Table 2-2

Successes		Frequency	
Number	Proportion: $\hat{\pi}$	Absolute	Relative: f
0	0.0000	0	0.00
1	0.0625	0	0.00
2	0.1250	0	0.00
3	0.1875	0	0.00
4	0.2500	0	0.00
5	0.3125	1	0.01
6	0.3750	0	0.00
7	0.4375	1	0.01
8	0.5000	5	0.05
9	0.5625	10	0.10
10	0.6250	17	0.17
11	0.6875	21	0.21
12	0.7500	20	0.20
13	0.8125	15	0.15
14	0.8750	7	0.07
15	0.9375	3	0.03
16	1.0000	0	0.00
		100	1.00

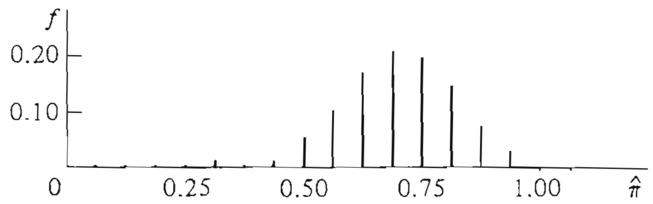

Figure 2-2

(iii) the sampling distribution for samples size 16 is considerably more symmetric than that for samples size 4.

These results have been obtained by repeated sampling from a dichotomous population (i.e., a population containing only two types of individuals) with a proportion of successes equal to 0.7. Only two sampling distributions, those corresponding to samples size 4 and size 16, have been derived. But even given this specific character of our experiments, the results clearly inspire certain generalizations about the sampling distribution of the proportion of successes observed in a sample ($\hat{\pi}$) as an estimator of the proportion of successes in the population (π). These generalizations are

1. $\hat{\pi}$ is an *unbiased estimator* of π (i.e., the mean of the sampling distribution of $\hat{\pi}$ is equal to the population parameter π).

values 0 to 9, sample means will, in general, not be intergers. Thus the sampling distribution will be a frequency distribution with classes defined by intervals and not by points. We may, of course, choose a single value such as the center of each interval to represent each class. The distribution obtained as a result of our experi-

Table 2-4

Value of sample mean: $\bar{x}$		Frequency	
Interval	Midpoint	Absolute	Relative: f
0.5 to 1.499	1	1	0.01
1.5 to 2.499	2	5	0.05
2.5 to 3.499	3	12	0.12
3.5 to 4.499	4	31	0.31
4.5 to 5.499	5	28	0.28
5.5 to 6.499	6	15	0.15
6.5 to 7.499	7	5	0.05
7.5 to 8.499	8	3	0.03
8.5 to 9.499	9	0	0.00
		100	1.00

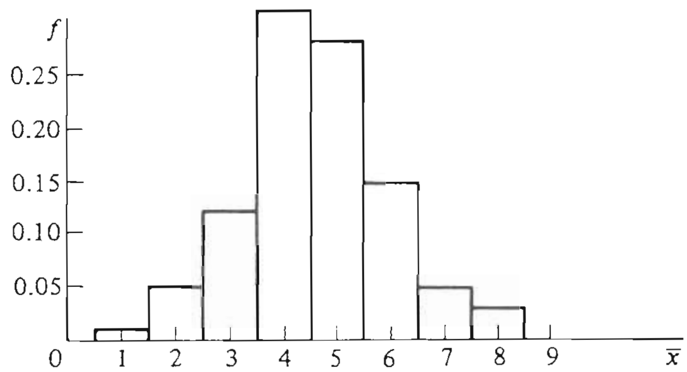

Figure 2-3

ment is shown in Table 2-4 and illustrated by Figure 2-3. The main characteristics of this distribution are

$$\text{Mean} = \sum_{i=1}^{9} f_i \bar{x}_i = 4.60.$$

$$\text{Standard deviation} = \sqrt{\sum_{i=1}^{9} f_i(\bar{x}_i - 4.60)^2} = 1.3638.$$

The results indicate that 59% of the estimated values $(0.31 + 0.28 = 0.59)$ fall within ± 1 of the true value of 4.5, while 86% of the estimates lie within ± 2 of 4.5.

