Chapter3 part1

Chapter 3 Joint Distributions

§3.1 Introduction

joint prob. distribution of two or more r.v's. defined on the same sample space.

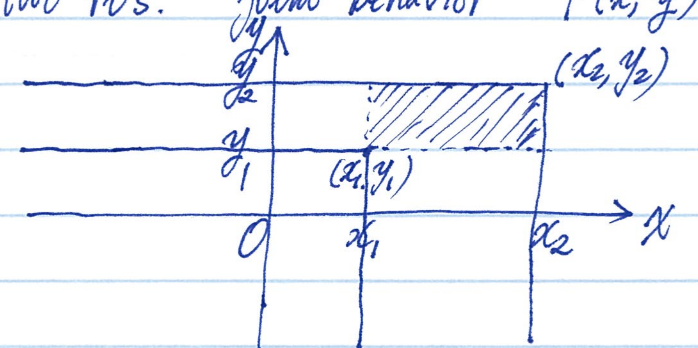
1°. In ecological studies, counts of several species, modeled as r.v's, are often made. One species is often the prey of another; clearly, the number of predators will be related to the number of prey.

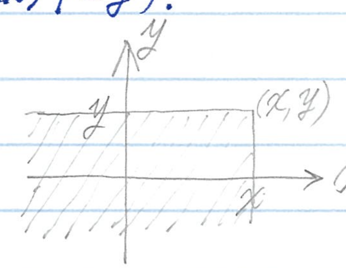
2°. The joint prob. dist of the x-, y- and z- components of wind velocity can be experimentally measured in studies of atmospheric turbulence.

3°. The joint dist. of values of various physiological variables in population of patients is often of interest in medical studies.

4°. A model for joint dist. of age and length in a population of fish can be used to estimate the age distribution from the length distribution. The age dist. is relevant to the setting of reasonable harvesting policies.

X, Y are two r.v's. joint behavior $F(x, y) = P(X \le x, Y \le y)$.





$$P(x_1 < X \le x_2,\ y_1 < Y \le y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1).$$

$$\left(\text{compare: } P(a \le X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a)\right)$$

$X_1, X_2, \cdots, X_n$; joint prob. dist.

$$F(x_1, x_2, \cdots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n).$$

§3.2 Discrete r.v.s

X, Y are discrete r.v's on the same sample space.

X takes on values $x_1, x_2, \cdots$; Y takes on values $y_1, y_2, \cdots$

joint frequency function, joint prob. mass. function: $p(x, y)$ is

$$p(x_i, y_j) = P(X = x_i, Y = y_j), \quad i, j = 1, 2, \cdots$$

Expl. A fair coin is tossed three times; Let X denote the number of heads on the first toss and Y the total number of heads. Find $p(x, y)$.

Ans: X takes 0 or 1. Y takes values of 0, 1, 2, 3.

one needs to find all prob. $p(x_i, y_j)$, $i = 1, 2$, $j = 1, 2, 3, 4$.

$X \backslash Y$	0	1	2	3
0	1/8	1/4	1/8	0
1	0	1/8	1/4	1/8

$p(0,0) = P(X=0, Y=0) = P(\text{1st } T, 3T\text{'s}) = P(3T\text{'s}) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}$

$p(0,1) = P(X=0, Y=1) = P(X=0)P(Y=1|X=0)$

$= \frac{1}{2}\cdot P(1\text{ H's in next 2 tosses}) = \frac{1}{2}\cdot\binom{2}{1}\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$

$p(0,2) = P(X=0)P(Y=2|X=0) = \frac{1}{2}\cdot\binom{2}{2}\frac{1}{2}\frac{1}{2} = 1/8$.

$p(0,3) = P(X=0)P(Y=3|X=0) = \frac{1}{2}\cdot 0 = 0$.

$p(1,0) = P(X=1, Y=0) = 0$.

$p(1,1) = P(X=1, Y=1) = P(X=1)P(Y=1|X=1) = \frac{1}{2}\cdot P(2\text{ T's in next 2 tosses}) = 1/8$.

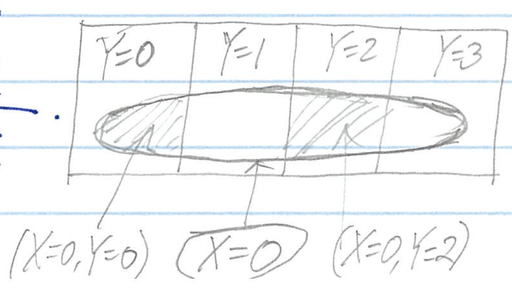
From joint p.m.f of $p(x,y)$, one could find p.m.f for X and Y individually.

for r.v. X, possible values are 0, 1.

$$P(X=0) = P(X=0, Y=0) + P(X=0, Y=1) + P(X=0, Y=2) + P(X=0, Y=3)$$
$$= \frac{1}{8} + \frac{1}{4} + \frac{1}{8} + 0 = \frac{4}{8} = 1/2.$$

similarly, $P(X=1) = \frac{1}{2}$.

