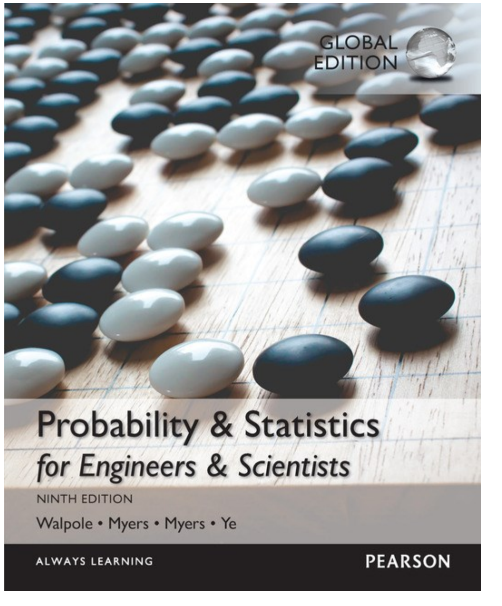
Chapter 5

Some Discrete Probability Distributions – part 1

School of Computing, Gachon Univ.
Joon Yoo



Introduction

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

- Consequently, discrete random variables associated with these experiments can be described by essentially the same probability distribution and therefore can be represented by a single formula.

So !!

- If you have an in-depth understand of some probability distributions, you can use them to describe many of the discrete random variables encountered in practice.

- In fact, one needs only a handful of important probability distributions to describe lots of real-life random phenomena encountered in practice.

  - For instance, in a study involving testing the effectiveness of a new drug, the number of cured patients among all the patients who use the drug approximately follows a ***binomial*** *distribution*.

G-WINL Wireless Intelligent Networking LAB

5.2 Binomial and Multinomial Distributions

- Question
  - Suppose a fair coin is tossed 10 times. Let X denote the number of heads in this 10 tosses. What is the probability distribution of the random variable X ?
    - (즉, 10번 시행 중 몇 개의 head (=성공) 횟수 X의 확률분포?)

- Question 2
  - Suppose a fair coin is tossed 10 times. If the number of heads in this 10 trials is 2. What is the probability of showing the head at 11-th toss ?

Introduction: Bernoulli* distribution

- An experiment often consists of **repeated trials**, each with two possible outcomes that may be labeled success or failure.
  - (instead of success & failure, many types of two outputs can be used; e.g., Yes / No, 0 / 1, T / F, or etc.)
- In such the experiment, each trial is called a **Bernoulli trial**.
- The process is referred to as a **Bernoulli process**.

*베르누이

WINL Wireless Intelligent Networking LAB

Bernoulli Process

- Strictly speaking, the Bernoulli process must possess the following properties:
  - 1. The experiment consists of n ***repeated*** trials.
  - 2. Each trial results in an outcome that may be classified as a success or a failure.
  - 3. The probability of success, denoted by ***p***, remains **constant** from trial to trial.
  - 4. The repeated trials are **independent**.

Example

- The experiment consists of repeated trials. We flip a coin 2 times.
- Each trial can result in just two possible outcomes - heads or tails.
- The probability of success is constant 0.5 on every trial.
- The trials are independent; that is, getting heads on one trial does not affect whether we get heads on other trials

Bernoulli RV

- Property
  - Expectation

$$E[X] = 1 \cdot P\{X = 1\} + 0 \cdot P\{X = 0\} = p$$

  - Variance

$$\sigma^2 = E(X^2) - E(X)^2 = p\,(1-p)$$

$$\Uparrow$$

$$E(X^2) = 0^2 \times f(0) + 1^2 \times f(1) = p$$

Wireless Intelligent Networking LAB

Example

- Coin toss (success with p=1/4)
  - X=1 for success (Head) and X=0 for failure (Tail)
  - Find pmf

$$p(x) = \begin{cases} 3/4 & , x = 0 \\ 1/4 & , x = 1 \end{cases}$$

  - Mean $E[X]=p=1/4$
  - Variance $\mathrm{Var}(X)=p(1-p)=3/4 \times 1/4= 3/16$

- Example
  - Consider the set of Bernoulli trials where three items are selected at random from a manufacturing process, inspected, and classified as **defective** or non-defective.
  - A defective item is designated a success. **The number of successes is a random variable X** assuming integral values from 0 through 3. The eight possible outcomes and the corresponding values of X are

Outcome	NNN	NDN	NND	DNN	NDD	DND	DDN	DDD
x	0	1	1	1	2	2	2	3

  - Since the items are selected independently and we assume that the process produces 25% defectives, we have

$$P(NDN) = P(N)P(D)P(N) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{9}{64}.$$

  - Distribution of X = ?

x	0	1	2	3
$f(x)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

Binomial RV!

