§ 3.6 Marginal Distributions

Recall the example in the last section where I drew 2 tablets at random from a bottle containing 3 aspirins, 2 sedative and 4 laxative caplets.

We let

X = # of aspirins chosen

Y = # of sedatives chosen.

Got the following joint distribution.

$y \backslash x$	0	1	2	
0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{7}{12}$
1	$\frac{2}{9}$	$\frac{1}{6}$		$\frac{7}{18}$
2	$\frac{1}{36}$			$\frac{1}{36}$
	$\frac{5}{12}$	$\frac{1}{2}$	$\frac{1}{12}$	

Add the numbers in each column and each row.

The totals for each column are the probs that X will take on the values 0, 1, 2, i.e., we get the prob. distr. of X with

$$P(X=x) = g(x) = \sum_{y=0}^{2} f(x,y) \quad , \quad x = 0,1,2.$$

Similarly the row totals are the probs. that Y takes on the values 0, 1, 2 i.e. we get the prob distr of Y with

$$P(Y=y) = h(y) = \sum_{x=0}^{2} f(x,y) \quad , \quad y = 0,1,2.$$

We have the following def^n.

<u>Def^n</u>. If X, Y are discrete r.v.'s and $f(x,y)$ is the value of their joint PDF at (x,y), the f^n

$$g(x) = \sum_{y} f(x,y)$$

for each x in the range of X is called the <u>marginal distribution of X</u>.

The fn.

$$h(y) = \sum_x f(x,y)$$

for each y in the range of Y is called the <u>marginal distribution of Y</u>.

For cts. r.v.'s the sums are replaced by integrals and we have

<u>**Defn**</u> If X, Y are cts r.v.'s and $f(x,y)$ is the value of a joint PDF at (x,y), the fn

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \quad -\infty < x < \infty$$

is called the <u>marginal density of X</u>.

The fn

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx, \quad -\infty < y < \infty$$

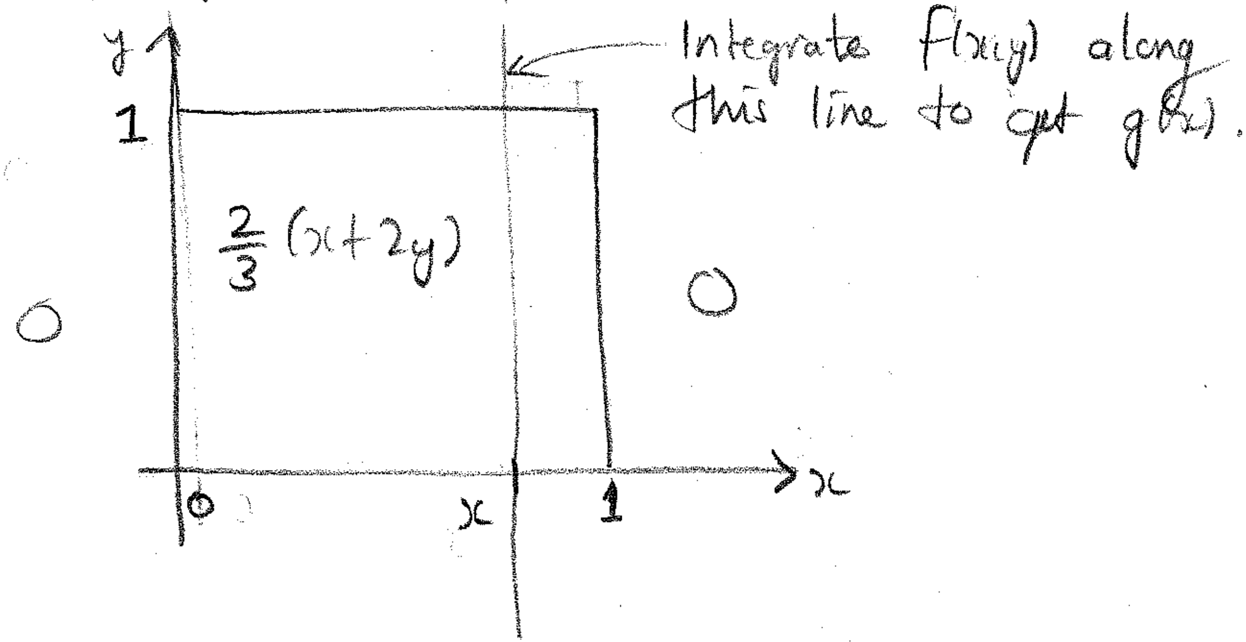
is called the <u>marginal density of Y</u>.