X	0	1
$p(x)$	$\frac{1}{2}$	$\frac{1}{2}$



The law of total probability!

Also, $P(Y=0) = P(X=0, Y=0) + P(X=1, Y=0) = 1/8 + 0 = 1/8$.

$P(Y=1) = P(X=0, Y=1) + P(X=1, Y=1) = 1/4 + 1/8 = 3/8$.

$P(Y=2) = 3/8$, $P(Y=3) = 1/8$.

Y	0	1	2	3
$p(y)$	1/8	3/8	3/8	1/8

$$P(Y=k) = \binom{3}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{3-k}, \quad k=0,1,2,3.$$

put together

$X \backslash Y$	0	1	2	3	$P(X=x)$
0	1/8	1/4	1/8	0	1/2
1	0	1/8	1/4	1/8	1/2
$P(Y=y)$	1/8	3/8	3/8	1/8	1

marginal frequency function

generally. (X, Y). $p(x_i, y_j) = P(X=x_i, Y=y_j) = p_{ij}$, $i=1,2,\cdots,n$. $j=1,2,\cdots,m$.

$$p_X(x_i) = P(X=x_i) = \sum_{j=1}^{m} P(X=x_i, Y=y_j) = \sum_{j=1}^{m} p_{ij} = p_{i\cdot}, \quad i=1,2,\cdots,n$$

$$p_Y(y_j) = P(Y=y_j) = \sum_{i=1}^{n} P(X=x_i, Y=y_j) = \sum_{i=1}^{n} p_{ij} = p_{\cdot j}, \quad j=1,2,\cdots,m.$$

Expl. Suppose that 3 balls are randomly selected from an urn containing 3 red, 4 white and 5 blue balls. Let X and Y denote, respectively, the number of red and white balls chosen, find the joint prob. mass function $p(x,y)$.

Ans: 3 R, 4 W, 5 B $\xrightarrow{\circ\circ\circ}$ 3 balls.

$X = \#$ of R. X takes on 0, 1, 2, 3.

$Y = \#$ of W. Y ... 0, 1, 2, 3.

$X \backslash Y$	0	1	2	3	$P(X=x)$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
$P(Y=y)$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

$$p(0,0) = P(X=0, Y=0) = \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}.$$

$$p(1,0) = P(X=1, Y=0) = \frac{\binom{3}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{30}{220}.$$

$$p(2,2) = P(X=2, Y=2) = 0$$

directly, $P(X=0) = \dfrac{\binom{9}{3}}{\binom{12}{3}} = \dfrac{84}{220}$. $P(X=1) = \dfrac{\binom{3}{1}\binom{9}{2}}{\binom{12}{3}} = \dfrac{108}{220}$.

$P(X=2) = \dfrac{\binom{3}{2}\binom{9}{1}}{\binom{12}{3}} = \dfrac{27}{220}$, $P(X=3) = \dfrac{\binom{3}{3}}{\binom{12}{3}} = \dfrac{1}{220}$.

Expl A. Multinomial distribution, a generalization of binomial dist.

Suppose n independent trials can result in one of r types of outcomes and that that each of on each trial the probabilities of the r outcomes are $p_1, p_2, \cdots p_r$. Let N_i be the total number of outcomes of type i in the n trials, $i=1,2,\cdots,r$. find the joint frequency function for $(N_1, N_2, \cdots, N_r)$.

e.g. 1. toss a die, one of 6 possible outcomes.

select a voter, democrate, republican, independent.

(or race: black, hispanic, asian, white.)

$P(N_1=n_1, N_2=n_2, \cdots, N_r=n_r) = ?$

for any particular sequence of trials giving rise to $N_1=n_1, N_2=n_2, \cdots N_r=n_r$ occurs with prob. $p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$. how many such sequence? $\binom{n}{n_1\ n_2 \cdots n_r} = \dfrac{n!}{n_1!\, n_2! \cdots n_r!}$

thus $P(N_1=n_1, N_2=n_2, \cdots, N_r=n_r) = \binom{n}{n_1!\ n_2! \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$.