WINL Wireless Intelligent Networking LAB

Binomial (이항) RV

- Extension of Bernoulli RV
  - Suppose that n independent Bernoulli trials are performed
    - X = the number of success that occur in the n trials

H with probability p H with probability p H with probability p …   H with probability p

n toss coins

Q: What is the probability $P(X=x)$?

Binomial RV & distribution

- **Binomial random variable X**
  - The number of *X* of successes in **n** Bernoulli trials is called a binomial random variable.
- Binomial distribution $b(x; n; p)$ *or* $b(x; p)$
  - The probability distribution of this discrete random variable X is called the binomial distribution, and its values will denoted by $b(x; n; p)$
    - since they depend on **the number of trials, n**, and the **probability of a success on a given trial, *p***.

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 (p.m.f.) of the binomial random variable X, **the number of successes in n** independent trials, is

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

*이항분포

- Example
  - The probability that a certain kind of component will survive a shock test is 3/4. Find the probability that exactly 2 of the next 4 components tested survive.

- Solution
  - Let X be the number of components that survives in the next 4 tested components and assuming that the tests are independent and *p* = ¾ for each of the 4 tests, we obtain

$$X \sim \text{Bin}(4, 3/4)$$

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

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

Wireless Intelligent Networking LAB

Where does the Name "*Binomial*" Come From?

=이항식

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

$$= b(0;n,p) + b(1;n,p) + b(2;n,p) + \cdots + b(n;n,p)$$

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

$$= 1 \quad (\text{when } p+q=1).$$

=n개중 p가 2차인 개수

- Frequently, we are interested in problems where it is necessary to find P (X < r) or P (a ≤ X ≤ b). Binomial sums

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

(see Table A.1)

Wireless Intelligent Networking LAB

- 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?

(use Table A.1)

- Solution: (use Table A.1)
  - Let X be the number of people that survive. X ~ Bin(15; 0.4).

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

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

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

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

G-WINL Wireless Intelligent Networking LAB

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

						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	0.0093	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	0.0950	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	0.0022
	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
	8	1.0000	0.9985	0.9925	0.9743	0.8577	0.5982	0.2839	0.0744	0.0070	0.0001
	9		0.9998	0.9984	0.9929	0.9417	0.7728	0.4728	0.1753	0.0267	0.0005
	10		1.0000	0.9997	0.9984	0.9809	0.8949	0.6712	0.3402	0.0817	0.0033
	11			1.0000	0.9997	0.9951	0.9616	0.8334	0.5501	0.2018	0.0170
	12				1.0000	0.9991	0.9894	0.9349	0.7541	0.4019	0.0684
	13					0.9999	0.9979	0.9817	0.9006	0.6482	0.2108
	14					1.0000	0.9997	0.9967	0.9739	0.8593	0.4853
	15						1.0000	0.9997	0.9967	0.9719	0.8147
	16							1.0000	1.0000	1.0000	1.0000

- 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 randomly picks 20 items from a shipment. What is the probability that there will be at least one defective item among these 20?

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

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

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

GWINL
Wireless Intelligent Networking LAB

- Example 5.3
  - (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 exactly 3 shipments each containing at least one defective device among the 20 that are selected and tested from the shipment?

(b) In this case, each shipment can either contain at least one defective item or not. Hence, testing of each shipment can be viewed as a Bernoulli trial with $p = 0.4562$ from part (a). Assuming independence from shipment to shipment and denoting by Y the number of shipments containing at least one defective item, Y follows another binomial distribution $b(y; 10, 0.4562)$. Therefore,

$$P(Y = 3) = \binom{10}{3} 0.4562^3 (1 - 0.4562)^7 = 0.1602.$$

Mean, Variance of b(x; n, p)

- Theorem 5.1

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

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

Proof

- Property
  - Binomial RV X = sum of Bernoulli RVs X_i

$$X = \sum_{i=1}^{n} X_i \qquad \text{where} \quad X_i = \begin{cases} 1 & \text{if the } i\text{th trial is a success} \\ 0 & \text{otherwise} \end{cases}$$