Ex. Supp X, Y have joint PDF

$$f(x,y) = \begin{cases} \frac{2}{3}(x+2y), & 0 < x, y < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Find the marginal densities of X and Y.

Soln Density looks like



If $0 < x < 1$,

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 \frac{2}{3}(x+2y)\,dy = \frac{2}{3}(x+1).$$

while $g(x) = 0$ elsewhere

Similarly, if $0 < y < 1$

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^1 \frac{2}{3}(x+2y)\,dx = \frac{1}{3}(1+4y)$$

while $h(y) = 0$ elsewhere.

Note that $g(x) \geq 0$ everywhere and

$$\int_{-\infty}^{\infty} g(x)\,dx = \int_0^1 \frac{2}{3}(x+1)\,dx = \left[\frac{x^2}{3} + \frac{2x}{3}\right]_0^1 = 1$$

Similarly, $h(y) \geq 0$ everywhere and

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

Thus g, h can both serve as the pdf's of cts r.v.'s by Thm 3.5 (P. 86).

When we have more than 2 r.v.'s we can talk not only of the marginal distr. of each one, but also of the <u>joint marginal distributions</u> of several of them at once.

e.g. If the discrete r.v.'s $X_1, \ldots, X_n$ have joint PDF $f(x_1, \ldots, x_n)$, then the marginal distribution of X_1 alone is given by

$$g(x_1) = \sum_{x_2} \cdots \sum_{x_n} f(x_1, x_2, \ldots, x_n)$$

('sum out' all the other variables).

for all values in the range of X_1, while the joint marginal distribution of X_1, X_2, X_3 is given by

$$m(x_1, x_2, x_3) = \sum_{x_4} \sum_{x_5} \cdots \sum_{x_n} f(x_1, \ldots, x_n)$$

(again sum out all the other variables).

For cts r.v.s the situation is similar with integrals replacing sums (as usual).

e.g. If $X_1, \ldots, X_n$ are cts with joint PDF $f(x_1, \ldots, x_n)$, the marginal density of X_2 alone is given by

$$h(x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_3 \cdots dx_n,$$

$$-\infty < x_2 < \infty$$

('integrate out' $x_1, x_3, \ldots, x_n$)

while the joint marginal density of X_1 and X_n is given by

$$\varphi(x_1, x_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, dx_2\, dx_3 \cdots dx_{n-1},$$

$$-\infty < x_1, x_n < \infty$$

(integrate out $x_2, \ldots, x_{n-1}$)

Ex. For the trivariate prob. density

$$f(x_1, x_2, x_3) = \begin{cases} (x_1+x_2)e^{-x_3}, & 0<x_1, x_2<1,\ 0<x_3 \\ 0 & \text{elsewhere} \end{cases}$$

(see earlier in § 3.4), find the joint marginal density of X_1 and X_3 and the marginal density of X_1 alone.

Soln. For $0<x_1<1, \quad 0<x_3$

$$m_1(x_1, x_3) = \int_0^1 (x_1+x_2)e^{-x_3}\, dx_2 = (x_1+\tfrac{1}{2})e^{-x_3}$$

while $m_1(x_1, x_3) = 0$ elsewhere. For $0<x_1<1$

$$g(x_1) = \int_0^\infty \int_0^1 f(x_1, x_2, x_3)\, dx_2\, dx_3 = \int_0^\infty m(x_1, x_3)\, dx_3$$

$$= \int_0^\infty (x_1+\tfrac{1}{2})e^{-x_3}\, dx_3 = x_1 + \tfrac{1}{2}.$$

while $g(x_1) = 0$ elsewhere.

<u>Note</u>. the short cut

$$g(x_1) = \int_0^\infty m(x_1, x_3)\, dx_3$$

which saved us having to (re)do one integration.