Next we present the derived sampling distribution of sample median. Since the sample size is an odd number and all values of X are integers, sample median will

Table 2-6

Value of sample mean: $\bar{x}$		Frequency	
Interval	Midpoint	Absolute	Relative: f
0.5 to 1.499	1	0	0.00
1.5 to 2.499	2	1	0.01
2.5 to 3.499	3	14	0.14
3.5 to 4.499	4	34	0.34
4.5 to 5.499	5	32	0.32
5.5 to 6.499	6	16	0.16
6.5 to 7.499	7	3	0.03
7.5 to 8.499	8	0	0.00
8.5 to 9.499	9	0	0.00
		100	1.00

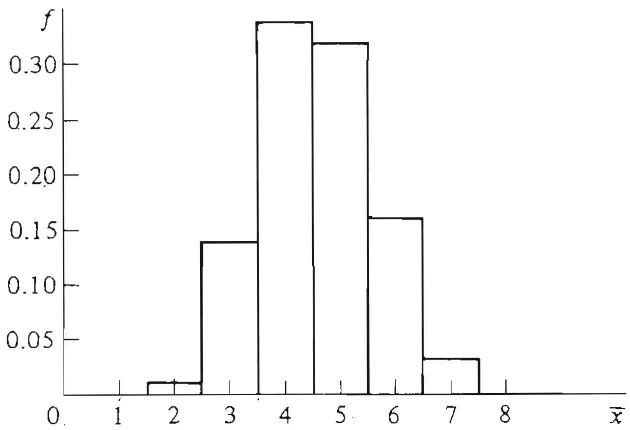

Figure 2-5

are

$$\text{Mean} = \sum_{i=2}^{7} f_i \bar{x}_i = 4.57.$$

$$\text{Standard deviation} = \sqrt{\sum_{i=2}^{7} f_i(\bar{x}_i - 4.57)^2} = 1.0416.$$

The results of our sampling experiments in case B can be summarized as follows: (i) the mean of the sampling distribution of sample mean is approximately equal to the population mean for both sample sizes examined; (ii) the dispersion of the sampling distribution of sample mean for samples size 10 is less than that for samples size 5, the variance of the latter (1.3638^2) being almost twice as large as that of the former (1.0416^2); and (iii) the mean of the sampling distribution of sample median is also approximately equal to the population mean, but its variation is greater than that of the sample mean in samples of equal size.

Table 4-5

Proportion of successes: $\hat{\pi}$	Experimental $f(\hat{\pi})$	Theoretical $f(\hat{\pi})$
0	0	0
1/16	0	0.00
2/16	0	0.00
3/16	0	0.00
4/16	0	0.00
5/16	0.01	0.00
6/16	0	0.01
7/16	0.01	0.02
8/16	0.05	0.05
9/16	0.10	0.10
10/16	0.17	0.17
11/16	0.21	0.21
12/16	0.20	0.20
13/16	0.15	0.15
14/16	0.07	0.07
15/16	0.03	0.02
1	0	0.00
Mean	0.7006	0.7
Standard deviation	0.1191	0.1146

Thus, in this case, distributions based on 100 samples come quite close to those that would result from an infinite number of samples. Furthermore, the generalizations made on the basis of experimental results—namely unbiasedness, consistency, relative change in the standard deviation, and decrease in skewness—all proved to be correct. However, the experimental results compare unfavorably with the theoretical ones in two respects. In the first place, the experimental results fail to give us any formulas for variance, measure of skewness, and, of course, the individual probabilities. In the second place, and this is much more important, there is no guarantee at all that the generalizations deduced from the experimental distributions are, in fact, valid. The conclusions are not proved, only suggested by the results of isolated experiments.

4-2 Normal Distribution as the Limiting Case of Binomial Distribution

One of the findings of Section 4-1 was that the binomial distribution tends to be increasingly more symmetric as n (size of sample) increases, regardless of the value of π. Even the distributions with π close to zero (or to unity), which for small n are very skewed, tend to become symmetric when n is somewhat larger. This point is demonstrated by Figure 4-1 which shows the binomial probabilities for $\pi = 0.10$ for

