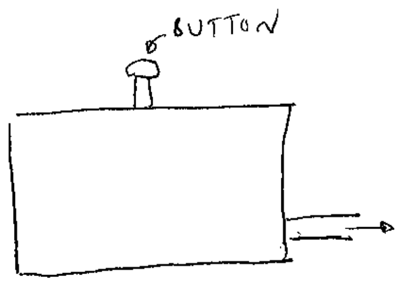
RANDOM PROCESSES

RANDOM VARIABLE: EXAMPLES OF RANDOM VARIABLES (RV) ARE

(i) THE ROLL OF A DICE
(ii) THE HEIGHT OF A PERSON
(iii) THE NOISE VOLTAGE APPEARING ACROSS THE TERMINALS OF A RESISTOR.

UNLIKE THE DETERMINISTIC VARIABLE, THE VALUE OF THE RANDOM VARIABLE IS NOT CERTAIN. LET US SUPPOSE THAT THE MARKS OBTAINED BY STUDENTS IN A COURSE IS RANDOM. LET THERE BE N STUDENTS IN THE CLASS. IMAGINE THAT MARKS ARE INSIDE A BOX WITH A BUTTON ON IT. EVERY TIME A BUTTON IS PUSHED, THE BOX OUTPUTS THE MARKS OF A STUDENT AS SHOWN BELOW:



SAMPLE x OF R.V. X

LET US DENOTE THE MARKS x_n. AFTER REPEATED BUTTON PUSHES, WE OBTAIN THE SEQUENCE $x_1, x_2 \cdots x_N$. THE OUTPUT OF THE BOX WHICH GENERATES THE ENTIRE SET OF SAMPLES x_n IS CALLED X. THE QUANTITY X IS A RANDOM VARIABLE.

THE SET OF ALL POSSIBLE OUTCOMES (i.e. MARKS x_n) IS REFERRED TO AS THE ENSEMBLE.

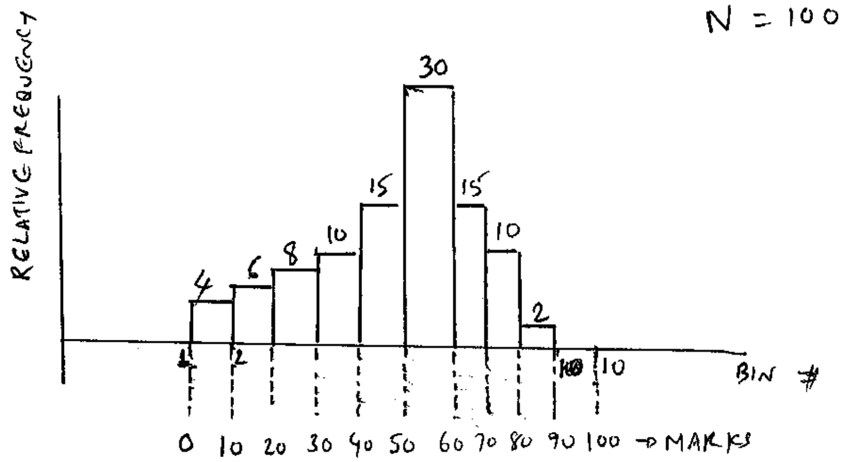
WE NOW DISCRETIZE THE MARKS OBTAINED BY STUDENTS INTO "BINS".

SUPPOSE THAT THE WIDTH OF A BIN IS 10. i.e.

BIN #1, 0 - 10
BIN #2, 10.1 - 20
⋮
BIN #10, 90.1 - 100 (A+)

A STUDENT WHO HAS GOT 83 AND ANOTHER STUDENT WHO HAS 89 CORRESPOND TO THE SAME BIN (BIN #9). WE NOW PUSH THE BUTTON ON THE BOX AND DEPENDING ON THE OUTPUT x_n OF THE BOX, WE ASSIGN IT TO THE APPROPRIATE BIN. AFTER GETTING ALL THE SAMPLES, WE PLOT THE NUMBER OF STUDENTS FALLING INTO THE SAME BIN VS THE BIN NUMBER, AS SHOWN BELOW.

