The normal approximation to the binomial distribution.

As usual, let X denote the number of successes in n trials of a binomial experiment, with $P(\text{success}) = p$.

We've seen that:

(1) $\mu = E(X) = np$,

(2) $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (0 \le k \le n)$;

(3) $\sigma^2 = var(X) = np(1-p)$;
$\sigma = std(X) = \sqrt{np(1-p)}$.

Example 1.

Let X be the number of successes in 6 trials of a binomial experiment, with $P(\text{success}) = 0.4$.

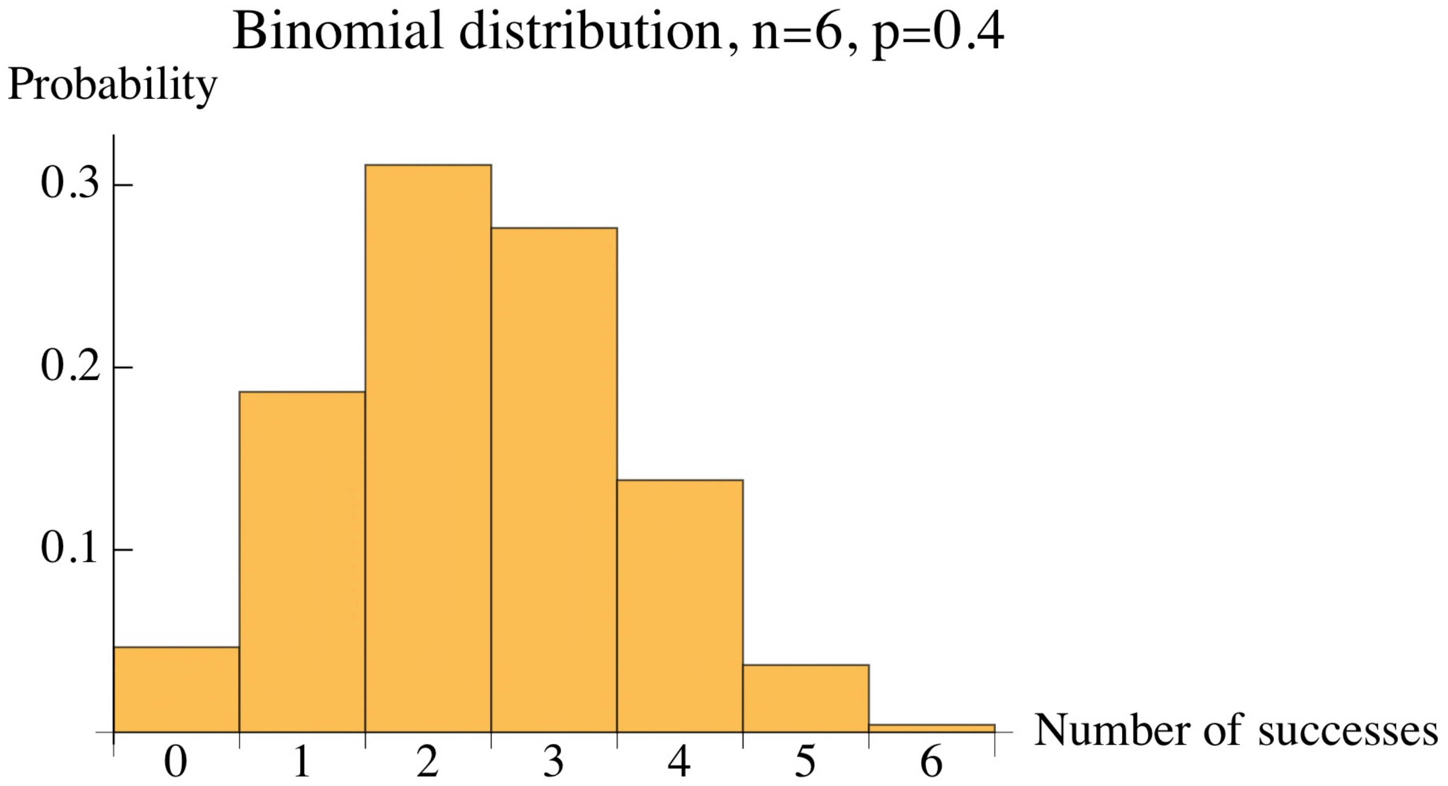
Find the probability mass function for X; find $E(X)$, $var(X)$, $std(X)$.

