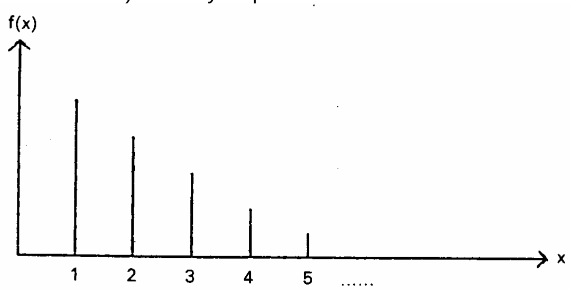
UNIT 20: Random Variables. Discrete and Continuous Probability Distributions

Specific Objectives:

1. To be able to find the expectations and variances of discrete and continuous probability distributions.
2. To learn Binomial and Normal distribution and their daily life applications.
3. To recognize the property of linear combination of independent normal variables.

Detailed Content	Time Ratio	Notes on Teaching
20.1 Random Variables (a) Discrete probability functions	6	A formal treatment of random variable is not expected. Instead, teachers can introduce its preliminary idea by using simple examples such as throwing of coins (for discrete random variable) and life time of electric bulbs (for continuous random variable). Discrete probability function $f(x)$ can be introduced as $f(x)=P(X=x)$ where X is a discrete random variable and x is a fixed value of a random variable through familiar examples such as throwing of 2 coins: $f(x)=\begin{cases}0.25 & \text{for } x=0\\ 0.5 & \text{for } x=1\\ 0.25 & \text{for } x=2\\ 0 & \text{otherwise}\end{cases}$ X (the number of heads obtained) is a discrete random variable which can take the values 0, 1 or 2. Emphasis should be laid on the conditions $f(x)\geq 0$ and $\sum f(x)=1$ Teachers should remind students that capital letter X is usually reserved for random variable and the lower case x for values the random variable can assume. The following is another example of discrete random variable. *Example* X is the number of attempts required to get a 'six' in a throw of a die. The discrete probability function $f(x)$ is $f(x)=P(X=x)=\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{x-1}$. Clearly,

Detailed Content	Time Ratio	Notes on Teaching
		$\sum f(x)=\sum_{x=1}^{\infty}P(X=x)=\frac{\frac{1}{6}}{1-\frac{5}{6}}=1$ Representing the discrete probability function graphically (in the form of bar chart or histogram as shown below) certainly helps students to visualize the concept. 
(b) Probability density functions		At this stage, students should have a clear picture of the discrete probability function. We can extend this idea to continuous random variable and introduce the continuous probability density function (p.d.f.) $f(x)$. Students should note that $f(x)\geq 0$ and $\int_{-\infty}^{\infty}f(x)\,\mathrm{d}x=1$ Students are expected to know that a continuous random variable X can take any value within a specified range and it is related to p.d.f. $f(x)$ by $P(a\leq x<b)=\int_{a}^{b}f(x)\,\mathrm{d}x$

Detailed Content	Time Ratio	Notes on Teaching
		In fact, graphs of p.d.f. can be interpreted as frequency curves of continuous data in statistics. Also students should note that $P(X=a)=0$ and $P(a \le X < b) = P(a < X < b) = P(a < X \le b) = P(a \le X \le b)$ The following are some examples. *Example* 1 (rectangular distribution) The p.d.f. of X is defined as $f(x)=\begin{cases} k & \text{for } 0 < x \le 4 \quad k \text{ is a constant} \\ 0 & \text{otherwise} \end{cases}$ k can be determined from $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$. Also $P(-2 < X \le 1) = \int_0^1 f(x)\,\mathrm{d}x$ and $P(X \ge 3) = 1 - \int_0^3 f(x)\,\mathrm{d}x = \int_3^4 f(x)\,\mathrm{d}x$ Students may be asked to find M in terms of b if P($X \le M$) = b. They should note that when b = 0.5, M is the median. *Example 2* The scheduled time of arrival of a flight to a certain city is 8:00 a.m. However, the actual time of arrival is (8 + X) am, where X is a random variable having the following p.d.f.: $f(x)=\begin{cases} \dfrac{3(4-x^2)}{32} & \text{for } -2 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$ Possible questions include finding the probability that a flight will be between 7:00 a.m. and 8:00 a.m. and between 9:00 a.m. and 10:00 a.m. It is worthwhile to spare some time to discuss with students the meaning of the term cumulative distribution function $\phi(t)$.