What is the marginal dist. of any N_i? $P(N_i=n_i) = ?$

N_i can be considered as binomial r.v. ["Success" = "type i", "others" = "failure"]

$$P(N_i=n_i) = p_{N_i}(n_i) = \binom{n}{n_i} p_i^{n_i} \cdot (1-p_i)^{n-n_i}.$$

provide an example: $n=6$, 4 S', 2 F's $S_1S_2S_3S_4F_5F_6$ $p^4(1-p)^2$

$S_1S_2S_3F_4S_5F_6$

$S_1F_2F_3S_4S_5S_6$ $\binom{6}{4}$

$\cdots$

$n=6$, 3 D' $D_1D_2D_3R_4R_5I$

2 R' $R_1R_2D_3D_4D_5I$ $\binom{6}{3\ 2\ 1} = \dfrac{6!}{3!2!1!}$

1 I $\cdots$

$p_D^3 p_R^2 p_I$

§3.3 Continuous Random Variables

$(X, Y) \sim f(x,y)$, $F(x,y) = P(X \le x, Y \le y)$.

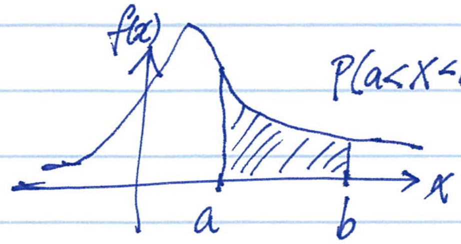
joint density function, j.p.d.f. — joint c.d.f.

1°. $f(x,y) \geq 0$

2°. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$.

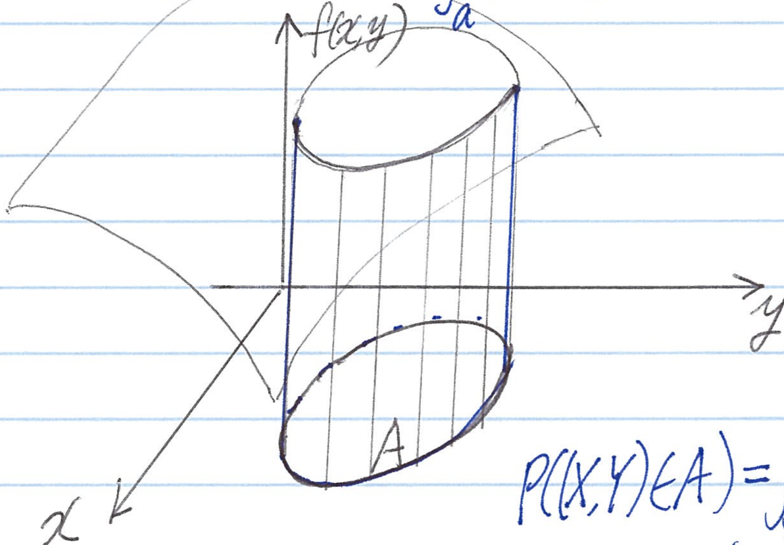
$P((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy$.

two-dimensional set



Compare:

$X \sim f(x)$, $P(a<X<b) = \int_a^b f(x)\,dx$,



$P((X,Y) \in A) = \iint_{(x,y)\in A} f(x,y)\,dx\,dy$

In particular, if $A = \{(x,y): X \le x, Y \le y\}$

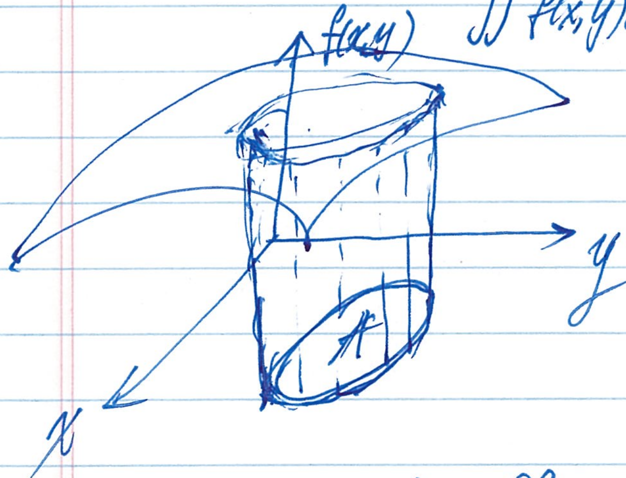
then $F(x,y) = P(X \le x, Y \le y) = P((X,Y) \in A) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dv\,du$

From fundamental theorem of multivariable calculus,

$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$. (compare: $f(x) = F'(x) = \frac{d}{dx}F(x)$).

