The Mean and Variance for a Binomial Random Variable

Let n be a positive integer, let p be a real number satisfying $0 \le p \le 1$, and let $q = 1 - p$. Let X count the number of successes in n independent trials where each trial has probability p of success. The possible values for X are $0, 1, 2, \ldots, n-1, n$. We have shown in class that the *pdf* for X is

$$f(k) = \Pr(X = k) = \binom{n}{k} p^k q^{n-k} \quad \text{for } k = 0, 1, 2, \ldots, n.$$

Recall from algebra that the binomial expansion formula is

$$(x+y)^m = \sum_{j=0}^{m} \binom{m}{j} x^j y^{m-j}$$

for each nonnegative integer m. Consequently,

$$\sum_{all\ \ k} f(k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = (1)^n = 1,$$

as expected for a *pdf*.

The mean of X is given by

$$\begin{aligned} \mu &= E(X) = \sum_{all\ \ k} k \cdot f(k) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} \end{aligned}$$

since the $k = 0$ term is 0. Thus,

$$\begin{aligned} \mu &= \sum_{k=1}^{n} k \cdot \frac{n!}{k!(n-k)!} p^{k-1} p q^{n-k} = p \sum_{k=1}^{n} k \cdot \frac{n \cdot (n-1)!}{k \cdot (k-1)!(n-k)!} p^{k-1} q^{n-k} \\ &= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} \ . \end{aligned}$$

Recognizing that $n - k = (n-1) - (k-1)$,

$$\begin{aligned} \mu &= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\,[(n-1)-(k-1)]!} p^{k-1} q^{(n-1)-(k-1)} \\ &= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} \ . \end{aligned}$$

Let $j = k - 1$. Then as k ranges from 1 to n, it follows that j ranges from 0 to $n-1$. (Here we are using the fact that n is positive, and hence that $n - 1 \geq 0$.) Substituting,

$$\mu = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{(n-1)-j} = np\,(p+q)^{n-1} = np(1)^{n-1} = np.$$

We turn next to determining σ^2 for X. We proved in class that if $n = 1$, then $\sigma^2 = pq$, so we will assume that $n \geq 2$. Recall that one handy trick for computing $\sigma^2 = E\left((X-\mu)^2\right)$ is the fact that $E\left((X-\mu)^2\right) = E(X^2) - \mu^2$. Here we will use an alternative trick. Specifically, $E\left(X\left(X-1\right)\right) = E\left(X^2 - X\right) = E(X^2) - \mu$, so an alternative expression for $E(X^2)$ is $E\left(X\left(X-1\right)\right) + \mu$. Thus

$$\sigma^2 = E\left(X\left(X-1\right)\right) + \mu - \mu^2.$$

Since we know for a binomial random variable that $\mu = np$, all that remains to be computed is $E\left(X\left(X-1\right)\right)$.

$$\begin{aligned} E\left(X\left(X-1\right)\right) &= \sum_{all\ \ k} k(k-1)\cdot f(k) = \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k q^{n-k} \end{aligned}$$

since $k(k-1) = 0$ for $k = 0$ and $k = 1$. Then proceeding as before,

$$\begin{aligned} E\left(X\left(X-1\right)\right) &= \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k q^{n-k} = \sum_{k=2}^{n} k(k-1)\frac{n!}{k!(n-k)!} p^k q^{n-k} \\ &= \sum_{k=2}^{n} k(k-1)\frac{n(n-1)\cdot(n-2)!}{k(k-1)\cdot(k-2)!(n-k)!} p^{k-2} p^2 q^{n-k} \\ &= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} q^{n-k}\ . \end{aligned}$$

Recognizing that $(n-k) = (n-2) - (k-2)$,

$$\begin{aligned} E\left(X\left(X-1\right)\right) &= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!\left[(n-2)-(k-2)\right]!} p^{k-2} q^{(n-2)-(k-2)} \\ &= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} q^{(n-2)-(k-2)} \end{aligned}$$

Let $j = k - 2$. Then as k ranges from 2 to n, it follows that j ranges from 0 to

$n-2$. (Notice we are using the fact that $n \geq 2$ here.) Substituting,

$$\begin{aligned} E\left(X\left(X-1\right)\right) &= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j q^{(n-2)-j} \\ &= n(n-1)p^2 \left(p+q\right)^{n-2} = n(n-1)p^2 \left(1\right)^{n-2} \\ &= n(n-1)p^2. \end{aligned}$$

Finally,

$$\begin{aligned} \sigma^2 &= E\left(X\left(X-1\right)\right) + \mu - \mu^2 \\ &= n(n-1)p^2 + np - \left(np\right)^2 \\ &= n^2p^2 - np^2 + np - n^2p^2 \\ &= np - np^2 \\ &= np(1-p) = npq. \end{aligned}$$

Summarizing:

Theorem 1 *Let X be a binomial random variable distributed according to $B(n,p)$ where n is a positive integer, where p is a real number satisfying $0 \leq p \leq 1$, and where $q = 1-p$. Then $\mu = np$ and $\sigma^2 = npq$.*

In a few sections, we will encounter the moment generating function (MGF) for X. This is the function

$$M(t) = E(e^{tX})$$

defined for all t near $t = 0$. That is,

$$\begin{aligned} M(t) &= E(e^{tX}) = \sum_{all\ \ k} e^{tk} \cdot f(k) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} e^{tk} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \left(e^t p\right)^k q^{n-k} \\ &= \left(e^t p + q\right)^n \end{aligned}$$

Summarizing,

Theorem 2 *Let X be a binomial random variable distributed according to $B(n,p)$ where n is a positive integer, where p is a real number satisfying $0 \leq p \leq 1$, and where $q = 1-p$. Then the moment generating function for X is $M(t) = (e^t p + q)^n$ for all t near $t = 0$.*