Detailed Content	Time Ratio	Notes on Teaching
		$\phi(t) = \sum_{x \le t} f(x)$ in discrete case and $\phi(t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x$ in continuous case Examples such as the two shown below may be used to illustrate these two cases. 1. *Discrete* In a throw of 2 dice, the probability of getting a sum greater than 10 is $1-\phi(10)$. ($\phi(10)$ is the probability that the sum is equal to or smaller than 10.) 2. *Continuous* If $\phi(a)$ denotes the probability that the life time of an electric bulb is smaller than a, then $P(X < a) = \phi(a)$, $P(a < X < b) = \phi(b) - \phi(a)$ and $P(X > a) = 1 - \phi(a)$.
20.2 Expectations and Variances	5	A brief revision on the meaning and physical significance of mean and standard deviation will facilitate students' learning the concepts of expectation. The meaning of expectation can be introduced through simple example such as that shown below. *Example* A man has a probability p = 0.01 of winning a prize x = \$200. We say that his chance is worth px = (\$200)·(0.01) = \$2. Then the teacher can extend this idea to n discrete values of X. Teachers should define the expectation of a discrete random variable ($E(X) = \sum px$) and that of a continuous random variable ($E(X) = \int_{-\infty}^{\infty} xf(x)\,\mathrm{d}x$). Teachers may also discuss with students the definition of expectation of a function of X. The following shows the two definitions. $E[g(x)] = \sum pg(x)$ discrete random variable $E[g(x)] = \int_{-\infty}^{\infty} f(x)g(x)\,\mathrm{d}x$ continuous random variable . In the case of discrete random variables, students are expected to know the meanings of $E(X)$ (= μ) and $E\left[(X-\mu)^2\right]$ (= Var(X) = σ^2). In particular, teachers should indicate that μ is a measure of central tendency while σ^2 is a measure of dispersion of X about μ. Interesting examples can be discussed.

Detailed Content	Time Ratio	Notes on Teaching

Example

The probability of a candidate passing an examination at anyone attempt is 0.4. If he fails, he carries on entering until he passes and each entry costs him \$120. Teachers may discuss with students the expected cost of his passing the examination.

Calculations involving fair games, expected gain/loss are best illustrated by real-life examples. The following are two of them.

Example 1

In an investment, a man can make a profit of \$5 000 with a probability of 0.62 or a loss of \$8 000 with a probability of 0.38.

$E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$ can be calculated from

$\mu = \$(5\,000)\,(0.62) + \$(-8000)\,(0.38) = \$60$

$\sigma^2 = (5000 - 60)^2\,(0.62) + (-8000 - 60)^2\,(0.38)$

μ is called the expected gain.

Example 2

A gambling machine has four windows and each of them displays one of the four different colours: red, orange, yellow and blue. Each of the colours is equally likely to be displayed and the colour displayed by the machine on one window is independent of the colour displayed on the other windows. A man pays \$*a* for a game. He receives \$5 if all the colours displayed in the four windows are different. He receives \$30 if all the colours displayed are the same. In all other cases, he loses. \$*X* is the net amount he received in playing a game.

In this example, teachers can discuss the following with students.

1. When $E(X) = 0$, the game is a fair game. What is the fair price (i.e. \$*a*)?
2. Suppose $a = 1$, what are $E(X)$ and Var(*X*)?

Most of the students should realize that $E(X) < 0$ in most of the gambling games. Teachers may also ask students to work out the new μ and σ^2 when all the money is doubled and to find the relations between the new and old parameters. For abler students, teachers may ask them to guess the value of $E(Y)$ where \$*Y* is the net amount the man receives if he plays the games twice.

Detailed Content	Time Ratio	Notes on Teaching

In the case of continuous random variables, examples showing the steps in calculating $E(X)$ and Var(*X*) should be provided.

Example

Orange juice is delivered to a fast food shop every morning. The daily demand for orange juice is a continuous random variable *X* distributed with a probability density function $f(x)$ of the form

$$f(x) = \begin{cases} ax(b-x) & \text{for } 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

The mean daily demand is 0.625 units.

a and *b* can be calculated from the equation

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1 \quad \text{and} \quad \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = 0.625$$

The orange juice container at this fast food shop is filled to their total capacity of 0.8 units every morning.

