5 The Binomial Distribution

The binomial distribution plays a very important role in many life science problems. In order to develop this distribution, now we look at a related distribution called Bernouilli distribution.

5.1 Bernoulli Distribution (P.43)

Many life science experiments result in responses which have only two possible outcomes "success" (S) and "failure" (F). Such responses are called *dichotomous*. For example, a doctor is interested to know whether the recent medical examination gives 'positive' or 'negative' result for cancer for his patient.

Examples:
1. Gender of a new born: "boy" (B) or "girl" (G).
2. Result of an experiment: "success" (S) or "failure" (F).
3. Result of an examination: "pass" (P) or "fail" (F).

Definition: A random variable whose responses are dichotomous is called a *Bernoulli random variable.*

Note: In many problems, it is easy to use 1 for "success" (S) and 0 for "failure" (F).

In example 3, let $X = 1$ if the examination mark M is over 50 and 0 otherwise. Then X is the result of dichotomising the random variable M such that

$$X = \begin{cases} 1 \text{ if } M \geq 50, & \text{(or pass/success)} \\ 0 \text{ if } M < 50. & \text{(or fail/failure).} \end{cases}$$

In general $X = 1$ denotes the event of a success (S).

- Let $P(X = 1) = P(\text{Success}) = p$.
- Therefore, clearly $X = 0$ denotes the event of a failure (F) and $P(X = 0) = P(\text{Failure}) = 1 - p$, where $0 \leq p \leq 1$.

Let $p(r) = P(X = r)$. Therefore, the above can be written as

$$p(r) = \begin{cases} p & \text{if } r = 1, \\ 1 - p & \text{if } r = 0. \end{cases}$$

Therefore, the probability distribution of a Bernoulli RV X can be given as

x	1	0
$P(X = x)$	p	$1 - p$

Then

- mean of X is

$$E(X) = 1 \times p + 0 \times (1 - p) = p$$

- variance of X is

$$\text{Var}(X) = p(1 - p)$$

Note: $E(X^2) = 1^2 \times p + 0^2 \times (1 - p) = p$ and therefore,
$\text{Var}(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1 - p)$

5.2 Binomial distribution (P.43-48)

Suppose that we repeat Bernoulli trials n (fixed) times independently under the same conditions. An experiment involving such independent Bernoulli trials is called a *binomial experiment*.

Example: Throw a die 4 times and consider the event of observing an even number. This gives a sequence of *independent trials* each with only two possible outcomes, "even" (S) and "odd" (F).

Definition: Let X denote the total number of "successes" (S) in n *independent trials* with $p = P(S)$ be the probability of a S in each trial. Then X is said to be a *binomial random variable* and has a *binomial distribution* with parameters n and p.

We write $X \sim B(n, p)$ where X can take the values $0, 1, 2, \ldots, n$.

In the die example, $n = 4$ and $p = \frac{1}{2}$ and write $X \sim B(4, \frac{1}{2})$.

Summary:

In a binomial experiment,

1. it consists of a fixed number n of *identical* trials,
2. there are only 2 possible outcomes in each trial, denoted by "S" and "F", and
3. the trials are independent with the same probability of "S" p.

5.2.1 Probability mass function

Consider the following example to understand the probability distribution of the number of successes associated with $X \sim B(n, p)$.

Example: When $n = 4$, it is clear from a tree diagram that there are $2^4 = 16$ possible outcomes altogether. Let X be number of successes (S). All 16 outcomes are given below:

Outcome	x	Probability	Outcome	x	Probability	No. of case
SSSS	4	p^4	FFFF	0	$(1-p)^4$	each 1
SSSF	3	$p^3(1-p)$	SFSS	3	$p^3(1-p)$	
SSFS	3	$p^3(1-p)$	FSSS	3	$p^3(1-p)$	4
SSFF	2	$p^2(1-p)^2$	FFSS	2	$p^2(1-p)^2$	
SFSF	2	$p^2(1-p)^2$	FSFS	2	$p^2(1-p)^2$	
SFFS	2	$p^2(1-p)^2$	FSSF	2	$p^2(1-p)^2$	6
FFFS	1	$p(1-p)^3$	FSFF	1	$p(1-p)^3$	
FFSF	1	$p(1-p)^3$	SFFF	1	$p(1-p)^3$	4

