Definitions. A **binomial experiment** has the following properties:

1. There is a fixed number, n, of identical trials.
2. Each trial results in one of two mutually exclusive outcomes – success (S) and failure (F). [Such a trial is called a Bernoulli trial.]
3. The probability of success on a single trial is p and remains the same from trial to trial. The probability of a failure on a single trial is q. [Note that $0 \le p, q \le 1$, and $q = 1 - p$ or $p + q = 1$.]
4. The trials are independent.

Then if the random variable Y is the number of successes observed during the n trials, we say that Y has a **binomial distribution with parameters n and p**, and we write $Y \sim \text{bin}(n, p)$.

Y is also called a **Bernoulli random variable** when $n = 1$, and we often write $Y \sim \text{Bernoulli}(p)$.

2

Binomial experiment:

- There are n identical trials.
- Each trial results in success (S) or failure (F).
- $P(S) = p$ and remains the same from trial to trial. $P(F) = q$.
- The trials are independent.

Example 1. Suppose that the probability of germination of a seed of a particular species of beet is 0.8, and that 100 seeds are planted. The number of seeds that germinate is observed.

- Is this a binomial experiment?
- If not, what more is needed?
- If so, what is are n, p, and q?

3

Binomial experiment:

- There are n identical trials.
- Each trial results in success (S) or failure (F).
- $P(S) = p$ and remains the same from trial to trial. $P(F) = q$.
- The trials are independent.

Example 2. Suppose that 40% of a large population of registered voters favor candidate Jones. A random sample of 50 voters will be selected, and Y, the number favoring Jones, is to be observed.

- Is this a binomial experiment?
- If not, what more is needed?
- If so, what are n, p, and q?

4

Binomial experiment:

- There are n identical trials.
- Each trial results in success (S) or failure (F).
- $P(S) = p$ and remains the same from trial to trial. $P(F) = q$.
- The trials are independent.

Example 3. Roll a fair die seven times. Let Y be the number of rolls in which the number showing on the top face is a one.

- Is this a binomial experiment?
- If not, what more is needed?
- If so, what are n, p, and q?

5

Example 3. Seven tosses of fair die; Y = the number of ones.
Find the support of Y and its probability function $p(y)$.

The support of Y is {0, 1, 2, 3, 4, 5, 6, 7}.

Since the tosses are independent trials, the probability of any sequence, $(t_1, t_2, t_3, t_4, t_5, t_6, t_7)$, where each t_i is "one" or "non-one," is

$P(t_1)\,P(t_2)\,P(t_3)\,P(t_4)\,P(t_5)\,P(t_6)\,P(t_7)$, where $P(t_i) = \begin{cases} 1/6 & \text{if } t_i = \text{"one"} \\ 5/6 & \text{if } t_i = \text{"non-one"} \end{cases}$

so that any sequence with y "ones" and $7-y$ "non-ones" has probability

$$\left(\frac{1}{6}\right)^y \left(\frac{5}{6}\right)^{7-y}$$

How many such sequences are there? There are $\binom{7}{y}$ sequences with exactly y "ones." So the probability function of Y is

$$p(y) = P(Y = y) = \begin{cases} \binom{7}{y}\left(\frac{1}{6}\right)^y \left(\frac{5}{6}\right)^{7-y}, & \text{for } y = 0, 1, 2, 3, 4, 5, 6, 7 \\ 0, & \text{elsewhere} \end{cases}$$

6

Definition. A random variable, Y, has a **binomial distribution with parameters n and p** if $p(y) = P(Y = y) = \begin{cases} \binom{n}{y} p^y q^{n-y}, & \text{for } y = 0, 1, 2, \ldots, n \\ 0, & \text{elsewhere} \end{cases}$

where $n \geq 1$, $0 < p < 1$, and $q = 1 - p$.

Question: Does this define a probability distribution? We ***must*** check this for every definition we make. We need to check two things:

(1) $p(y) \geq 0$ for all y (2) $\sum_{y=0}^{n} p(y) = 1$

(1) is clear here since $\binom{n}{y}$, p, and q are all positive.

(2) follows from the binomial theorem (hence the name binomial distr.):

$$\sum_{y=0}^{n} p(y) = \sum_{y=0}^{n} \binom{n}{y} p^y q^{n-y} = (p+q)^n = 1^n = 1.$$

7

Example (p. 111 #3.41, modified): A multiple-choice examination has 15 questions, each with five possible answers, only one of which is correct. Suppose that a student answers each of the questions with an independent random guess. What is the probability that he answers:

a. exactly five questions correctly?
b. at least five questions correctly?

Solution. First restate the question in terms of random variables:
Let the random variable Y be the number of questions answered correctly. Then $Y \sim \text{bin}(15, 0.2)$ and we are asked for

a. $P(Y = 5)$, and
b. $P(Y \geq 5) = p(5) + p(6) + p(7) + \cdots + p(13) + p(14) + p(15)$.

8

Example (p. 111 #3.41 modified): What is the probability that he answers:

a. exactly five questions correctly?