The probability P that in a given day, the fast food shop cannot meet the demand for orange juice is given by $P = 1 - \phi(0.8) = \int_{0.8}^{1} f(x)\,\mathrm{d}x$

For abler students, teachers may guide them to prove the two formulae

$E[ag(X) + b] = aE[g(X)] + b$ and $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ where *a*, *b* are constants.

Also, it is not difficult for an average student to show that

$$E[g(X) + h(X)] = E[g(X)] + E[h(X)]$$

and $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$

The following example shows the use of the above formulae.

Example

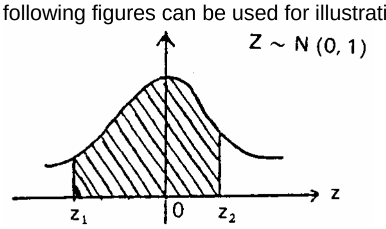
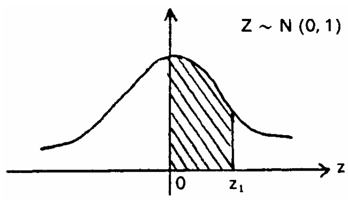
Given $Z = 2X^2 - 3X + 5$ where *X* is a random variable with mean μ and variance σ^2.

E(*Z*) can be obtained from
$$\begin{aligned} E(Z) &= E(2X^2 - 3X + 5) = E(2X^2) - E(3X) + E(5) \\ &= 2(\mu^2 + \sigma^2) - 3\mu + 5 \end{aligned}$$

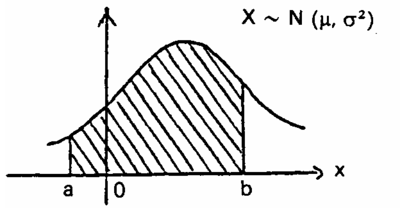
Detailed Content	Time Ratio	Notes on Teaching
20.3 Binomial Distribution (a) Bernoulli trials, Binomial probability	7	Teachers can introduce Bernoulli trials by using the familiar example of tossing a fair coin. Teachers should emphasize that in a Bernoulli trial, there are only two possible outcomes. Repeated Bernoulli trials play an important role in probability and statistics especially when the probabilities of the two possible outcomes are the same for each trial. Students should know that the probability associated with r successes in the n trials is given by the expression $P(r \text{ successes}) = C_r^n p^r q^{n-r}$ Teachers can easily quote numerous examples in our daily life to illustrate the Binomial probability. *Example* 1 A die is thrown n times. In order that the probability of getting at least one 'six' is greater than 0.99, n should satisfy the following inequality: $1-\left(\frac{5}{6}\right)^n > 0.99$ *Example* 2 *r* balls are randomly distributed in n cells. Students may be asked to find the probability P_k that a specified cell contains exactly *k* balls ($k \le r$). In this case, students are expected to know that $P_k = P(k$ successes in *r* Bernoulli trials) with $p = \frac{1}{n}$.
(b) Binomial distribution		At this stage, teachers may introduce that Binomial distribution can be considered as a repeated Bernoulli trial with the same probability of success. Teachers may also introduce the notation B (n, p) for the distribution. Students are expected to know the formulae $E(X) = np$ and $\text{Var}(X) = npq$ for Binomial distribution. The probability graph of a Binomial distribution with different values of *n* and *p* can be shown. Students should be able to see that when $p = q = 0.5$, the graph is symmetric. For abler students, the mode of Binomial distribution can also be discussed.