Table 4-9

$\bar{x}$	$f(\bar{x})$		$\bar{x}f(\bar{x})$	$\bar{x}^2 f(\bar{x})$
0	$P(0, 0)$	$= 1/10^2$	0.00	0.000
1/2	$P(0, 1) + P(1, 0)$	$= 2/10^2$	0.01	0.005
2/2	$P(0, 2) + P(2, 0) + P(1, 1)$	$= 3/10^2$	0.03	0.030
3/2	$P(0, 3) + P(3, 0) + P(1, 2) + P(2, 1)$	$= 4/10^2$	0.06	0.090
4/2	. . .	$= 5/10^2$	0.10	0.200
5/2	. . .	$= 6/10^2$	0.15	0.375
6/2	. . .	$= 7/10^2$	0.21	0.630
7/2	. . .	$= 8/10^2$	0.28	0.980
8/2	. . .	$= 9/10^2$	0.36	1.440
9/2	$P(0, 9) + P(9, 0) + P(1, 8) + P(8, 1)$ $+ P(2, 7) + P(7, 2) + P(3, 6)$ $+ P(6, 3) + P(4, 5) + P(5, 4)$	$= 10/10^2$	0.45	2.025
10/2	$P(1, 9) + P(9, 1) + P(2, 8) + P(8, 2)$ $+ P(3, 7) + P(7, 3) + P(4, 6)$ $+ P(6, 4) + P(5, 5)$	$= 9/10^2$	0.45	2.250
⋮	⋮	⋮	⋮	⋮
17/2	$P(8, 9) + P(9, 8)$	$= 2/10^2$	0.17	1.445
18/2	$P(9, 9)$	$= 1/10^2$	0.09	0.810
	Sum	1	4.50	24.375

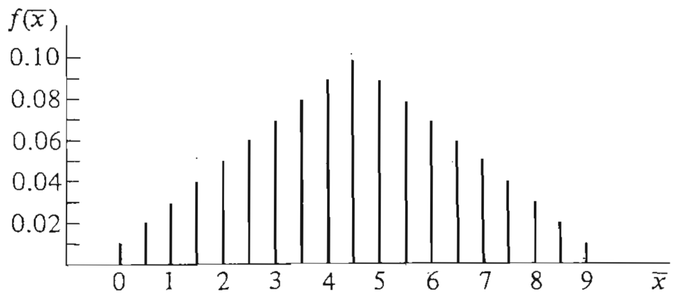

Figure 4-12

The distribution is perfectly symmetric around the point 4.5. Its graphical representation is given in Figure 4-13.

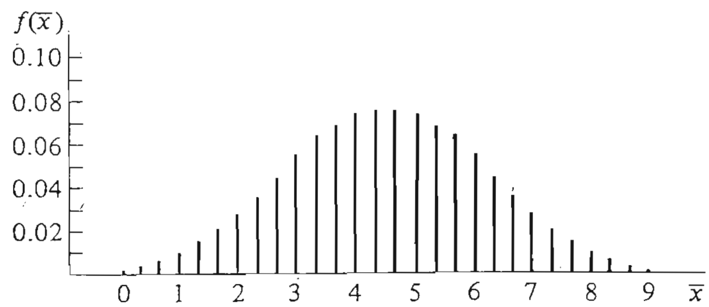

Figure 4-13

	Sample size 5	
Interval: $\bar{x}$	Experimental $f(\bar{x})$	Theoretical $f(\bar{x})$
0.5 to 1.499	0.01	0.01
1.5 to 2.499	0.05	0.05
2.5 to 3.499	0.12	0.16
3.5 to 4.499	0.31	0.28
4.5 to 5.499	0.28	0.28
5.5 to 6.499	0.15	0.16
6.5 to 7.499	0.05	0.05
7.5 to 8.499	0.03	0.01
8.5 to 9.499	0.00	0.00
Mean	4.60	4.5
Standard deviation	1.3638	1.2845

	Sample size 10	
Interval: $\bar{x}$	Experimental $f(\bar{x})$	Theoretical $f(\bar{x})$
0.5 to 1.499	0.00	0.00
1.5 to 2.499	0.01	0.02
2.5 to 3.499	0.14	0.12
3.5 to 4.499	0.34	0.36
4.5 to 5.499	0.32	0.36
5.5 to 6.499	0.16	0.12
6.5 to 7.499	0.03	0.02
7.5 to 8.499	0.00	0.00
8.5 to 9.499	0.00	0.00
Mean	4.57	4.5
Standard deviation	1.0416	0.9083