Then the probability distribution of or pmf 9probability mass function) of X is

$$\begin{aligned}
P(X=4) &= p^4 \\
P(X=3) &= 4p^3(1-p) \\
P(X=2) &= 6p^2(1-p)^2 \\
P(X=1) &= 4p(1-p)^3 \\
P(X=0) &= (1-p)^4
\end{aligned}$$

Binomial Coefficients

The number of ways of selecting 2 items from 4 is denoted by $\binom{4}{2}$ and this value is 6. These are called the number of combinations or the *binomial coefficients*. That is, $\binom{4}{2} = 6$ gives the number of *combinations* choosing 2 trials for "S" from 4 trials. This is also denoted by ${}_4C_2$ and read as "4 choose 2".

There are many ways to calculate $\binom{4}{2}$. In a calculator, check for the button ${}_nC_r$ or $\binom{n}{r}$. Press 4 and then ${}_nC_r$ or $\binom{n}{r}$ followed by 2. Do you get 6?

Exercise: Find ${}_6C_3$, ${}_7C_2$, and ${}_9C_7$ from your calculator.

Answer: ____________

Return back to any sequence of x "S" and $n - x$ "F", i.e.,

$$\underbrace{S\ S\ \cdots\ S}_{x\ S'\text{s}}\ \underbrace{F\ F\ \cdots\ F}_{(n-x)\ F'\text{s}}.$$

It occurs at the same probability of $p^x(1-p)^{n-x}$. The number of combinations of choosing x trials for "S" from the totally n trials is $\binom{n}{x}$. Therefore, the probability of exactly x successes out of n independent trials is given by

$$P(X = x) = \binom{n}{x} p^x(1-p)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

Clearly, a binomial random variable is a sum of n independent Bernoulli random variables.

If $Y_i, i = 1, 2, \ldots, n$, are independent Bernoulli random variables with parameter p, then the sum of the Y_i's follows a binomial distribution, that is,

$$X = \sum_{i=1}^{n} Y_i = Y_1 + Y_2 + \ldots + Y_n \sim B(n, p).$$

When $n = 1$, X follows a Bernoulli distribution with parameter p as a special case of the binomial distribution.

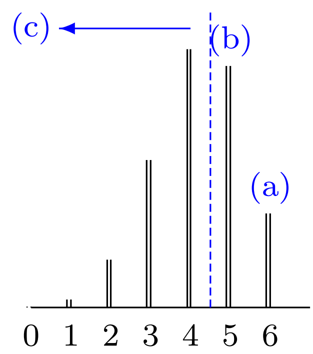
Example: A biologist estimates that the chance of germination for a type of bean seed is 0.7. A student was given 6 seeds. Let X be the number of seeds germinated from 6 seeds. Assuming that the germination of seeds are independent, explain why the distribution of X is binomial. What are the values of n and p? What are the probabilities that he gets

(a) all seeds germinated,

(b) just one seed not germinated, and

(c) at most four seeds germinated?

Solution: Since the germination of 6 seeds are *independent* and the outcome is *binary*, germinated or not, with the *same* probability of germination being 0.7, the distribution of X is binomial, i.e. $X \sim B(6, 0.7)$ with $n = 6$ and $p = 0.7$.