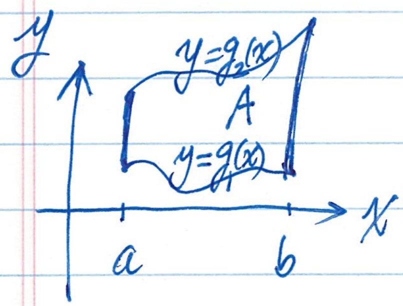
$(X,Y) \sim f(x,y)$, $f(x,y) \geq 0$

$\iint f(x,y)\,dx\,dy = 1$, total volume of the solid under surface is 1. (or graph of $f(x,y)$)

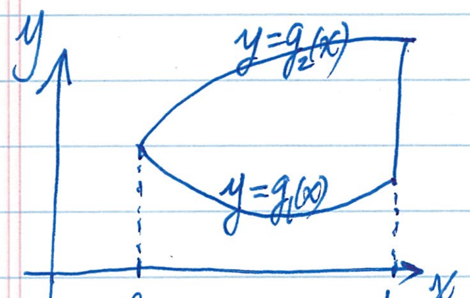


$$P\{(X,Y) \in A\} = \iint_A f(x,y)\,dx\,dy$$

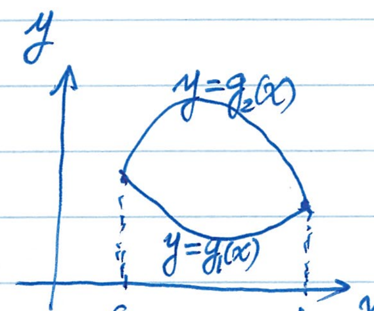
Type I



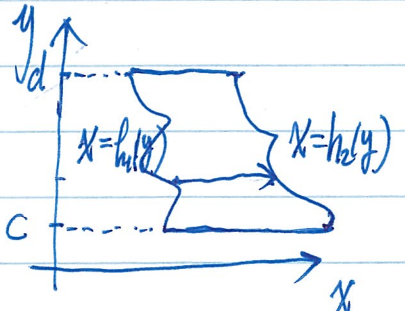
$$= \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx$$



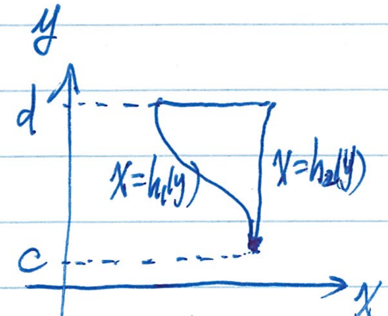
$$= \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx$$



Type II

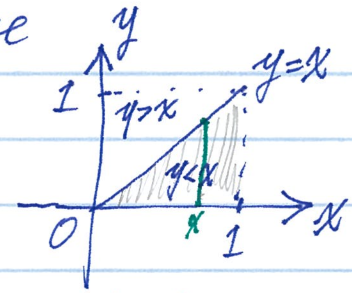


$$\int_c^d \int_{h_1(y)}^{h_2(y)} f(x,y)\,dx\,dy$$



Expl A. $(X,Y) \sim f(x,y) = \begin{cases} \frac{12}{7}(x^2+xy), & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & , \text{otherwise} \end{cases}$

find $P(X>Y)$.



Ans. $P(X>Y) = P((X,Y) \in A)$,

where $A = \{(x,y) \mid 0 \le y \le x \le 1\}$

inner integral

$$= \iint_A f(x,y)\,dx\,dy = \int_0^1 \int_0^x \frac{12}{7}(x^2+xy)\,dy\,dx$$

double integral over region A   iterated integral   first integrated with respect to y, holding x as fixed.

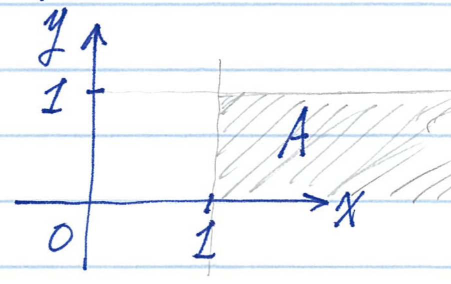
$$= \frac{12}{7}\int_0^1 \left[\int_0^x (x^2+xy)\,dy\right] dx$$

$$= \frac{12}{7}\int_0^1 \left[x^3 + x\cdot\frac{y^2}{2}\Big|_0^x\right] dx = \frac{12}{7}\int_0^1 \left[x^3+\frac{x^3}{2}\right] dx = \frac{18}{7}\int_0^1 x^3\,dx$$

$$= \frac{18}{7}\cdot\frac{1}{4}x^4\Big|_0^1 = \frac{18}{7}\cdot\frac{1}{4} = \frac{9}{14}.$$

Expl. $(X,Y) \sim f(x,y) = \begin{cases} 2e^{-x}\cdot e^{-2y}, & 0<x<\infty, 0<y<\infty \\ 0, & \text{else.} \end{cases}$

Find 1°. $P(X>1, Y<1)$. 2°. $P(X<Y)$. 3°. $P(X<a)$.



