EXAMPLE 4 Consider sampling with replacement from an urn containing M balls, K of which are defective. Let X represent the number of defective balls in a sample of size n. The individual draws are Bernoulli trials where "defective" corresponds to "success," and the experiment of taking a sample of size n with replacement consists of n repeated independent Bernoulli trials where $p = P[\text{success}] = K/M$; so X has the binomial distribution

$$\binom{n}{x}\left[\frac{K}{M}\right]^x\left[1-\frac{K}{M}\right]^{n-x} \quad \text{for} \quad x = 0, 1, \ldots, n, \tag{6}$$

which is the same as $P[A_k]$ in Eq. (3) of Subsec. 3.5 of Chap. I, for $x = k$. ////

The sketches in Fig. 3 seem to indicate that the terms $f_X(x; n, p)$ increase monotonically and then decrease monotonically. The following theorem states that such is indeed the case.

2.3 Hypergeometric Distribution

Definition 4 Hypergeometric distribution A random variable X is defined to have a *hypergeometric distribution* if the discrete density function of X is given by

$$f_X(x; M, K, n) = \begin{cases} \dfrac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} & \text{for} \quad x = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$$= \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} I_{\{0, 1, \ldots, n\}}(x)$$

(7)==

where M is a positive integer, K is a nonnegative integer that is at most M, and n is a positive integer that is at most M. Any distribution function defined by the density function given in Eq. (7) above is called a *hypergeometric distribution*. ////

Theorem 5 If X is a hypergeometric distribution, then

$$\mathscr{E}[X] = n \cdot \frac{K}{M} \qquad \text{and} \qquad \operatorname{var}[X] = n \cdot \frac{K}{M} \cdot \frac{M-K}{M} \cdot \frac{M-n}{M-1} \tag{8}$$

PROOF

$$\mathscr{E}[X] = \sum_{x=0}^{n} x \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} = n \cdot \frac{K}{M} \sum_{x=1}^{n} \frac{\binom{K-1}{x-1}\binom{M-K}{n-x}}{\binom{M-1}{n-1}}$$

$$= n \cdot \frac{K}{M} \sum_{y=0}^{n-1} \frac{\binom{K-1}{y}\binom{M-1-K+1}{n-1-y}}{\binom{M-1}{n-1}}$$

$$= n \cdot \frac{K}{M},$$

using $\sum_{i=0}^{m} \binom{a}{i}\binom{b}{m-i} = \binom{a+b}{m}$ given in Appendix A.

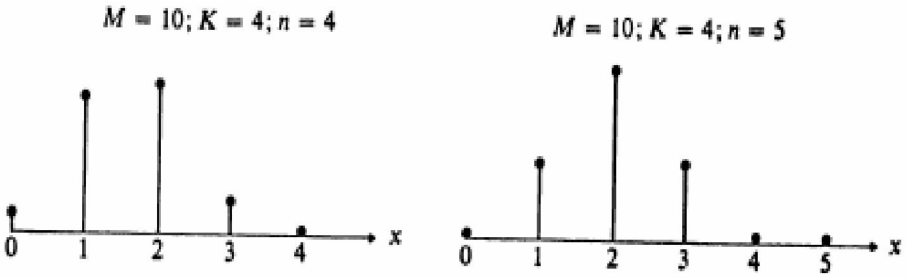

FIGURE 4
Hypergeometric densities.

(8)===

$$\mathscr{E}[X(X-1)]$$

$$= \sum_{x=0}^{n} x(x-1) \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}}$$

$$= n(n-1)\frac{K(K-1)}{M(M-1)} \sum_{x=2}^{n} \frac{\binom{K-2}{x-2}\binom{M-K}{n-x}}{\binom{M-2}{n-2}}$$

$$= n(n-1)\frac{K(K-1)}{M(M-1)} \sum_{y=0}^{n-2} \frac{\binom{K-2}{y}\binom{M-2-K+2}{n-2-y}}{\binom{M-2}{n-2}} = n(n-1)\frac{K(K-1)}{M(M-1)}.$$

Hence

$$\operatorname{var}[X] = \mathscr{E}[X^2] - (\mathscr{E}[X])^2 = \mathscr{E}[X(X-1)] + \mathscr{E}[X] - (\mathscr{E}[X])^2$$

$$= n(n-1)\frac{K(K-1)}{M(M-1)} + n\frac{K}{M} - n^2\frac{K^2}{M^2}$$

$$= n\frac{K}{M}\left[(n-1)\frac{K-1}{M-1} + 1 - \frac{nK}{M}\right]$$

$$= \frac{nK}{M}\left[\frac{(M-K)(M-n)}{M(M-1)}\right].$$ ////

Remark If we set $K/M = p$, then the mean of the hypergeometric distribution coincides with the mean of the binomial distribution, and the variance of the hypergeometric distribution is $(M-n)/(M-1)$ times the variance of the binomial distribution. ////