(a) $P(X = 6)$ = ____________________

(b) $P(X = 5)$ = ____________________

(c) $P(X \leq 4)$ = ____________

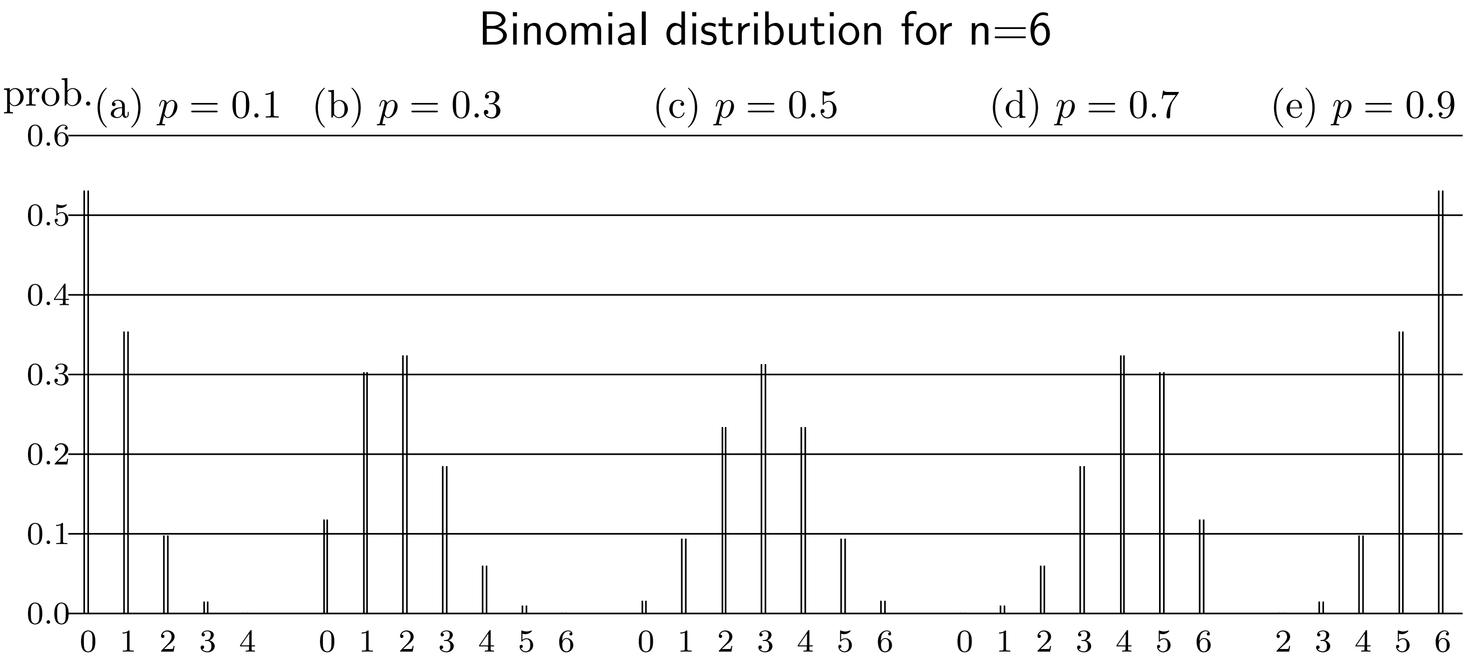
= ______________________________

= _________________________

Exercise: Book P.55, Q2.

5.3 Binomial plot

A binomial distribution can be plotted using a bar chart. The following plots show different binomial distributions when p varies.



