APPENDIX - 4

Fitting of curves: Least square method

A.4.1. Introduction

Let x is the independant variable and y the dep dent variable and let (x_i, y_i), $i=1,2,\ldots,n$, be a given of n pairs of values. The general problem in cu fitting is to determine, if possible, an analyti expression of the form $y=f(x)$ for the function relationship suggested by the given data.

A.4.2. Fitting of a straight line

Let $y = a + bx$ ... (1)

be the straight line of best fit to the set of n points (x_i, y_i), $i=1,2,\ldots,n$.

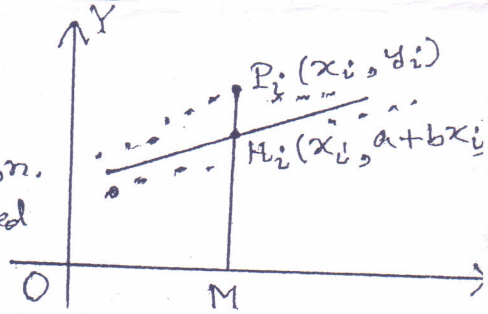


Fig. 1

The term 'best fit' is interpreted in accordance with Legender's Principle of Least Square which consists in minimizing the sum of the squares of the deviations of the actual value from its estimated values as given by the line of b fit.

Let $P_i(x_i, y_i)$ be any general point in the scatter d ram (see Fig. 1). Draw $P_iM \perp$ to X-axis meeting the li (1) in $H_i(x_i, a+bx_i)$.

$\therefore P_iH_i = P_iM - H_iM = y_i - (a+bx_i)$,

which is called the error of estimation.

According to the principle of least squares, we have to determine a and b so that

$$E = \sum_{i=1}^{n} P_iH_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \quad \text{is minimum.}$$

For minimum:

$$\left.\begin{array}{l}\frac{\partial E}{\partial a}=0 \Rightarrow -2\sum_i (y_i-a-bx_i)=0 \Rightarrow \sum_i y_i = na+b\sum x_i \\ \frac{\partial E}{\partial b}=0 \Rightarrow -2\sum x_i(y_i-a-bx_i)=0 \Rightarrow \sum_i x_i y_i = a\sum_i x_i + b\sum_i x_i^2\end{array}\right\} \cdots (2)$$

Equations (2) are known as <u>normal equations</u> for estimating a and b.

All the quantities $\sum x_i, \sum y_i, \sum x_i y_i, \sum x_i^2$ can be obtained from the given set of points (x_i, y_i) and so the equations (2) can be solved for a and b. With these values of a and b so obtained equation (1) is the line of best fit to the given set of points (x_i, y_i), $i=1,2,\dots,n$.

<u>Illustration 1</u>. Fit a straight line to the following data

x_i :	1	2	3	4	6	8
y_i :	2.4	3	3.6	4	5	6

<u>Solution</u>. Let the required straight line be $y=a+bx$. Then the normal equations are

$\sum y_i = 6a + b\sum x_i$ and $\sum x_i y_i = a\sum x_i + b\sum x_i^2$

x_i	y_i	$x_i y_i$	x_i^2
1	2.4	2.4	1
2	3	6	4
3	3.6	10.8	9
4	4	16	16
6	5	30	36
8	6	48	64
$\sum x_i = 24$	$\sum y_i = 24$	$\sum x_i y_i = 113.2$	$\sum x_i^2 = 130$

$\therefore$ $24 = 6a + 24b$,
$113.2 = 24a + 130b$.

Solving: $a = 1.976$, $b = 0.506$

$\therefore$ Straight line of best fit is $y = 1.976 + 0.506$

A.4.3 Fitting of second degree (parrabolic) curve

Let $y = a + bx + cx^2$ be the second degree parrabola of b fit to the set of n points (x_i, y_i), $i = 1, 2, \ldots, n$. Using pr ciple of least square, we have to determine a, b, c that $E = \sum_i (y_i - a - bx_i - cx_i^2)^2$ is minimum.

For minimum:

$$\left.\begin{array}{l} \frac{\partial E}{\partial a} = 0 \Rightarrow \sum y_i = na + b\sum x_i + c\sum x_i^2 \\ \frac{\partial E}{\partial b} = 0 \Rightarrow \sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3 \\ \frac{\partial E}{\partial c} = 0 \Rightarrow \sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 \end{array}\right\} \text{Normal equations}$$

Illustration 2. Fit a second degree parrabola to the follo data:

x_i :	0	1	2	3	4
y_i :	1	1.8	1.3	2.5	6.3

Solution. Let the required parabola of best fit is $y = a + bx$

	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
	0	1	0	0	0	0	0
	1	1.8	1	1	1	1.8	1.8
	2	1.3	4	8	16	2.6	5.2
	3	2.5	9	27	81	7.5	22.5
	4	6.3	16	64	256	25.2	100.8
$\sum$	10	12.9	30	100	354	37.1	130.3

Normal equations: $12.9 = 5a + 10b + 30c$

$37.1 = 10a + 30b + 100c$

$130.3 = 30a + 100b + 354c$

Solving: $a = 1.42$, $b = -1.07$, $c = 0.55$.