N = 100



THE NUMBER OF STUDENTS OCCUPYING THE SAME BIN IS CALLED "RELATIVE FREQUENCY" AND THIS KIND OF PLOT IS CALLED THE "HISTOGRAM". THIS GRAPH INDICATES THAT THERE ARE 4 STUDENTS WHO HAVE THEIR MARKS IN THE RANGE 0 – 10 AND SO ON. OBVIOUSLY, THE SUM OF ALL RELATIVE FREQUENCIES OF BINS SHOULD BE N.

WE DENOTE $x(k)$, $k = 1, 2, \cdots K$ AS THE STUDENTS' MARKS – (DISCRETIZED) WHERE K IS THE TOTAL NUMBER OF BINS, AND $f(k)$ AS THE RELATIVE FREQUENCY (i.e. NO. OF STUDENTS IN THE BIN k).

<u>MEAN AND VARIANCE</u> :

GIVEN A SEQUENCE x_n OF THE RANDOM VARIABLE, THE MEAN m_x IS GIVEN AS

$$m_x(N) = \frac{1}{N} \sum_{n=1}^{N} x_n \longrightarrow (1)$$

HOWEVER, IN THE ABOVE DEFINITION THE MEAN DEPENDS EXPLICITLY ON ~~TH~~ 'N'. TO MAKE THE MEAN INDEPENDENT OF THE NUMBER OF SAMPLES, WE DEFINE IT AS

$$m_x = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} x_n . \longrightarrow (2)$$

ALTERNATIVELY, WE CAN CALCULATE THE MEAN USING THE HISTOGRAM. NOTE THAT THE STUDENTS OCCUPYING A BIN HAVE THE SAME MARKS. THEREFORE, TO CALCULATE THE AVERAGE, WE CAN MULTIPLY THE MARKS BY THE RELATIVE FREQUENCY AND ADD OVER ALL THE BINS, i.e.

$$m_x'(K) = \lim_{N \to \infty} \sum_{k=1}^{K} x(k)\, f(k) \cdot \frac{1}{N} \longrightarrow (3)$$

THE MEAN CALCULATED USING (2) AND (3) ARE DIFFERENT BECAUSE THE MEAN OF (3) DEPENDS ON THE TOTAL NUMBER OF BINS (OR EQUIVALENTLY THE BIN WIDTH). HOWEVER, AS THE WIDTH OF THE BIN DECREASES (i.e. $K \to \infty$), THE ERROR DUE TO DISCRETIZATION DECREASES, TOO, AND THE MEAN CALCULATED USING (3) APPROACHES THAT USING (2), i.e.

$$\lim_{K \to \infty} m_x'(K) = m_x = \lim_{\substack{N \to \infty \\ K \to \infty}} \sum_{k=1}^{K} x(k) \frac{f(k)}{N}$$

IN THE LIMIT OF THE WIDTH OF A BIN GOING TO ZERO, WE CAN REPLACE $x(k)$ BY A CONTINUOUS VARIABLE x, $\frac{f(k)}{N}$ by $f(x)$ AND THE SUM $\longrightarrow$ integral, i.e.

$$m_x = \int_{-\infty}^{\infty} x\, f(x)\, dx \longrightarrow (4)$$

THE FUNCTION $f(x) = \lim_{N\to\infty,\, K\to\infty} \frac{f_K}{N}$ IS CALLED THE PROBABILITY DENSITY FUNCTION (p.d.f.). SINCE THE SUM OF RELATIVE FREQUENCY IS EQUAL TO N, IT FOLLOWS THAT

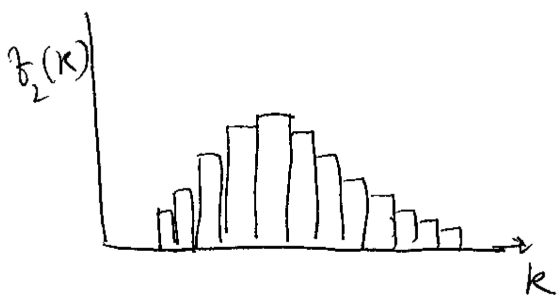
$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \qquad \longrightarrow (5)$$

THE OPERATION IN (4) IS ALSO CALLED THE EXPECTATION OF THE RANDOM VARIABLE. THE EXPECTATION OPERATOR IS DENOTED AS $E(x)$.