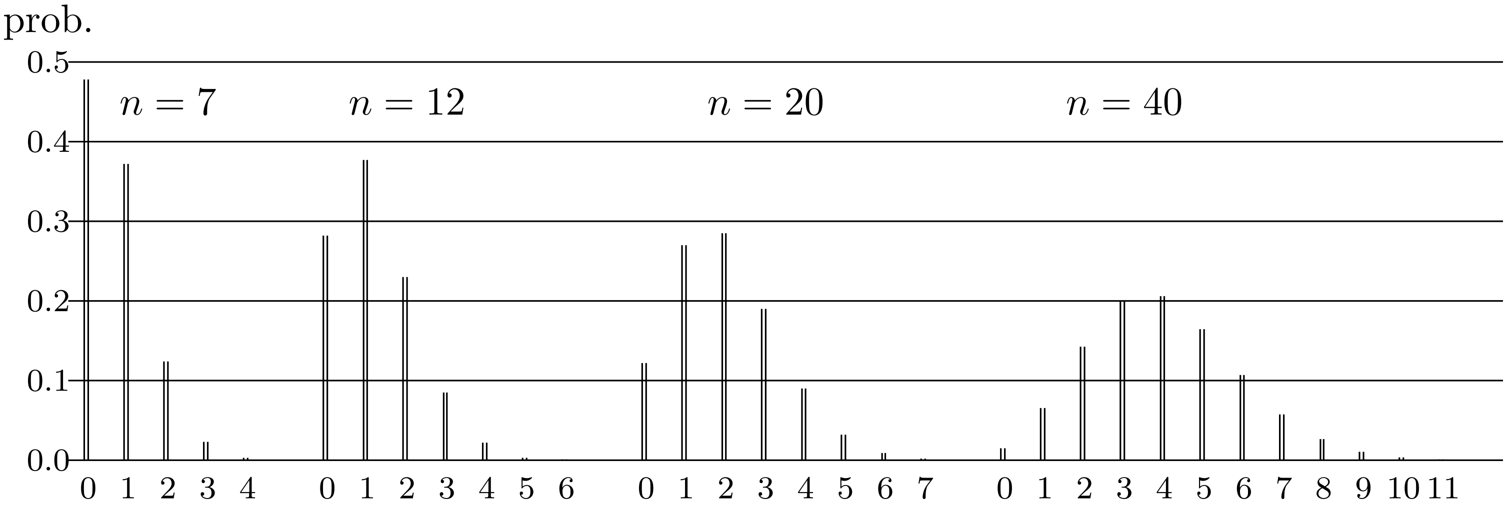
1. Plot (c) is symmetric. Plots (a) and (b) are skewed (have a heavy tail) to the right since p is low and so low values occur at higher probabilities and high values occur at lower probabilities. On the other hand, plots (d) and (e) are skewed to the left.

2. Plots (b) & (d) and (a) & (e) are mirror image to each other as their p sum to 1.

The following gives another set of plots when $p = 0.1$ and n increases. Clearly when n is small, the distribution is skewed but it becomes *more symmetric* as n increases.

Binomial distribution for p=0.1

As $n \to \infty$, the distribution shifts to the right, spreads out, and gradually approaches a symmetric distribution, even though p=0.1 is small.



5.4 Use of table and computer (P.47-48)

Example: A pharmaceutical firm has discovered a new diagnostic test that has 90% chance to indicate a positive result for a patient who are infected by a certain disease. If it is tried on 5 infected patients, find the probability that 4 will be detected.

Solution: Let X be the number of patients out of 5 infected patients who are diagnosed by the test. Since the results are *independent* across infected patients and the outcome is *binary* with the *same* probability of positive result, the distribution of X is binomial, i.e. $X \sim B(5, 0.9)$ with $n = 5$ and $p = 0.9$.

Clearly there are $\binom{5}{4} = 5$ ways of choosing 4 "S" from 5 patients. Therefore,

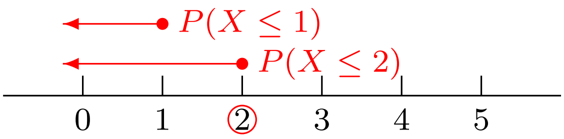
$$P(X = 4) = \underline{\hspace{6cm}}$$

Use of computer: In R,

`pbinom(x,n,p)` gives $\Pr(X \leq x)$, and
`dbinom(x,n,p)` gives $\Pr(X = x)$.

Attend the tutorial for the use of R.