$\therefore$ Parabola of best fit is $y = 1.42 - 1.07x + 0.55x^2$.

Appendix-2

Two-Dimensional Distributions

A2.1 Two-Dimensional Random Variables

So far we have defined only one random variable on a sample space. It is also possible to define two or more random variables on the same sample space. For example, if we are interested in recording the height and weight of every person in a certain locality, then study of two random variables is needed to describe such experiments mathematically.

Definition 1. Let X and Y be two random variables defined on the same sample space S. Then the function (X, Y) that assigns a point in $R^2 (= R \times R)$ is said to be a two-dimensional random variable.

Definition 2. The distribution function of the two-dimensional random variable (X, Y) is a real valued function F defined by the relation:

$$F(x, y) = P(X \leq x, Y \leq y), \quad \forall\, x, y \in R.$$

It is also called the joint distribution of the random variables X and Y.

Definition 3. A two-dimensional random variable (X, Y) is said to be discrete if it takes at most a countable number of points in R^2. Therefore, if (X, Y) is a two-dimensional discrete random variable, then the joint probability mass function of (X,Y) is denoted and defined as

$$f_{X,Y}(x_i, y_i) = P(X = x_i, Y = y_i).$$

such that $\sum f_{X,Y}(x_i, y_i) = 1$, where the summation is taken over all possible values of (X,Y).

<u>Definition 4</u> A two-dimensional random variable (X,Y) is said to be <u>continuous</u> if there exists a non-negative function $f_{X,Y}$ (called the probability density function) which satisfies the relation:

$$F(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\,dt\,ds, \ \forall (x,y) \in R^2.$$

<u>Properties</u>

(i) $P(a < X \le b, c < Y \le d) = \int_{c}^{d} \left[\int_{a}^{b} f(x,y)\,dx \right] dy = \int_{a}^{b} \left[\int_{c}^{d} f(x,y)\,dy \right] dx$

(ii) $F(x,y) = \int_{-\infty}^{y} \left[\int_{-\infty}^{x} f(t,s)\,dt \right] ds = \int_{-\infty}^{x} \left[\int_{-\infty}^{y} f(t,s)\,ds \right] dt, \ \forall (x,y) \in R^2$

(iii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\,dx \right] dy = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\,dy \right] dx = 1.$

Definition 5 (Marginal Distribution)

Given a distribution function F of (X, Y), the one-dimensional distribution functions (called marginal distributions) F_X and F_Y can be obtained as follows:

$$\{X \le x, Y < \infty\} = \{X \le x\} \cap \{Y < \infty\} = \{X \le x\} \cap S = \{X \le x\}$$

$$\Rightarrow P(X \le x, Y < \infty) = P(X \le x) \Rightarrow F_X(x) = P(X \le x) = \lim_{y \to \infty} F(x, y).$$

Similarly, $F_Y(y) = P(Y \le y) = \lim_{x \to \infty} F(x, y)$.

For discrete distribution:

$$\{X = x_i\} = \{X = x_i, Y = y_1\} \cup \{X = x_i, Y = y_2\} \cup \dots = \bigcup_j \{X = x_i, Y = y_j\}$$

Since the events on the right are mutually exclusive, we have

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j) \Rightarrow p_{i.} = \sum_j p_{ij}$$

Similarly, $P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) \Rightarrow p_{.j} = \sum_i p_{ij}$

Illustrative Examples - I

Ex. 1 Determine the value of the constant k such that the function $f(x,y)$ given by

$$f(x,y) = \begin{cases} k \dfrac{1+x+y}{(1+x)^4(1+y)^4}, & 0 \le x, y < \infty. \\ 0, & \text{otherwise} \end{cases}$$

is a probability density function of bivariate distribution (X, Y). Also find the marginal distributions of X and Y.

(BESUS - 2012).

Solution

We have, $k \displaystyle\int_0^\infty \int_0^\infty \frac{1+x+y}{(1+x)^4(1+y)^4}\, dx\, dy = 1$

$$\Rightarrow k \int_0^\infty \frac{1}{(1+y)^4}\left[\int_0^\infty \frac{1+x+y}{(1+x)^4}\, dx\right] dy = 1$$

$$\Rightarrow k \int_0^\infty \frac{1}{(1+y)^4}\left[\int_0^\infty \frac{dx}{(1+x)^3} + y \int_0^\infty \frac{dx}{(1+x)^4}\right] dy = 1$$

$$\Rightarrow k \int_0^\infty \frac{1}{(1+y)^4}\left[-\frac{1}{2(1+x)^2} - \frac{y}{3(1+x)^3}\right]_{x=0}^{\infty} dy = 1$$

$$\Rightarrow k \int_0^\infty \frac{1}{(1+y)^4}\left(\frac{1}{2} + \frac{1}{3}y\right) dy = 1 \Rightarrow \frac{k}{6}\int_0^\infty \frac{3+2y}{(1+y)^4}\, dy = 1$$