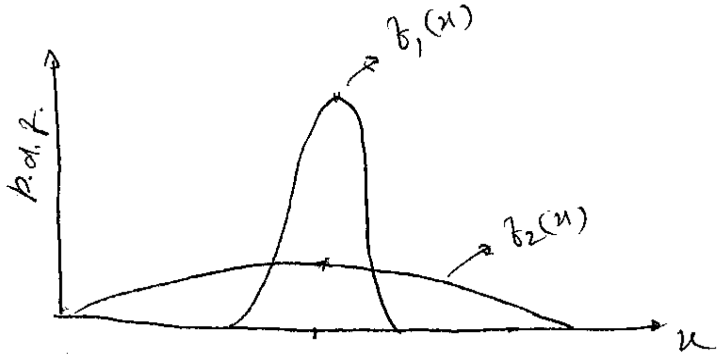
$$E(x) = m_x$$

VARIANCE:

CONSIDER TWO HISTOGRAMS SHOWN BELOW. $f_1(K) \to f_1(x)$ AND $f_2(K) \to f_2(x)$ AS $K \to \infty$.



PROBABILITY DENSITY FUNCTIONS CORRESPONDING TO $f_1(x)$ AND $f_2(x)$ ARE SHOWN BELOW:



IF THE MEAN OF A HISTOGRAM OR A PROBABILITY DENSITY FUNCTION ALONE IS SPECIFIED, WE CAN NOT DETERMINE THE SPREAD OF MARKS. FOR EXAMPLE THE PROBABILITY DENSITY FUNCTIONS SHOWN IN THE FIGURE HAVE THE SAME MEAN, BUT THE DEVIATION OF MARKS FROM THE MEAN ARE DIFFERENT, i.e. these p.d.f.s HAVE DIFFERENT WIDTHS. THE WIDTH OF THE DISTRIBUTION CAN BE QUANTIFIED IN SEVERAL WAYS. THE SIMPLEST WAY IS TO FIND THE <u>ABSOLUTE</u> DEVIATION OF THE MARKS FROM THE MEAN AND THEN AVERAGE IT ACCORDING TO THE FORMULA:

$$\text{SPREAD} : \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} |x_n - m_x| \qquad m_x = \text{MEAN}$$

HOWEVER, THIS QUANTITY IS DIFFICULT TO DEAL WITH MATHEMATICALLY SINCE IT IS NOT DIFFERENTIABLE AT $x = m_x$. INSTEAD, WE ~~ADOPT~~ DEFINE A MEASURE CALLED 'VARIANCE' DENOTED AS σ_x^2 IN THE FOLLOWING FORM:

$$\sigma_X^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} (x_n - m_X)^2$$

VARIANCE GIVES THE MEASURE OF VARIATIONS OF THE MARKS FROM THE MEAN. IT IS THE MEAN SQUARED VALUE OF THE DEVIATION OF THE MARKS FROM THE MEAN. THE STANDARD ~~VA~~ DEVIATION σ_X IS DEFINED AS THE ~~SHORT~~ SQUARE ROOT OF THE VARIANCE, WHICH IS THE ROOT MEAN SQUARE VALUE OF THE DEVIATIONS.

IF WE USE HISTOGRAM, INSTEAD OF THE ~~Eq.~~ DIRECT CALCULATION ~~OF~~ AS IN ~~Eq~~ EQ. (4), VARIANCE IS GIVEN BY

$$\sigma_X^2 = \lim_{\substack{N\to\infty \\ K\to\infty}} \sum_{k=1}^{K} (x(k) - m_X)^2 \cdot f(k)$$

$$= \int_{-\infty}^{\infty} (x - m_X)^2 \cdot f(x)\, dx \longrightarrow (6)$$

THE QUANTITY, VARIANCE IS ALSO THE ~~AVERAGE~~ EXPECTATION OF THE QUANTITY $(x - m_X)^2$ DENOTED AS $E(x - m_X)^2$.

$$E(x - m_X)^2 = E[x^2 - 2m_X x + m_X^2]$$

$$= E(x^2) - 2m_X E(x) + m_X^2 \quad \left[\text{SINCE } m_X \text{ IS CONST; } E(m_X^2) = m_X^2\right]$$

$$= E(x^2) - 2m_X m_X + m_X^2$$

$$= E(x^2) - m_X^2 \longrightarrow (7)$$