Detailed Content	Time Ratio	Notes on Teaching
(c) Applications		Binomial probability distribution is useful in describing many real-life events. The following are three possible applications. *Example* 1 A student sits for a test which contains only 4 multiple choice questions. With his knowledge of the subject, he has a probability of 0.7 of knowing the correct answer of each question. There are 5 options in each question, thus the student has the probability 0.2 of getting the correct answer in each question through guessing. He has attempted all the questions. The probability that the student knows the correct answers of 3 questions is $C_3^4(0.7)^3(0.3)$. Since the student can get the correct answer of a question simply by guessing, *P*(correct answer for a question) = *p* can be calculated from the two cases (a) he knows the question and (b) he guesses it correctly. Suppose *X* is the number of correct answer(s) obtained, students may be asked to calculate $E(X)$ (= $4p$), $\text{Var}(X)$ (= $4p(1-p)$) and $P(X = 1)$. ($= C_1^4 p^1(1-p)^3$) The following questions can also be raised. 1. 2 marks will be awarded for a correct answer and 1 mark will be deducted for a wrong answer. Suppose *Y* is the total score obtained by the student, calculate *E*(*Y*) and Var(*Y*). 2. Given that the student only knows the correct answers of 3 questions, what is the probability that the student obtains full marks? 3. Given that the student only gets one correct answer, what is the probability that he gets it through guessing? *Example* 2 5% of light bulbs are defective. A large batch of light bulbs is tested according to the following rules. (a) A sample of 10 light bulbs is tested.

Detailed Content	Time Ratio	Notes on Teaching
		(i) If two or more light bulbs are defective, then the whole batch is rejected. (ii) If there is no defective light bulb, the whole batch is accepted. (iii) If there is only one defective light bulb, try rule (b). (b) Another sample *of 10* bulbs is tested. If there is no defective bulb, the whole batch is accepted; otherwise it is rejected. If *X* is the number of light bulbs examined, then it is not difficult to find $P(X = 20) = 10\ (0.95)^9\ (0.05)$ and $P(X = 10) = 1 - 10\ (0.95)^9\ (0.05)$. Students may be asked to find $E(X)$ and $\text{Var}(X)$. *Example 3* 10% of the items produced by a machine are defective. The items are packed in large batches, and a batch is accepted if a sample *of* n items from it contains no defectives; otherwise it is rejected. The least value of n to ensure the probability that the batch will be rejected is at least 0.95 satisfies $(0.9)^n < 1 - 0.95$. If $n = 10$, then *P* (the batch being accepted) = $(0.9)^{10}$ = P. The chance that of 8 batches being inspected, 5 will be rejected = $C_3^8 p^3 (1-p)^5$.
20.4 **Normal Distribution** (a) Basic definitions	10	Normal distribution is a very important example *of* continuous probability distribution. The p.d.f. $f(x)$, i.e. $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$ should be introduced, but detailed explanation is unnecessary. Students are expected to recognize that $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$, but the proof is not necessary. It is worthwhile for teachers to discuss with students why normal distribution is commonly used in many subjects.

Detailed Content	Time Ratio	Notes on Teaching
		1. Easy to use. 2. Can be used as an approximation to other distributions. Graphs with different μ and σ can be introduced. Students should realize that all the graphs shown are bell-shaped and are symmetric about $x = \mu$. The notation $N(\mu, \sigma^2)$ which means a normal distribution with mean = μ and variance = σ^2 may be introduced.
(b) Standard normal curve and the use of normal table		The normal distribution depends on μ and σ. Students should find that it is difficult to tabulate the probability function of each normal distribution with a different set *of* parameters. Therefore, it is necessary to express the random variable in standard unit, using the transformation $Z = \frac{X-\mu}{\sigma}$. Students should have no difficulty in seeing that $E(Z) = 0$, $\text{Var}(Z) = 1$ and $$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{x-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = P(z_1 < Z < z_2)$$ The following figures can be used for illustration.  The two shaded parts have equal area. In the following figure, the area of the shaded part is $P(0 < Z < z_1)$.

Detailed Content	Time Ratio	Notes on Teaching



This area, for different values of z_1, is put into a table called normal distribution table (The table only gives values up to z_1 = 3.59). Adequate practice is necessary for ensuring that students can use the table properly.

Example
X is $N(8, 4)$

$$P(6 < X < 9) = P\left(\frac{6-8}{2} < \frac{X-8}{2} < \frac{9-8}{2}\right)$$
$$= P(-1 < X < 0.5)$$
$$= P(0 < Z < 1) + P(0 < Z < 0.5)$$
$$P(X > 9) = P(Z > 0.5) = 0.5 - P(0 < Z < 0.5)$$

In $P(X < k) = 0.87$, k can be obtained with greater accuracy if method of linear interpolation is used.

Teachers can remind students that in solving many of the problems, they have to make use of symmetry and laws of complementary probability. Moreover, for z_1 involving more than 3 significant figures, the method of linear interpolation should be used.

