

SUMMARY: Discrete Random Variables (with Binomial Distribution)

Approach to answering “DRV” Questions

Example:

For a game, three dice are thrown. If ‘1’ or ‘6’ appears you will be paid \$1, but if neither ‘1’ nor ‘6’ appears, you will pay \$5.

Find the probability distribution of the amount that you can earn in 1 game.

How much would you expect to lose in nine of these games?

Solution:

Let the random variable X denote the amount earned in one game.

Step 1:
Define the random variable.

x	-5	1
$P(X = x)$	$\left(\frac{4}{6}\right)\left(\frac{4}{6}\right)\left(\frac{4}{6}\right) = \frac{8}{27}$	$1 - \frac{8}{27} = \frac{19}{27}$

Step 2:
Identify possible outcomes.

Step 3:
Calculate probability of each outcome.

$$E(X) = \sum x P(x = X) = -5 \times \frac{8}{27} + 1 \times \frac{19}{27} = -\frac{7}{9}$$

$$E(X_1 + X_2 + \dots + X_9) = E(X) \times 9 = -7$$

Step 4:
Use probability distribution to answer subsequent parts.

Since the expectation is negative, I am expected to lose \$7 in nine games.

Expectation and Variance

Expectation (or mean or expected value) of X (usually denoted by μ)	$E(X) = \sum_{\text{all } x} x P(X = x)$ If $g(X)$ is any function of the discrete random variable X, then $E(g(X)) = \sum_{\text{all } x} [g(x) P(X = x)].$
Variance of X (usually denoted by σ^2)	$\text{Var}(X) = E[(X - \mu)^2] = \sum_{\text{all } x} (x - \mu)^2 P(X = x)$ We usually use $\text{Var}(X) = E(X^2) - [E(X)]^2$.
Standard deviation of X (usually denoted by σ)	It is the positive root of the variance of X. $\sigma = \sqrt{\text{Var}(X)}$

If a, b , and $a_1, a_2, \dots, a_n$ are real constants, and X and Y denote independent discrete random variables (with $X_1, X_2, \dots, X_n$ denoting repeated independent observations of X), then

(a) $E(a) = a$ (b) $E(aX) = aE(X)$ (c) $E(aX \pm b) = aE(X) \pm b$ (d) $E(aX \pm bY) = aE(X) \pm bE(Y)$ (e) $E(a_1X_1 \pm \dots \pm a_nX_n)$ $= (a_1 \pm \dots \pm a_n)E(X)$	(a) $\text{Var}(a) = 0$ (b) $\text{Var}(aX) = a^2\text{Var}(X)$ (c) $\text{Var}(aX \pm b) = a^2\text{Var}(X)$ (d) $\text{Var}(aX \pm bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$ (e) $\text{Var}(a_1X_1 \pm \dots \pm a_nX_n)$ $= (a_1^2 + \dots + a_n^2)\text{Var}(X)$
---	---

Binomial Distribution

If the discrete random variable X represents the number of successes in n trials of an experiment, with each trial being a success with probability p , then we write $X \sim B(n, p)$.

$$E(X) = np, \text{ Var}(X) = np(1-p) \text{ (in MF27).}$$

The probability that there are r number of successes in the n trials is given by

$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r} \text{ (in MF27).}$$

Conditions under which a binomial distribution can be a suitable model:

- (1) The experiment consists of n independent repeated trials.
- (2) The result of each trial is only one of two possible outcomes, labelled “success” or “failure”.
- (3) The probability of “success” of each trial remains constant and is denoted by p .

If the question asks for **assumptions**, you should answer using points (1) and (3) above, in context.

Example of answering in context:

According to the Health Promotion Board, 28% of Primary 1 students in Singapore are short-sighted. A sample of 20 Primary 1 students are randomly selected for a survey.

Explain why a binomial distribution is appropriate to model the number students in the sample who are short-sighted.

Solution:

A binomial distribution is appropriate because

- there are only two possible outcomes: either a student is short-sighted or he is not. (2)
- a randomly chosen student being short-sighted is independent of any other student being short-sighted, since the whole cohort of Primary 1 students in Singapore is large. (1)
- the probability of a student being short-sighted is constant at 0.28. (3)

Usage of GC:

- To find probabilities:
 - (1) Key in binompdf for $P(X = r)$.
 - (2) Key in binomcdf for $P(X \leq r)$.

Example of dealing with inequalities:

If X is a discrete random variable,

$$P(2 < X < 9) = P(3 \leq X \leq 8)$$

$$= P(X \leq 8) - P(X \leq 2)$$

- To find unknowns:

Key in binompdf or binomcdf into **[y=]** and use table or graph to find the unknown:

Given	In G.C, use	Graph/Table
$X \sim B(n, \frac{1}{20})$, $P(X \leq 10) \geq 0.995$	$Y = \text{binomcdf}\left(X, \frac{1}{20}, 10\right)$	Use ‘ table ’ (since n is a positive integer)
$X \sim B(10, p)$, $P(X = 3) = 0.2013$	$Y = \text{binompdf}(10, X, 3)$	Use ‘ graph ’ - set window, Xmin, Xmax, Ymin, Ymax to between 0 and 1 (since p is a probability)
$X \sim B(18, 0.1)$, $P(X < r) > 0.9$	$Y = \text{binomcdf}(18, 0.1, X - 1)$	Use ‘ table ’ (since r is a positive integer)