Stat 400, section 3.4 The Binomial Probability Distribution

Notes by Tim Pilachowski

Definition of Bernoulli trials which make up a **binomial** experiment:

The number of trials, n, in an experiment is fixed in advance.
There are exactly two events/outcomes for each trial, usually labeled success (S) and failure (F).
Trials are independent from one trial to the next, i.e. the outcome of one trial doesn't affect the next.
$P(S) = p$ must be the same for each trial. [$P(F) = 1 - p$ is often designated as q].

Example A-1. Suppose that a box contains 3 blue blocks and 2 yellow blocks. You pick four blocks ***without replacement***. Define success as "picking a blue block". Is this is a binomial experiment?

number of trials is fixed? $n =$

exactly two events/outcomes for each trial?

trials are independent?

$P(S) = p$ is the same for each trial? $p =$

$P(F) = q =$?

Example A-2. Suppose that a box contains 3 blue blocks and 2 yellow blocks. You pick four blocks ***with replacement***. Define success as "picking a blue block". Is this is a binomial experiment?

number of trials is fixed? $n =$

exactly two events/outcomes for each trial?

trials are independent?

$P(S) = p$ is the same for each trial? $p =$

$P(F) = q =$?

possible outcomes:

Let random variable X = number of successes, i.e. number of blues picked.

x					
$P(X = x)$					

For a binomial distribution – n trials, $P(S) = p$, X = number of successes, $x = 0, 1, 2, \dots, n$

$$P(X = x) = b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} = \dfrac{n!}{x!(n-x)!} p^x q^{n-x} & x = 0, 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

SEMI-IMPORTANT: In the real world, sampling with replacement is not always possible. The rule of thumb is, "If the sample size n is at most 5% of the population size N, then the experiment can be treated as though it were exactly a binomial experiment."

Example A-2 revisited. Suppose that a box contains 3 blue blocks and 2 yellow blocks. You pick four blocks with replacement. Define success as "picking a blue block". X = number of successes

n = p = q =

$P(X = 0)$ =

$P(X = 1)$ =

$P(X = 2)$ =

$P(X = 3)$ =

$P(X = 4)$ =

Example A-2 yet again. Suppose that a box contains 3 blue blocks and 2 yellow blocks. You pick five blocks with replacement. Define success as "picking a blue block". X = number of successes

x	0	1	2	3	4	5
$P(X \le x)$ decimal						

Appendix Table A.1 in your text has cumulative binomial probability distribution tables which provide approximate probabilities for various values of n, p and x. If the numbers you need happen to be there, you can use the tables instead of doing the calculations by hand.

Note that the tables give the *cumulative distribution function* value for random variable $X \sim \text{Bin}(n, p)$,

$$P(X \le x) = B(x; n, p) = \sum_{y=0}^{x} b(x; n, p) \qquad x = 0, 1, 2, \ldots, n$$

If we were to go to the tables, which one would we need?

Example A-2 variation. Suppose that a box contains 3 blue blocks and 2 yellow blocks. You pick fifteen blocks with replacement. Define success as "picking a blue block". X = number of successes

$n =$ $p =$ $q =$

Which binomial probability table will we use?

a) What is the probability of picking between 6 and 10 blue blocks, inclusive, i.e. $P(6 \le X \le 10)$?

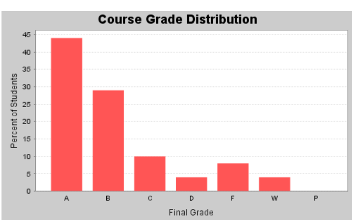
using the cumulative binomial probability distribution table:

b) What is the probability of picking at least 10 blue blocks?

The calculation would look like this:

$$P(X \ge 10) = P(X = 10) + P(X = 11) + P(X = 12) + P(X = 13) + P(X = 14) + P(X = 15)$$
$$= \frac{15!}{10!5!}0.6^{10}0.4^{5} + \frac{15!}{11!4!}0.6^{11}0.4^{4} + \frac{15!}{12!3!}0.6^{12}0.4^{3} + \frac{15!}{13!2!}0.6^{13}0.4^{2} + \frac{15!}{14!1!}0.6^{14}0.4^{1} + \frac{15!}{15!0!}0.6^{15}0.4^{0}$$
$$\approx 0.40321555041484$$

using the cumulative binomial probability distribution table:



Example B. A Math 220 class, taught in the Fall of 2010 at UMCP, had the following grade distribution. Define success as a grade of "C" or better. Pick 15 students at random. What is the probability that between 11 and 14 have a grade of C or better? *answer*: ≈ 0.5090

$n =$ $p =$

$q =$

calculations:

The expected value (mean) of a binomial probability distribution is a simple formula:

$$\text{If } X \sim \text{Bin}(n, p), \text{ then } E(X) = np.$$

It is reasonable to expect that a previously-observed proportion p will still hold for any sample of size n.