  - Expectation
    - Using the independence of X_i

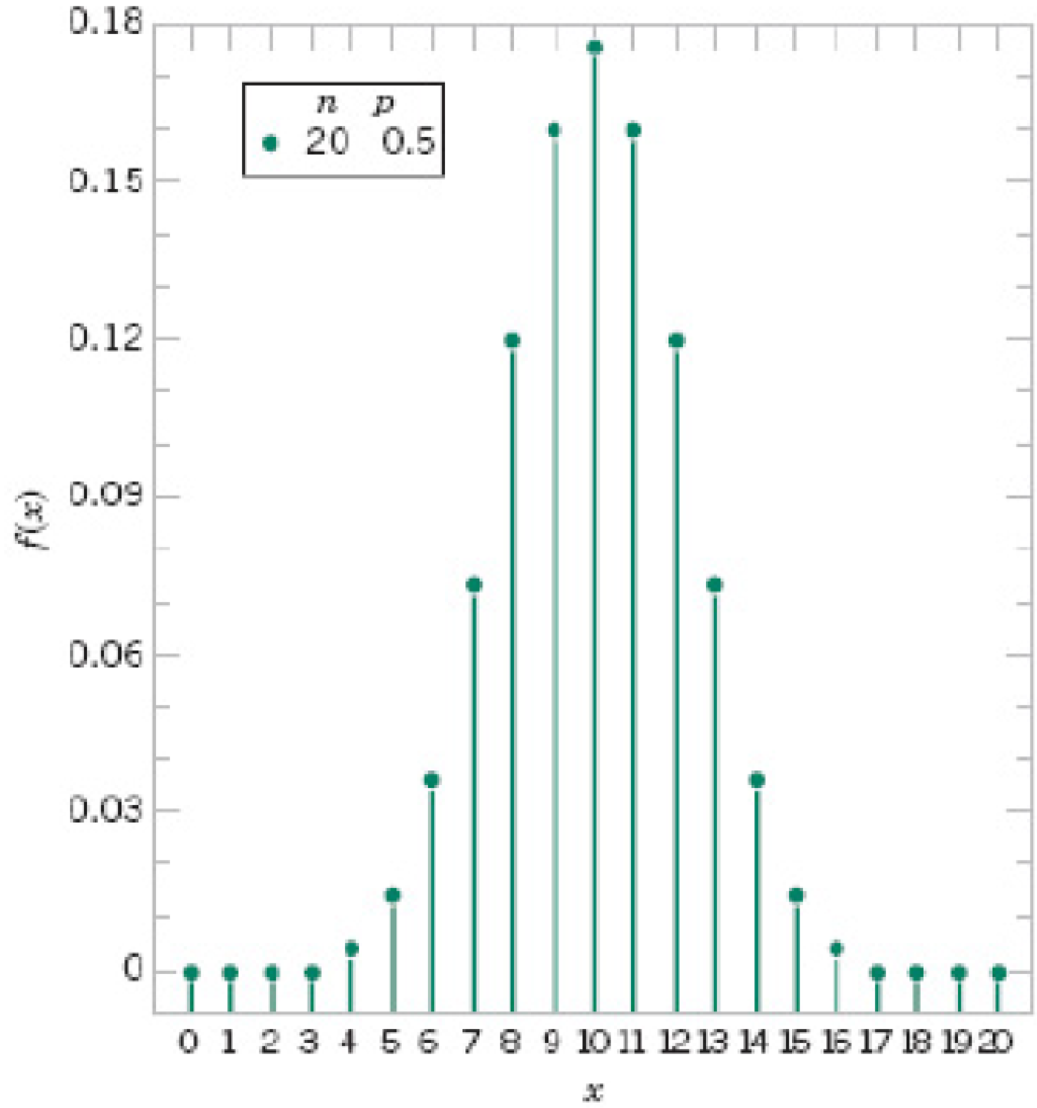
$$E[X] = \sum_{i=1}^{n} E[X_i]$$
$$= np$$

$$\boxed{\begin{aligned} E[X_i] &= P\{X_i = 1\} = p \\ \mathrm{Var}(X_i) &= E[X_i^2] - p^2 \\ &= p(1-p) \end{aligned}}$$

  - Variance

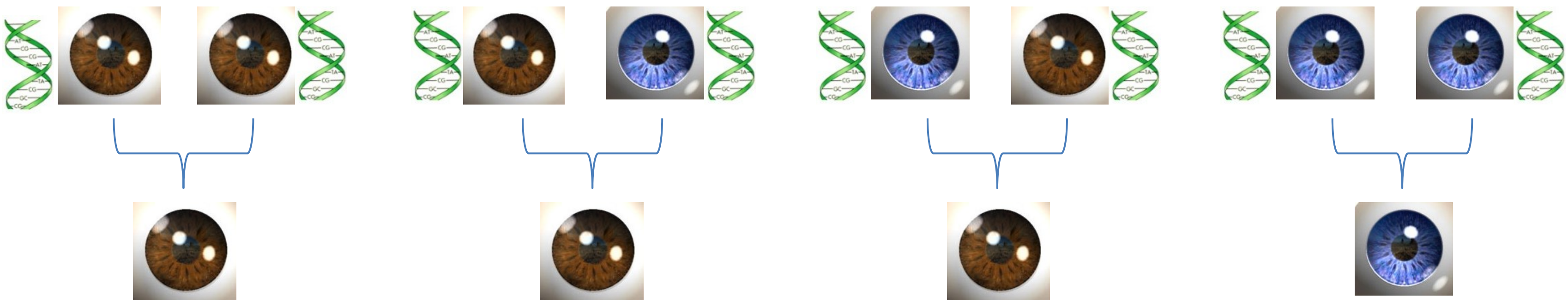
$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) \quad \text{since the } X_i \text{ are independent}$$
$$= np(1-p) = npq$$

- Probability function for a binomial random variable with n = 20 and p = 0.5.