Detailed Content	Time Ratio	Notes on Teaching
(c) Applications		Standard normal distribution is essential in daily applications. Teachers should provide adequate demonstration. Examples like the following may be used.

Detailed Content	Time Ratio	Notes on Teaching

Example 1
A manufacturer uses a machine to produce resistors. He found that 10% of the resistors are less than 95Ω and 20% of the resistors are above 110Ω. The distribution of the resistances X is normal.
μ and σ can be calculated from the two equations

$$P(X < 95) = P\left(Z < \frac{95-\mu}{\sigma}\right) = 0.1$$
$$P(X > 110) = P\left(Z > \frac{110-\mu}{\sigma}\right) = 0.2$$

Example 2
Suppose X, the length of a rod, is a normally distributed random variable with mean μ and variance 1. If X does not meet certain specifications, then the manufacturer will suffer a loss. Specifically, the profit M (per rod) is the following function of X.

$$M = \begin{cases} 3 & \text{if } 8 \le X \le 10 \\ -1 & \text{if } X < 8 \\ -5 & \text{if } X > 10 \end{cases}$$

The expected profit, $E(M)$, is given by $E(M) = 8\phi(10-\mu) - 4\phi(8-\mu) - 5$ where $\phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt$ is the cumulative probability function.

Suppose that the manufacturing process can be adjusted so that different values of μ may be achieved. The value of μ corresponding to maximum profit can be determined by differentiating $E(M)$ with respect to μ.

Example 3
A factory produces soft drinks contained in bottles. The normal volume contained in a bottle is 1.25 litres. However, due to random fluctations in the automatic bottling machine, the actual volume per bottle varies according to a normal distribution. It is observed that 15% of the bottles contain less than 1.25 litres whereas 10% contain more than 1.30 litres.

Detailed Content	Time Ratio	Notes on Teaching
		Students should have no difficulty in finding the mean μ and standard deviation σ of the volume distribution. The cost in cents of producing a bottle containing *x* litres of soft drinks is $C = 36 + 62x + 5x^2$ where *x* is the random variable having the above distribution. The expected cost of a bottle = $E(C)$ where $E(C) = 36 + 62\mu + 5\left(\mu^2 + \sigma^2\right)$. The expected cost of 20 000 bottles is 20 000$E(C)$.
(d) Binomial approximated to normal distribution		Students should be made clear that the binomial probability can be calculated by using normal approximation only when *n* is large. In this case, the mean and variance can be taken as *np* and *npq* respectively. Students should also be reminded that 'end continuity corrections' is required in this approximation. *Example* A coin is tossed 400 times. If X represents the number of heads obtained, then X is $B(400, 0.5)$. When it approximates to $N(200, 100)$, $P(195 \le X \le 210) = P\left(\frac{194.5 - 200}{10} < Z < \frac{210.5 - 200}{10}\right)$ Students may be interested to know that $P(195 \le X \le 210) \ne P(195 < X < 210)$.
20.5 **Linear Combination of Independent Normal Variables**	6	Students should recognize that the sum of scalar multiples of independent normal variables is also normal. From this, it is not hard to see that: If X and Y are two independent normal variables such that $X \sim N\left(\mu_1, \sigma_1^2\right)$ and $Y \sim N\left(\mu_2, \sigma_2^2\right)$, then $aX + bY \sim N\left(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2\right)$ for any real values *a* and *b*. The above result can be extended easily to n independent normal variables. Teachers should also quote examples to illustrate the usefulness of the above fact in daily-life application.

Detailed Content	Time Ratio	Notes on Teaching
		Example 1 Cakes are sold in packets of 6. The mass of each cake is a normal variable with mean 25 g and standard deviation 2 g. The mass of the packing material is a normal variable with mean 30 g and standard deviation 4 g. Find the distribution of the total mass of each packet of cakes and hence find the probability that the total mass of a packet is less than 142 g. *Example 2* The thickness, A cm, of a paperback is normally distributed with mean 2 cm and variance 0.63 cm^2. The thickness, B cm, of a hardback is normally distributed with mean 5 cm and variance 1.42 cm^2. The distribution of $X = 2A - B$ can be determined and hence the probability that a randomly chosen paperback is less than half the thickness of a randomly chosen hardback can be evaluated by using the standard normal distribution table.
	27	