Use of binomial table: It gives $P(X \leq x)$ for $n = 2, \ldots, 12$ and $p = 0.1, 0.2, \ldots, 0.9$. For example, using the binomial tables with $n = 5$, $p = 0.4$ and $x = 1, 2$, we have



$P(X = 2) = \underline{\hspace{10cm}}$

Table 1: Binomial Distribution Table

Percentage point $P(X \leq x)$ for binomial distribution with parameters n and p.
Blank entries are 0.0000 or 1.0000 as appropriate.

p		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
n	x									
2	0	0.8100	0.6400	0.4900	0.3600	0.2500	0.1600	0.0900	0.0400	0.0100
	1	0.9900	0.9600	0.9100	0.8400	0.7500	0.6400	0.5100	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.7290	0.5120	0.3430	0.2160	0.1250	0.0640	0.0270	0.0080	0.0010
	1	0.9720	0.8960	0.7840	0.6480	0.5000	0.3520	0.2160	0.1040	0.0280
	2	0.9990	0.9920	0.9730	0.9360	0.8750	0.7840	0.6570	0.4880	0.2710
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.2401	0.1296	0.0625	0.0256	0.0081	0.0016	0.0001
	1	0.9477	0.8192	0.6517	0.4752	0.3125	0.1792	0.0837	0.0272	0.0037
	2	0.9963	0.9728	0.9163	0.8208	0.6875	0.5248	0.3483	0.1808	0.0523
	3	0.9999	0.9984	0.9919	0.9744	0.9375	0.8704	0.7599	0.5904	0.3439
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.5905	0.3277	0.1681	0.0778	0.0313	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.5314	0.2621	0.1176	0.0467	0.0156	0.0041	0.0007	0.0001	0.0000

Exercise: Let $X \sim B(5, 0.9)$. Find
(a) $P(X \leq 4)$; (b) $P(X = 4)$.

Solution: From the binomial table with $n = 5$, $p = 0.9$, $x = 3, 4$,

$$\begin{aligned}
\text{(a) } P(X \leq 4) &= ______ \\
\text{(b) } P(X = 4) &= ________________ \\
&= __________________
\end{aligned}$$

Exercises:

1. Check if $X \sim B(7, 0.2)$, $P(X = 3) = 0.1149$;
2. Check if $X \sim B(11, 0.1)$, $P(X = 4) = 0.0158$;

Example: The proportion of students wearing spectacles is 40%. Let X be the number of students wearing spectacles in a random sample of 10 students. Use the tables to find

(a) $P(X \leq 2)$; (b) $P(2 \leq X < 5)$; (c) $P(X > 2)$;
(d) $P(X = 3)$; (e) a such that $P(X \geq a) = 0.1662$.

Solution: We have $X \sim B(10, 0.4)$ with $n =$ ___ and $p =$ ___.

Table 1: Binomial Distribution Table (cont.)

Percentage point $P(X \leq r)$ for binomial distribution with parameters n and p.
Blank entries are 0.0000 or 1.0000 as appropriate.

p		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
n	r									
10	0	0.3487	0.1074	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000
	1	0.7361	0.3758	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000
	2	0.9298	0.6778	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	0.0000
	3	0.9872	0.8791	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.8497	0.6331	0.3770	0.1662	0.0473	0.0064	0.0001
	5	0.9999	0.9936	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7		0.9999	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8		1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9			1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10				1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

(a) $P(X \leq 2)$ = ______

(b) $P(2 \leq X < 5)$ = ______

= ______

(c) $P(X > 2)$ = ______

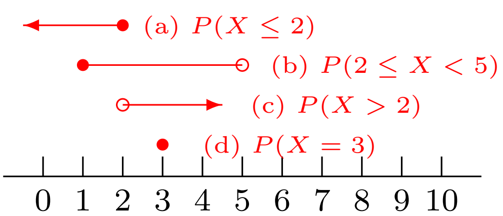
(d) $P(X = 3)$ = ______

= ______

(e) $P(X \geq a)$ = 0.1662

$\Leftrightarrow$ ______ = ______

$\Leftrightarrow$ ______ = ______



Read Examples on P.49-50.
Exercises: Book P.55 Q4 and P.56 Q6.