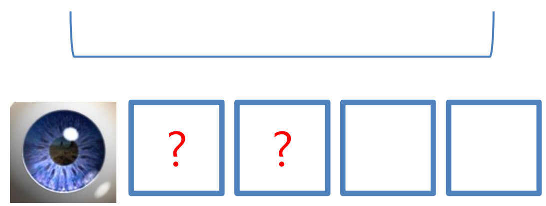


Example

- The color of one's eyes is determined by **a single pair of genes**, with the gene for **brown eyes being dominant** over the one for blue eyes.
- When two people mate, the resulting offspring receives one randomly chosen gene from each of its parents' gene pair. If the **eldest child of a pair of brown-eyed parents has blue eyes**, what is the probability that exactly **two of the four other children (none of whom is a twin) of this couple also have blue eyes**?



Probability that it receives the blue-eyed gene ½, from both parents 1/2 x ½ =1/4



$$\binom{4}{2}(1/4)^2(3/4)^2 = 27/128$$

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

Areas of Application of Binomial Distribution

- An industrial engineer is keenly interested in the “proportion defective” in an industrial process. Often, quality control measures and sampling schemes for processes are based on the binomial distribution.
- This distribution applies to any industrial situation where an outcome of a process is dichotomous (양분된) and the results of the process are independent, with the probability of success being constant from trial to trial.
- The binomial distribution is also used extensively for medical and military applications. In both fields, a success or failure result is important. For example, “cure” or “no cure” is important in pharmaceutical work, and “hit” or “miss” is often the interpretation of the result of firing a guided missile.

Wireless Intelligent Networking LAB

Multinomial Experiment

- The binomial experiment becomes a multinomial experiment if we let each trial have **more than 2 possible outcomes**.
- Examples
  - The classification of a manufactured product as being **light**, **heavy**, or **acceptable**
  - The recording of accidents at a certain intersection according to the day of the week

  - The drawing of a card from a deck with replacement is also a multinomial experiment if the 4 suits are the outcomes of interest.

Multinomial Experiment

- In general, if a given trial can result in any one of k possible outcomes $E_1, E_2, \ldots, E_k$ with probabilities $p_1, p_2, \ldots, p_k$, then the **<u>multinomial distribution will give the probability</u>** that E_1 occurs x_1 times, E_2 occurs x_2 times, . . . , and E_k occurs x_k times in n *independent trials*, where

$$x_1 + x_2 + \cdots + x_k = n.$$

We shall denote this joint probability distribution by

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n).$$

Clearly, $p_1 + p_2 + \cdots + p_k = 1$, since the result of each trial must be one of the k possible outcomes.

Multinomial Experiment (cont.)

- The total number of orders yielding similar outcomes for the ***n*** **trials** is equal to the number of partitions of *n* items into ***k*** groups with x_1 in the first group; $\mathbf{x_2}$ in the second group, … ; and $\mathbf{x_k}$ in the *k*-th group. This can be done in

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1!\,x_2!\cdots x_k!} \qquad \text{ways}$$

Wireless Intelligent Networking LAB

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

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

with

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

- **Example**
  - According to a genetics theory (유전이론), a certain cross of guinea pigs will result in **red**, **black**, and **white** offspring in the ration 8:4:4. Find the probability that among 8 offspring **5 will be red**, **2 black**, and **1 white**.

- **Solution**
  - Let $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ represent the number of offspring in red, black, and white, respectively.
    ($\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$) ~ Multinomial(8/16, 4/16, 4/16, 8)

$$P(X_1 = 5, X_2 = 2, X_3 = 1) = \binom{8}{5,2,1} \left(\frac{8}{16}\right)^5 \left(\frac{4}{16}\right)^2 \left(\frac{4}{16}\right)^1.$$

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

- **Example 5.7**

The complexity of arrivals and departures of planes at an airport is such that computer simulation is often used to model the "ideal" conditions. For a certain airport with 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:

$$\begin{aligned} &\text{Runway 1:} \quad p_1 = 2/9, \\ &\text{Runway 2:} \quad p_2 = 1/6, \\ &\text{Runway 3:} \quad p_3 = 11/18. \end{aligned}$$

What is the probability that 6 randomly arriving airplanes are distributed in the following fashion?

Runway 1: 2 airplanes,
Runway 2: 1 airplane,
Runway 3: 3 airplanes

- Solution

Using the multinomial distribution, we have

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

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

End of slide