Chapter 5
Some Discrete Probability Distributions

吳明龍 副教授
成大資訊系/醫資所
minglong.wu@csie.ncku.edu.tw
辦公室:資訊系館 12 樓

5.1 Introduction

- Often, the observations generated by different statistical experiments have the same general type of behavior.

- In fact, one needs only a handful of important probability distributions to describe many of the discrete random variables encountered in practice.

5.1 Introduction

- Binomial distribution (Section 5.3): test the effectiveness of a new drug.
- Hypergeometric distribution (Section 5.4): test the number of defective items from a batch of production.
- Negative binomial distribution (Geometric distribution) (Section 5.5): the number of samples required to produce a false alarm
- Poisson distribution (Section 5.6): the number of white cells from a fixed amount of an individual's blood sample.

5.2 Discrete Uniform Distribution (8th Ed. only)

- **Discrete Uniform Distribution**: If the random variable X assumes the values x_1, x_2, …, x_k, with equal probabilities, then the discrete uniform distribution is given by

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

- 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

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

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

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

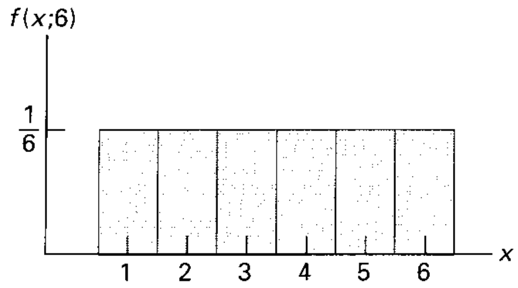

Figure 5.1 Histogram for the tossing of a die.

Discrete Uniform Distribution (8th Ed. only)

- Theorem 5.1: The mean and variance of the discrete uniform distribution $f(x; k)$ are

$$\mu = \frac{\sum_{i=1}^{k} x_i}{k}, \quad \sigma_X^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k}.$$

  – **Proof**

$$\mu = E(X) = \sum_{i=1}^{k} x_i f(x;k) = \sum_{i=1}^{k} \frac{x_i}{k} = \frac{\sum_{i=1}^{k} x_i}{k}$$

$$\sigma^2 = E[(X-\mu)^2] = \sum_{i=1}^{k} (x_i - \mu)^2 f(x;k) = \sum_{i=1}^{k} \frac{(x_i-\mu)^2}{k} = \frac{\sum_{i=1}^{k} (x_i-\mu)^2}{k}.$$

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

$$\mu = \frac{1+2+3+4+5+6}{6} = 3.5$$

$$\sigma^2 = \frac{(1-3.5)^2 + (2-3.5)^2 + \cdots + (6-3.5)^2}{6} = \frac{35}{12}$$

5.2 Binomial and Multinomial Distribution

- An experiment often consists of repeated trials, each with two possible outcomes that may be labeled success or failure.
- The most obvious application deals with the testing of items as they come off an assembly line, where each test/trial may indicate a defective or a nondefective item.
- The Bernoulli Process
  - The experiment consists of n repeated trials.
  - Each trial results in an outcome that may be classified as a success or a failure.
  - The probability of success, denoted by p, remains constant from trial to trial.
  - The repeated trials are independent.

Binomial and Multinomial 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

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

- Example 5.1: The probability that a certain kind of component will survive a given shock test is ¾. Find the probability that exactly 2 of the next 4 components tested survive.

$$b(2;4,\tfrac{3}{4}) = \binom{4}{2}\left(\tfrac{3}{4}\right)^2\left(\tfrac{1}{4}\right)^2 = \frac{4!}{2!2!}\cdot\frac{3^2}{4^4} = \frac{27}{128}.$$

Binomial and Multinomial Distribution

- Binomial distribution corresponds to the binomial expansion of $(q + p)^n$, i.e.,

$$(q+p)^n = \binom{n}{0}q^n + \binom{n}{1}pq^{n-1} + \binom{n}{2}p^2q^{n-2} + \cdots + \binom{n}{n}p^n$$
$$= b(0;n,p) + b(1;n,p) + b(2;n,p) + \cdots + b(n;n,p) = 1$$

- Example 5.2: 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 (a) at least 10 survive, (b) from 3 to 8 survive, and (c) exactly 5 survive?