Solution. First restate the question in terms of random variables:
Y = # of questions answered correctly ~ bin(15, 0.2) and we want

a. $P(Y = 5) = \binom{15}{5}(0.2)^5(1-0.2)^{15-5} = \binom{15}{5}(0.2)^5(0.8)^{10} = 0.10318$

We can also use the distributions on the TI-83/84/plus calculators:

2nd | DISTR | 0: binompdf(| 1 | 5 | , | . | 2 | , | 5 |)

or we can use Table 1 on pages 839-841 of the text, which tables $P(Y \leq \boldsymbol{a})$ for various values of $\boldsymbol{n}$ and $\boldsymbol{p}$ and for $\boldsymbol{a}$ between 0 and $\boldsymbol{n}$:

$P(Y = 5) = P(Y \leq 5) - P(Y < 5)$
$= P(Y \leq 5) - P(Y \leq 4)$ since Y only assumes integer values
$= 0.939 - 0.836 = 0.103$

The function $F(y) = P(Y \leq y)$ is the (**cumulative**) **distribution function** or **cdf** of the random variable Y.

9

Table 1. Binomial Probabilities pp. 839-841, Appendix 3

Tabulated values are $P(Y \le a) = \sum_{y=0}^{a} p(y)$. (Computations are rounded at third decimal place.)

(c) $n = 15$ (p. 840)

	p													
a	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	a
0	.860	.463	.206	.035	.005	.000	.000	.000	.000	.000	.000	.000	.000	0
1	.990	.829	.549	.167	.035	.005	.000	.000	.000	.000	.000	.000	.000	1
2	1.000	.964	.816	.398	.127	.027	.004	.000	.000	.000	.000	.000	.000	2
3	1.000	.995	.944	.648	.297	.091	.018	.002	.000	.000	.000	.000	.000	3
4	1.000	.999	.987	.836	.515	.217	.059	.009	.001	.000	.000	.000	.000	4
5	1.000	1.000	.998	.939	.722	.403	.151	.034	.004	.000	.000	.000	.000	5
6	1.000	1.000	1.000	.982	.869	.610	.304	.095	.015	.001	.000	.000	.000	6
7	1.000	1.000	1.000	.996	.950	.787	.500	.213	.050	.004	.000	.000	.000	7
8	1.000	1.000	1.000	.999	.985	.905	.696	.390	.131	.018	.000	.000	.000	8
9	1.000	1.000	1.000	1.000	.996	.966	.849	.597	.278	.061	.002	.000	.000	9
10	1.000	1.000	1.000	1.000	.999	.991	.941	.783	.485	.164	.013	.001	.000	10
11	1.000	1.000	1.000	1.000	1.000	.998	.982	.909	.703	.352	.056	.005	.000	11
12	1.000	1.000	1.000	1.000	1.000	1.000	.996	.973	.873	.602	.184	.036	.000	12
13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.833	.451	.171	.010	13
14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.794	.537	.140	14

10

Example (p. 111 #3.41 modified): What is the probability that he answers:
b. at least five questions correctly?

Solution. First restate the question in terms of random variables:
Y = # of questions answered correctly ~ bin(15, 0.2) and we want
$P(Y \ge 5) = p(5) + p(6) + p(7) + \cdots + p(13) + p(14) + p(15)$.

Rather than evaluate this, we can use the TI-83/84 or Table 1.
For both, recall that since Y only assumes integer values,

$$P(Y \ge 5) = 1 - P(Y < 5) = 1 - P(Y \le 4).$$

Table 1: $P(Y \le 4) = 0.836$, so $P(Y \ge 5) = 1 - P(Y \le 4)$
$= 1 - 0.836 = 0.164$.

TI-83/84/plus:
1 – 2nd DISTR A: binomcdf(1 5 , . 2 , 4)

Obtain $P(Y \ge 5) = 0.1642337246$.

11

Mean of a Binomial Random Variable, $Y \sim \text{bin}(n, p)$

$$\mu = E(Y) = \sum_{y=0}^{n} y p(y) \quad \text{by the definition of } E(Y)$$

$$= \sum_{y=0}^{n} y \binom{n}{y} p^y q^{n-y} \quad \text{by the definition of } p(y)$$

$$= \sum_{y=1}^{n} \frac{y \cdot n!}{y!(n-y)!} p^y q^{n-y} \quad \text{omit term with } y = 0; \ \binom{n}{y} = \frac{n!}{y!(n-y)!}$$

$$= \sum_{y=1}^{n} \frac{n!}{(y-1)!(n-y)!} p^y q^{n-y} \quad \text{write } y! = y(y-1)! \text{ and cancel } y$$

This looks similar to the sum that we had when showing that $\sum p(y) = 1$, except that the sum starts at $y = 1$, not at $y = 0$, and the coefficients of the terms are not binomial coefficients.

So, we'll change the index of summation and manipulate the coefficients to obtain the sum over its support of the probabilities of another binomial distribution, forcing the resulting sum to be 1.

