactly 1 defective is found in the sample if there are 3 defectives in the entire lot?

- Solution: Using hypergeometric distribution with $n = 5, N = 40, k = 3$ and $x = 1$;

$$h(x; N, n, k) =$$

$$h(1; 40, 5, 3) =$$

$$\frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0.3011$$

- So this plan is likely not desirable since it detects a bad lot (3 defectives) only about 30% of the time.

- **Theorem 5.3**:
The mean and variance of the hypergeometric distribution $h(x; N, n, k)$ are
$$\mu = \frac{nk}{N} \text{ and } \sigma^2 = \frac{N-n}{n-1} * n * \frac{k}{N} * \left(1 - \frac{k}{N}\right)$$

- **Example 5.14**:Find the mean and variance of the random variable of Example 5.12 ($n = 5, N = 40,$ *and* $k = 3$) and then use Chebyshev's theorem to interpret the interval $\mu \pm 2\sigma$

- Solution:
$$\mu = \frac{5*3}{40} \text{ and } \sigma^2 = \frac{40-5}{40-1} * 5 * \frac{3}{40} * \left(1 - \frac{3}{40}\right) \Rightarrow \sigma = 0.558$$
$$\mu \pm 2\sigma = 0.3775 \pm 2 * 0.558$$
it has a probability of at least 3/ 4 of falling between -0.741 and 1.491.

- That is, at least three fourths of the time, the 5 components include less than 2 defectives.

- **Relationship to the Binomial Distribution**. If n is small compared to N, the nature of the N items changes very little in each draw. (when $\frac{n}{N} \leq 0.05$)

- $\mu = np = \frac{nk}{N}$ *and* $\sigma^2 = npq = n * \frac{k}{N} * \left(1 - \frac{k}{N}\right)$, where $\frac{N-n}{n-1}$ is negligible when n is smallrelative to N.
- The binomial distribution may be viewed as a large population edition of the hypergeometric distributions.
- **Example 5.15**: A manufacture of automobile tires reports that among a shipment of 5000 sent to a local distributor, 1000 are slightly faulty.
- If one purchases 10 of these tires at random from the distributor, what is the probability that exactly 3 are faulty?
- Solution: $h(x; N, n, k) =?$

$$h(3; 5000, 10, 1000) \approx$$

$$\sum_{x=0}^{3} b(x; 10, 0.2) - \sum_{x=0}^{2} b(x; 10, 0.2)$$
$$= 0.8791 - 0.6778 = 0.2013$$

- **Multivariate HypergeometricDistribution**: If N items can be partitioned into the k cells $A_1, A_2, \ldots, A_k$ with $a_1, a_2, \ldots, a_k$ elements, respectively, then the probability distribution of the random variable $X_1, X_2, \ldots, X_k$, representing the number of elements selected from $A_1, A_2, \ldots, A_k$ in a random sample of size n, is

$$f(x_1, x_2, \ldots, x_k; a_1, a_2, \ldots, a_k, N, n) = \frac{\begin{pmatrix} a_1 \\ x_1 \end{pmatrix}\begin{pmatrix} a_2 \\ x_2 \end{pmatrix}\cdots\begin{pmatrix} a_k \\ x_x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

- with $\sum_{i=1}^{k} x_i = n$ and $\sum_{i=1}^{k} a_i = N$
- **Example 5.16**: A group of 10 individuals are used for a biological case study.
- The group contains 3 people with blood type O, 4 with blood type A, and 3 with blood type B.
- What is the probability that a random sample of 5 will contain 1 person with blood type O, 2 with blood type A, and 2 with blood type B?

- Solution:

$$f(1,2,2;3,4,3,10,5) = \frac{\binom{3}{1}\binom{4}{2}\binom{3}{2}}{\binom{10}{5}} = \frac{3}{14}$$