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

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

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

Acrobat Reader - [binominal.pdf]

File Edit View Tools Window Help

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	0.0000
	8		0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9		0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0023
	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	0.7941
	15							1.0000	1.0000	1.0000	1.0000
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	
	7	0.9999	0.9930	0.9729	0.9256	0.7161	0.4018	0.1423	0.0257	0.0015	0.0000

Page 1 of 1 200% 8.97 x 12.06 in

Binomial and Multinomial Distribution

- Example 5.3: 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%
  a) 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?
  b) 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?

Solution

(a) Denote by X the number of defective devices among the 20, X follows $b(x;20,0.03)$.

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

(b) Denote by Y the number of shipments containing at least one defective item, Y follows $b(y;10,0.4562)$

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

Binomial and Multinomial Distribution

- Theorem 5.1: The mean and variance of the binomial distribution $b(x; n, p)$ are $\mu = np$ and $\sigma^2 = npq$
  - **Proof**

Let the outcome on the jth trial represented by a Bernoulli random variable I_j.

The number of successes in a binomial experiment is denoted by

$$X = I_1 + I_2 + \cdots + I_n$$

$$E(I_j) = 0 \cdot q + 1 \cdot p = p$$

$$E(X \pm Y) = E(X) \pm E(Y).$$

$$\mu_X = E(X) = E(I_1) + E(I_2) + \cdots + E(I_n) = \underbrace{p + p + \cdots + p}_{n \text{ terms}} = np.$$

$$\sigma_{I_j}^2 = E[(I_j - p)^2] = E(I_j^2) - p^2 = (0^2 \cdot q + 1^2 \cdot p) - p^2 = p(1-p) = pq.$$

$$\sigma_X^2 = \sigma_{I_1}^2 + \sigma_{I_2}^2 + \cdots + \sigma_{I_n}^2 = \underbrace{pq + pq + \cdots + pq}_{n \text{ terms}} = npq.$$

$$\sigma^2_{a_1X_1 + a_2X_2 + \cdots + a_nX_n} = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + \cdots + a_n^2\sigma_{X_n}^2.$$

Binomial and Multinomial Distribution

- Example 5.5: Find the mean and variance of the binomial random variable of Example 5.2 (n = 15, p = 0.4), and then use Chebyshev's theorem to interpret the interval $\mu \pm 2\sigma$.

(Example 5.2: The probability that a patient recovers from a rare blood disease is 0.4.)

- **Solution**

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

$\mu = 15 \cdot 0.4 = 6$

$\sigma^2 = 15 \cdot 0.4 \cdot 0.6 = 3.6, \sigma = \sqrt{3.6} = 1.897$

$*$ Chebyshev's theorem:

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

The interval $\mu \pm 2\sigma = 6 \pm 2 \cdot 1.897 \Rightarrow 2.206$ to 9.794

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

Binomial and Multinomial Distribution

- Example 5.6: It is conjectured that an impurity exists in 30% of all drinking wells in a city. 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** If the conjecture is correct, is it likely that we could have found 6 or more impure wells?

$$P(X \geq 6) = \sum_{x=0}^{10} b(x;10,0.3) - \sum_{x=0}^{5} b(x;10,0.3)$$
$$= 1 - 0.9527 = 0.0473$$

As a result, it is unlikely (4.7% 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.**

Binomial and Multinomial Distribution

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

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

$$\text{with } \sum_{i=1}^{k} x_i = n \quad \text{and} \quad \sum_{i=1}^{k} p_i = 1.$$

Binomial and Multinomial Distribution

- Example 5.7: The complexity of arrivals and departures into an airport are such that computer simulation is often used to model the "ideal" conditions. 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: Runway 1: p_1= 2/9, Runway 2: p_2= 1/6, 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 multinomial distribution

$$f\left(2,1,3;\frac{2}{9},\frac{1}{6},\frac{11}{18},6\right)=\binom{6}{2,1,3}\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{11}{18}\right)^3=\frac{6!}{2!1!3!}\cdot\frac{2^2}{9^2}\cdot\frac{1}{6}\cdot\frac{11^3}{18^3}=0.1127$$