12

Mean of a Binomial Random Variable, $Y \sim \text{bin}(n, p)$

$$\mu = E(Y) = \sum_{y=1}^{n} \frac{n!}{(y-1)!(n-y)!} p^y q^{n-y}$$

$$= \sum_{x=0}^{n-1} \frac{n!}{x!\,(n-(x+1))!} p^{x+1} q^{n-(x+1)} \quad \longleftarrow \text{substituting } x = y-1, \text{so } y = x+1$$

$$= \sum_{x=0}^{n-1} \frac{n!}{x!\,(n-1-x)!} p^{x+1} q^{n-1-x}$$

$$= np \sum_{x=0}^{n-1} \frac{(n-1)!}{x!\,((n-1)-x)!} p^x q^{(n-1)-x} \quad \longleftarrow \text{factoring } n \text{ and } p \text{ out of the sum}$$

$$= np \sum_{x=0}^{n-1} \binom{n-1}{x} p^x q^{(n-1)-x} \quad \longleftarrow \text{since } \binom{n-1}{x} = \frac{(n-1)!}{x!\,((n-1)-x)!}$$

$$= np \sum_{x=0}^{n-1} p_X(x) \quad \longleftarrow \ p_X(x) = \binom{n-1}{x} p^x q^{(n-1)-x}, \text{ the p.f. of the bin}(n\text{-}1, p) \text{ distribution}$$

$$= np \cdot 1 \quad \longleftarrow \ \sum_{x=0}^{n-1} x p_X(x) = 1 \quad \text{since we are summing a probability function over all the values in its support}$$

$$= np$$

13

Mean and Variance of a Binomial Random Variable, $Y \sim \text{bin}(n, p)$

Thus, if $Y \sim \text{bin}(n, p)$, $\mu = E(Y) = n\,p$.

A similar (but more complicated) trick can be used to show that

$$\sigma^2 = V(Y) = n\,p\,q = n\,p\,(1 - p),$$

but it will be easier using some later results. So, for now we just assume it. (The text derives it using the more complicated trick.)

Also remember the **computing formula for variance**

$$V(Y) = E(Y^2) - [\,E(Y)\,]^2$$

which implies the formula

$$E(Y^2) = V(Y) + [\,E(Y)\,]^2$$

both are **extremely useful**.

Example (p. 113 #3.58) A sale involves four items randomly selected from a large lot known to contain 10% defectives. Let Y denote the number of defective items among the four sold. The purchaser of the items will return the defectives for repair, and the repair cost is given by the formula

$$C = 3Y^2 + Y + 2.$$

Find the expected repair cost. [Answer: 3.96]

Solution. $Y \approx \text{bin}(4, 0.1)$.

[This is only approximate, since each time an item is selected for sale, the proportion of defectives changes. However, since the lot is "large," the change is insignificant, and the proportion of defectives remains approximately 10%. So for all practical purposes, Y can be considered to be a binomial random variable.]

Example (p. 113 #3.58) A sale involves four items randomly selected from a large lot known to contain 10% defectives. Let Y denote the number of defective items among the four sold. The purchaser of the items will return the defectives for repair, and the repair cost is given by the formula

$$C = 3Y^2 + Y + 2.$$

Find the expected repair cost. [Answer: 3.96]

Solution. $Y \approx \text{bin}(4, 0.1)$.

Since $C = 3Y^2 + Y + 2$,

$$E(C) = E(3Y^2 + Y + 2) = 3E(Y^2) + E(Y) + 2.$$

We know $E(Y) = np = (4)(0.1) = 0.4$. But how do we find $E(Y^2)$?

Solve the computing formula for variance, $V(Y) = E(Y^2) - [\,E(Y)\,]^2$,

for $E(Y^2) = V(Y) + [\,E(Y)\,]^2 = (4)(0.1)\,(0.9) + (0.4)^2 = 0.52$.

Therefore, $E(C) = 3E(Y^2) + E(Y) + 2 = 1.56 + 0.4 + 2 = 3.96$.

Distribution Functions

DISTR menu

To display the DISTR menu, press [2nd] [DISTR].

DISTR DRAW	
1:normalpdf(	Normal probability density
2:normalcdf(	Normal distribution probability
3:invNorm(	Inverse cumulative normal distribution
4:tpdf(	Student-t probability density
5:tcdf(	Student-t distribution probability
6:χ^2pdf(	Chi-square probability density
7:χ^2cdf	Chi-square distribution probability
8:Fpdf(	F probability density
9:Fcdf(	F distribution probability
0:binompdf(	Binomial probability
A:binomcdf(	Binomial cumulative density
B:poissonpdf(	Poisson probability
C:poissoncdf(	Poisson cumulative density
D:geometpdf(	Geometric probability
E:geometcdf(	Geometric cumulative density

Note: −1E99 and 1E99 specify infinity. If you want to view the area left of *upperbound*, for example, specify *lowerbound*=−1E99.

From the TI-83 Manual.

The TI-84 inserts invT([the Inverse cumulative Student t-distribution] at 4 and moves the rest down one.

The manuals are on TI's site at
http://education.ti.com/educationportal/sites/US/homePage/index.html