$$\Rightarrow \frac{k}{6}\int_0^\infty \frac{2(1+y)+1}{(1+y)^4}\, dy = 1 \Rightarrow \frac{k}{6}\left[-\frac{1}{(1+y)^2} - \frac{1}{3(1+y)^3}\right]_0^\infty = 1.$$

$$\Rightarrow \frac{K}{6}\left(1+\frac{1}{3}\right)=1 \Rightarrow K=\frac{9}{2}$$

$$f_X(x) = \frac{9}{2}\int_0^\infty \frac{1+x+y}{(1+x)^4(1+y)^4}\,dy\,, \quad 0 \le x < \infty$$

$$= \frac{9}{2(1+x)^4}\left[-\frac{1}{2(1+y)^2} - \frac{x}{3(1+y)^3}\right]_{y=0}^{\infty} = \frac{9}{2(1+x)^4}\left(\frac{1}{2}+\frac{x}{3}\right)$$

$$= \frac{3(3+2x)}{4(1+x)^4}\,; \quad 0 \le x < \infty.$$

Similarly, $f_Y(y) = \dfrac{3(3+2y)}{4(1+y)^4}\,; \quad 0 \le y < \infty.$

Ex. 2 The probability density function of a two-dimensional random variable (X, Y) is given by

$$f(x,y) = \begin{cases} K(x+y), & 0 < x+y < 1, x>0 \\ 0, & \text{elsewhere} \end{cases}$$

Find K and evaluate $P(X < \frac{1}{2}, Y > \frac{1}{4})$.

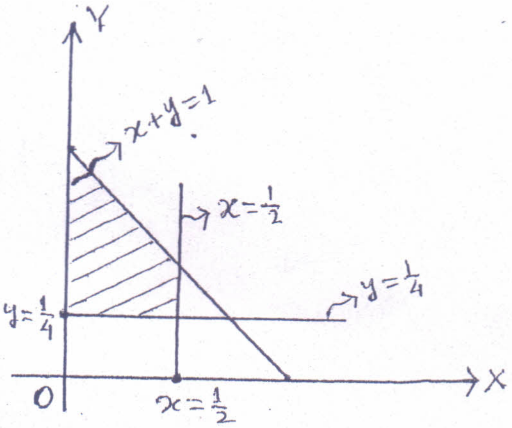
Solution We have, $K\displaystyle\int_0^1 \left[\int_{y=0}^{1-x} (x+y)\,dy\right] dx = 1$;

or, $K\displaystyle\int_0^1 \left[xy + \frac{y^2}{2}\right]_{y=0}^{1-x} dx = 1$

or, $K\displaystyle\int_0^1 \left\{x(1-x) + \frac{1}{2}(1-x)^2\right\} dx = 1$

or, $K\displaystyle\int_0^1 \left(-\frac{x^2}{2}+\frac{1}{2}\right) dx = 1$, or, $K\left(-\frac{1}{6}+\frac{1}{2}\right)=1 \Rightarrow K=3.$

$$P\left(X<\tfrac{1}{2}, Y>\tfrac{1}{4}\right) = P\left(X<\tfrac{1}{2}, \tfrac{1}{4}<Y<1-x\right)$$



$$= 3\int_{x=0}^{1/2}\int_{y=\frac{1}{4}}^{1-x} (x+y)\,dx\,dy$$

$$= 3\int_0^{1/2} \left[xy + \frac{y^2}{2}\right]_{y=\frac{1}{4}}^{1-x} dx$$

$$= 3\int_0^{1/2} \left[x(1-x) + \frac{1}{2}(1-x)^2 - \frac{x}{4} - \frac{1}{32}\right] dx = 3\int_0^{1/2} \left[\frac{15}{32} - \frac{x^2}{2} - \frac{x}{4}\right] dx$$

$$= 3\left[\frac{15}{64} - \frac{1}{48} - \frac{1}{32}\right] = \frac{105}{192} = \frac{35}{64}$$

$\cancel{3\left(\frac{15}{64}\right)}$ $\cancel{\frac{3}{16}}$

<u>Ex. 3</u>. Two random variables X and Y are jointly distributed as follows:

$$f(x,y) = \frac{2}{\pi}(1-x^2-y^2), \text{ for } 0 < x^2+y^2 < 1.$$

Find the marginal distribution of X.

<u>Solution</u> ~~We have~~

Now, $x^2+y^2=1 \Rightarrow y = \pm\sqrt{1-x^2}$ and $1-x^2 > 0$, i.e., $x^2<1$, i.e, $|x|<1$.

$$\therefore f_X(x) = \frac{2}{\pi}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} (1-x^2-y^2)\,dy = \frac{2}{\pi}\left[(1-x^2)y - \frac{y^3}{3}\right]_{y=-\sqrt{1-x^2}}^{\sqrt{1-x^2}}$$

$$= \frac{2}{\pi}\left[2(1-x^2)^{3/2} - \frac{2}{3}(1-x^2)^{3/2}\right] = \frac{2}{\pi}\cdot\frac{4}{3}(1-x^2)^{3/2}$$

$$= \frac{8}{3\pi}(1-x^2)^{3/2}; \quad |x|<1 \Leftrightarrow -1<x<1.$$