Solution

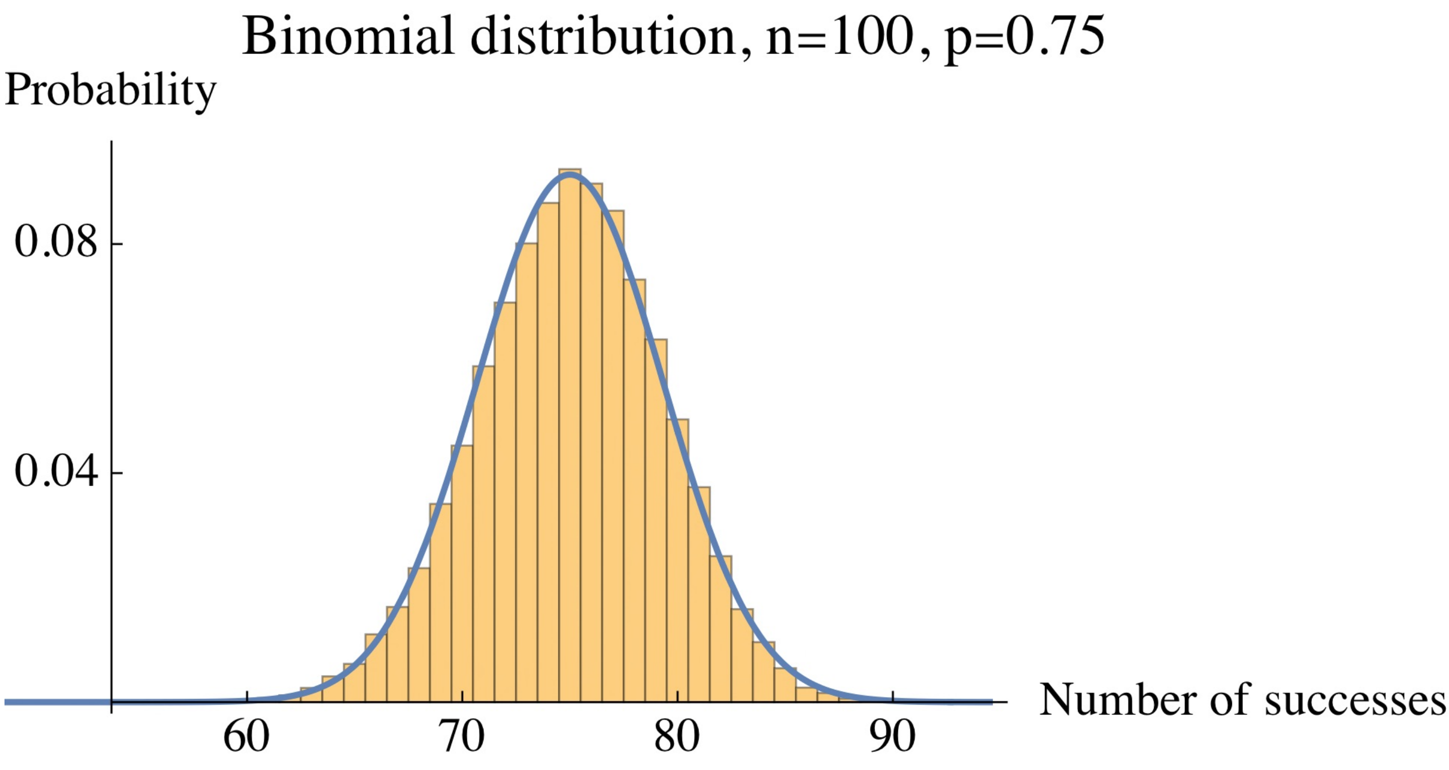
$P(X=0) = \binom{6}{0} \cdot 0.4^0 \cdot 0.6^6 = 0.047$
$P(X=1) = \binom{6}{1} \cdot 0.4^1 \cdot 0.6^5 = 0.187$
$P(X=2) = \binom{6}{2} \cdot 0.4^2 \cdot 0.6^4 = 0.311$
$P(X=3) = \quad \cdots \quad = 0.276$
$P(X=4) = \quad \cdots \quad = 0.138$
$P(X=5) = \quad \cdots \quad = 0.037$
$P(X=6) = \quad \cdots \quad = 0.004$

Also $E(X) = 6 \cdot 0.4 = 2.4$,
$var(X) = 6 \cdot 0.4 \cdot 0.6 = 1.44$,

$std(X) = \sqrt{1.44} = 1.2.$



There's a bit of a (skewed) bell shape.
Now repeat with X = # of successes in 100 trials of a binomial experiment, with $p = 0.75$:



The binomial distribution looks normal!

This illustrates:

<u>Theorem NAB</u> (normal approximation to binomial).
Let X be the number of successes in n trials of a binomial experiment with $P(\text{success}) = p$. Assume $np \geq 10$ and $n(1-p) \geq 10$.

Let Y be $N(np, \sqrt{np(1-p)})$. That is, Y is normal with the same μ and σ as X. Also let j and k be integers, with $0 \leq j, k \leq n$. Then:

(a) $P(X \leq j) \approx P(Y < j + 0.5)$;
(b) $P(X \geq k) \approx P(Y > k - 0.5)$;
(c) $P(j \leq X \leq k) \approx P(k - 0.5 < Y < j + 0.5)$.

Note: the extra factors of 0.5 in Thm. NAB are "continuity correction factors." They help the continuous rv Y better match the discrete rv X.

<u>Example</u>
Let X be the number of successes in 100 trials of a binomial experiment with $P(\text{success}) = 0.75$. Approximate
(a) $P(X \leq 70)$;
(b) $P(X = 75)$.

Use the facts that, if Z is standard normal, then

$$P(Z < -1.039) = 0.1492,$$
$$P(-0.115 < Z < 0.115) = 0.0916$$

Solution.

(a) By Thm. NAB,

$$P(X \le 70) \approx P(Y < 70.5),$$

where Y is $N(np, \sqrt{np(1-p)}) = N(75, 4.33)$.

To compute $P(Y < 70.5)$, we standardize Y:

$$P(Y < 70.5) = P\left(\frac{Y-75}{4.33} < \frac{70.5-75}{4.33}\right)$$
$$= P\left(\frac{Y-75}{4.33} < -1.039\right) = 0.1492,$$

by the given $N(0,1)$ fact.

Similarly,

$$P(X = 75) = P(75 \le X \le 75) \approx P(74.5 < Y < 75.5)$$
$$= P\left(\frac{74.5-75}{4.33} < \frac{Y-75}{4.33} < \frac{75.5-75}{4.33}\right)$$
$$= P\left(-0.115 < \frac{Y-75}{4.33} < 0.115\right) = 0.0916.$$

Note: exact computation of the binomial probabilities gives

$$P(X \le 70) = 0.1495,$$
$$P(X = 75) = 0.0918.$$