Using some extended algebra we can derive a formula for variance of a binomial probability distribution:

$$V(X) = npq = np(1 - p).$$

Then standard deviation is, as before, the square root of variance:

$$\sigma_x = \sqrt{npq} = \sqrt{np(1-p)}\,.$$

Example A-2 yet again. Suppose that a box contains 3 blue blocks and 2 yellow blocks. You pick four blocks with replacement. Define success as "picking a blue block". X = number of successes

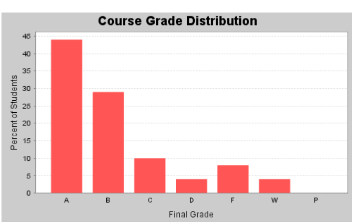
What is the expected number of blue blocks (i.e. the mean)? *answer*: 2.4

What are the variance and standard deviation? *answers*: 0.96, ≈ 0.9798

Example A-2 variation. Suppose that a box contains 3 blue blocks and 2 yellow blocks. You pick fifteen blocks with replacement. Define success as "picking a blue block". X = number of successes

What is the expected number of blue blocks (i.e. the mean)? *answer*: 9

What are the variance and standard deviation? *answers*: 3.6, ≈ 1.8974



Example B revisited. A Math 220 class, taught in the Fall of 2010 at UMCP, had the following grade distribution. Define success as a grade of "C" or better. Pick 15 students at random.

What is the expected number who have a C or better? *answer*: ≈ 12.54

What are the variance and standard deviation? *answers*: ≈ 2.0566, ≈ 1.4341

In Example A-2 above the probability p was theoretical. In Example B we used an empirical observation based on prior experience.

Example C. From prior experience and testing, Shockingly Good, Inc. has determined that 2 out of every 90 spark plugs produced is defective. The company picks 20 spark plugs at random from the production line. Define random variable X = number of good spark plugs.

How large does the population need to be to consider this a binomial experiment?

$p =$ $n =$

a) What is the probability that exactly 1 is defective? *answer*: ≈ 0.2900

b) What is the probability that at most 1 is defective? *answer*: ≈ 0.9280

c) What is the probability that at least 2 are defective? *answer*: ≈ 0.0720

d) When 1800 spark plugs are produced, what is the expected number of number of good spark plugs?
answer: 1760

e) When 1800 spark plugs are produced, what are the variance and standard deviation for X = number of good spark plugs? *answers*: ≈ 39.111, ≈ 6.2539

Appendix Table A-1, Cumulative Binomial Probabilities*

$$P(X \le x) = B(x; n, p) = \sum_{y=0}^{x} b(y; n, p) \qquad x = 0, 1, 2, \ldots, n \;**$$

*This is a partial reproduction. Your text has more values for n and p. I also included a row for $x = n$ (which your text does not include) to emphasize that, for a cumulative distribution like this, $B(x; n, p) = 1$.

** A table entry = 0 does not indicate an impossibility. All table entries are actually positive, but rounded to three significant digits appear as 0.000 .

n	*x*	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**
5	0	0.774	0.590	0.328	0.168	0.078	0.031	0.010	0.002	0.000	0.000	0.000
	1	0.977	0.919	0.737	0.528	0.337	0.188	0.087	0.031	0.007	0.000	0.000
	2	0.999	0.991	0.942	0.837	0.683	0.500	0.317	0.163	0.058	0.009	0.001
	3	1.000	1.000	0.993	0.969	0.913	0.813	0.663	0.472	0.263	0.081	0.023
	4	1.000	1.000	1.000	0.998	0.990	0.969	0.922	0.832	0.672	0.410	0.226
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0	0.599	0.349	0.107	0.028	0.006	0.001	0.000	0.000	0.000	0.000	0.000
	1	0.914	0.736	0.376	0.149	0.046	0.011	0.002	0.000	0.000	0.000	0.000
	2	0.988	0.930	0.678	0.383	0.167	0.055	0.012	0.002	0.000	0.000	0.000
	3	0.999	0.987	0.879	0.650	0.382	0.172	0.055	0.011	0.001	0.000	0.000
	4	1.000	0.998	0.967	0.850	0.633	0.377	0.166	0.047	0.006	0.000	0.000
	5	1.000	1.000	0.994	0.953	0.834	0.623	0.367	0.150	0.033	0.002	0.000
	6	1.000	1.000	0.999	0.989	0.945	0.828	0.618	0.350	0.121	0.013	0.001
	7	1.000	1.000	1.000	0.998	0.988	0.945	0.833	0.617	0.322	0.070	0.012
	8	1.000	1.000	1.000	1.000	0.998	0.989	0.954	0.851	0.624	0.264	0.086
	9	1.000	1.000	1.000	1.000	1.000	0.999	0.994	0.972	0.893	0.651	0.401
	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
15	0	0.463	0.206	0.035	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.829	0.549	0.167	0.035	0.005	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.964	0.816	0.398	0.127	0.027	0.004	0.000	0.000	0.000	0.000	0.000
	3	0.995	0.944	0.648	0.297	0.091	0.018	0.002	0.000	0.000	0.000	0.000
	4	0.999	0.987	0.836	0.515	0.217	0.059	0.009	0.001	0.000	0.000	0.000
	5	1.000	0.998	0.939	0.722	0.403	0.151	0.034	0.004	0.000	0.000	0.000
	6	1.000	1.000	0.982	0.869	0.610	0.304	0.095	0.015	0.001	0.000	0.000
	7	1.000	1.000	0.996	0.950	0.787	0.500	0.213	0.050	0.004	0.000	0.000
	8	1.000	1.000	0.999	0.985	0.905	0.696	0.390	0.131	0.018	0.000	0.000
	9	1.000	1.000	1.000	0.996	0.966	0.849	0.597	0.278	0.061	0.002	0.000
	10	1.000	1.000	1.000	0.999	0.991	0.941	0.783	0.485	0.164	0.013	0.001
	11	1.000	1.000	1.000	1.000	0.998	0.982	0.909	0.703	0.352	0.056	0.005
	12	1.000	1.000	1.000	1.000	1.000	0.996	0.973	0.873	0.602	0.184	0.036
	13	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.965	0.833	0.451	0.171
	14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.965	0.794	0.537
	15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000