5.5 Mean and variance of binomial distribution (P.51-52)

Theorem: If $X \sim B(n,p)$, then the mean and variance of X are given by

$$\mu = \mathrm{E}(X) = np \qquad \text{and} \qquad \sigma^2 = \mathrm{Var}(X) = np(1-p)$$

because $X = Y_1 + Y_2 + \cdots + Y_n$ is the sum of n independent Bernoulli r.v. Y_i each with $\mathrm{E}(Y_i) = p$ and $\mathrm{Var}(Y_i) = p(1-p)$.

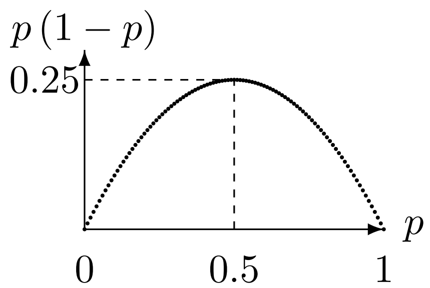
Example: Let $X \sim B(8, 0.60)$. Find $\mathrm{E}(X)$, $\mathrm{Var}(X)$ and $\mathrm{SD}(X)$.

Solution: We have $n = 8$ and $p = 0.6$.

$\mathrm{E}(X) =$ ______________________

$\mathrm{Var}(X) =$ ______________________________________

$\mathrm{SD}(X) =$ __________________



Note that $\mathrm{Var}(X) = np(1-p)$ increases from $p > 0$ and attains its maximum at $p = 0.5$ for a given n since the uncertainty is greatest when the success and failure are equally likely. $\mathrm{Var}(X)$ also increases with n for a given p.

The sum of 2 binomial random variables:

If $X_1 \sim B(n_1, p)$ & $X_2 \sim B(n_2, p)$, $X = X_1 + X_2 \sim B(n_1 + n_2, p)$.

since $X = \underbrace{Y_1 + \cdots + Y_{n_1+1}}_{n_1;\ \text{sum}=X_1 \sim B(n_1,p)} + \underbrace{Y_{n_1+1} + \cdots + Y_n}_{n_2;\ \text{sum}=X_2 \sim B(n_2,p)}$. Note that this result does *NOT* apply if p differs in X_1 and X_2.

Example: Let $X_1 \sim B(5, 0.4)$, $X_2 \sim B(7, 0.4)$ and $X_3 \sim B(7, 0.2)$. Find the distributions of $X_1 + X_2$ and $X_1 + X_3$.

Solution: $X_1 + X_2 \sim$ ____________________ but the distribution of $X_1 + X_3$ is ________

Example: (Soft drinks) Two rival soft drinks, C and P *taste the same*. In a blindfold test, 12 people are asked (independently) to state their preference for one or the other.

(a) What is the probability that the majority prefer P?
(b) How many people out of 12 people would prefer P?

Solution: Let X denote the number of people who prefer P out of 12 people. We have $X \sim B(12, 0.5)$ with $n =$ __ and $p =$ ___.

(a) $P(X \geq 7)$

= __________

=

____

=

____

= _______________ (Table with $n = 12,\ p = 0.5,\ x = 6$)

(b) $E(X)$ = ______________

Example: It is known that the computer disks produced by a company are defective with probability 0.02 independently of each other. Disks are sold in packs of 10. A money back guarantee is offered if a pack contains more than 1 defective disk.

(a) What proportion of sales result in the customers getting their money back?
(b) In a stock of 100 packs, how many packs are expected to result in the customers getting their money back?

Solution:
(a) Let X denote the number of defective disks in a pack of 10. We have $n =$ __ and $p =$____.
(b) Let Y denote the number of packs in a stock of 100 packs that result in customers getting their money back. We have $n' =$ ___ and $p =$ _________.

(a) $P(X > 1)$ = ____________

= ______________________________________

= __________________

(b) $E(Y)$ = _______________________________

Exercise: Book P.55 Q5 and P56 Q9.