Hypergeometric Distribution

- Binomial distribution: the sampling with replacement
  Hypergeometric distribution: the sampling without replacement
- 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.
- Hypergeometric distribution: 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.

$$h(x;N;n,k)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

Hypergeometric Distribution

- Example 5.9: 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 $x = 1$, $N = 40$, $n = 5$, and, $k = 3$

$$h(1;40,5,3) = \frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0.3011 \; .$$

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

Hypergeometric Distribution

- Theorem 5.2: The mean and variance of the hypergeometric distribution $h(x; N, n, k)$ are

$$\mu = \frac{nk}{N} \text{ and } \sigma^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}(1-\frac{k}{N})$$

  (the proof is shown in Appendix A24)

- Exapmle 5.11: Find the mean and variance of the random variable of Example 5.9 (n = 5, N = 40, and k = 3) and then use Chebyshev's theorem to interpret the interval $\mu \pm 2\sigma$.
  - **Solution**

$$\mu = \frac{5 \cdot 3}{40} = \frac{3}{8} = 0.375$$

$$\sigma^2 = (\frac{40-5}{39})(5)(\frac{3}{40})(1-\frac{3}{40}) = 0.3113 \Rightarrow \sigma = 0.558$$

$$\mu \pm 2\sigma = 0.375 \pm 2 \cdot 0.558$$

$$\Rightarrow \text{has a probability of at least } 3/4 \text{ of falling between } -0.741 \text{ and } 1.491.$$

A.24

PROOF To find the mean of the hypergeometric distribution, we write

$$E(X) = \sum_{x=0}^{n} x \frac{\boxed{\binom{k}{x}}\binom{N-k}{n-x}}{\binom{N}{n}} = k\sum_{x=1}^{n} \frac{(k-1)!}{(x-1)!(k-x)!} \cdot \frac{\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= k\sum_{x=1}^{n} \frac{\binom{k-1}{x-1}\binom{N-k}{n-x}}{\binom{N}{n}}.$$

Letting $y = x - 1$, we find that this becomes

$$E(X) = k\sum_{y=0}^{n-1} \frac{\binom{k-1}{y}\binom{N-k}{n-1-y}}{\boxed{\binom{N}{n}}}.$$

Writing

$$\binom{N-k}{n-1-y} = \binom{(N-1)-(k-1)}{n-1-y}$$

and

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N}{n}\binom{N-1}{n-1},$$

we obtain

$$E(X) = \frac{nk}{N}\sum_{y=0}^{n-1} \frac{\binom{k-1}{y}\binom{(N-1)-(k-1)}{n-1-y}}{\binom{N-1}{n-1}} = \frac{nk}{N},$$

since the summation represents the total of all probabilities in a hypergeometric experiment when $n - 1$ items are selected at random from $N - 1$, of which $k - 1$ are labeled success.

Hypergeometric Distribution

- 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 $\frac{n}{N} \le 0.05$)
  - $\mu = np = \frac{nk}{N}, \ \sigma^2 = npq = n \cdot \frac{k}{N}(1 - \frac{k}{N})$

    $(\frac{N-n}{N-1}$ is negligible when n is small relative to $N) \cdot$
  - The binomial distribution may be viewed as a large population edition of the hypergeometric distributions.
- Example 5.12: A manufacture of automobile tires reports that among a shipment of 5000 sent to a local distributor, 1000 are slightly blemished. If one purchases 10 of these tires at random from the distributor, what is the probability that exactly 3 are blemished?
  - **Solution**

$$h(3;5000,10,1000) \approx \underline{b(3;10,0.2)} = \sum_{x=0}^{3} b(x;10,0.2) - \sum_{x=0}^{2} b(x;10,0.2)$$

$$= 0.8791 - 0.6778 = 0.2013$$

Hypergeometric Distribution

- <u>Multivariate Hypergeometric Distribution</u>:
  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

$$f(x_1, x_2, \cdots, x_k; a_1, a_2, \cdots, a_k, N, n) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\binom{a_k}{x_k}}{\binom{N}{n}} \quad \text{with } \sum_{i=1}^{k} x_i = n \text{ and } \sum_{i=1}^{k} a_i = N$$

- Example 5.13: 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**

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