Ans. 1°. $P(X>1, Y<1) = P((X,Y) \in A)$,

where $A = \{(x,y) \mid x>1, y<1\}$

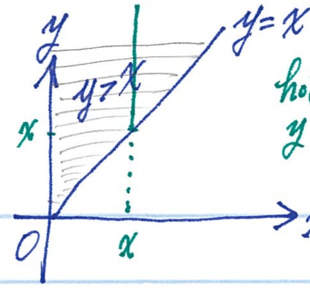
$$= \iint_A f(x,y)\,dx\,dy = \int_1^\infty \int_0^1 2e^{-x}e^{-2y}\,dy\,dx$$

$$= \int_1^\infty e^{-x}\int_0^1 2e^{-2y}\,dy\,dx = \int_1^\infty e^{-x}\cdot[-e^{-2y}]\Big|_0^1\,dx$$

$$= \int_1^\infty e^{-x}\cdot[(-e^{-2})-(-e^{0})]\,dx = \int_1^\infty e^{-x}(1-e^{-2})\,dx = (1-e^{-2})\int_1^\infty e^{-x}\,dx$$

$$= e^{-1}\cdot(1-e^{-2}).$$

2°. $P(X<Y)=\iint_A f(x,y)\,dx\,dy$, $\quad A=\{(x,y)\mid 0<x<y<\infty\}$



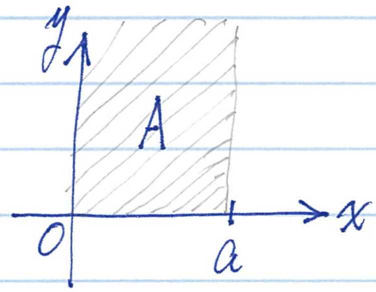
hold x as fixed, y from x to ∞!

$$=\int_0^\infty\int_x^\infty 2e^{-x}e^{-2y}\,dy\,dx$$

$$=\int_0^\infty e^{-x}\int_x^\infty 2e^{-2y}\,dy\,dx=\int_0^\infty e^{-x}\cdot\left(-e^{-2y}\right)\Big|_x^\infty dx=\int_0^\infty e^{-x}\cdot\left(0-(-e^{-2x})\right)dx$$

$$=\int_0^\infty e^{-x}\cdot e^{-2x}\,dx=\int_0^\infty e^{-3x}\,dx=\left(-\frac{1}{3}e^{-3x}\right)\Big|_0^\infty=0-\left(-\frac{1}{3}e^{-0}\right)=\frac{1}{3}$$

3°. $P(X<a)=\iint_A f(x,y)\,dx\,dy$



$$=\int_0^a\int_0^\infty 2e^{-x}e^{-2y}\,dy\,dx=\int_0^a e^{-x}\,dx=1-e^{-a}.$$

Marginal c.d.f, p.d.f for X and Y individually.

joint $(X,Y)\sim f(x,y)$, $\quad F(x,y)=P(X\le x,\ Y\le y)$.

marginal. $X\sim F_X(x)$, $f_X(x)$.

$$F_X(x)=P(X\le x)=P(X\le x,\ Y\le\infty)=\int_{-\infty}^{x}\underbrace{\int_{-\infty}^{\infty}f(u,v)\,dv}\,du$$

$$f_X(x)=F_X'(x)=\int_{-\infty}^{\infty}f(x,v)\,dv=\int_{-\infty}^{\infty}f(x,y)\,dy.$$

apply $\left[\int_{-\infty}^{x}g(u)\,du\right]_x'=g(x)$.

let $g(u)=\int_{-\infty}^{\infty}f(u,v)\,dv$

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

Thus, $F_X(x)=F(x,\infty)$; $F_Y(y)=F(\infty,y)$

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)\,dy,\quad f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\,dx.$$

Expl. B. $(X,Y)\sim f(x,y)=\begin{cases}\frac{12}{7}(x^2+xy), & 0\le x\le 1,\ 0\le y\le 1.\\ 0, & \text{else.}\end{cases}$

Find $f_X(x)$.

Ans: $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 \frac{12}{7}(x^2+xy)\,dy$

for $0 \le x \le 1$ only

$$= \frac{12}{7}\left(\int_0^1 x^2 dy + \int_0^1 xy\,dy\right)$$

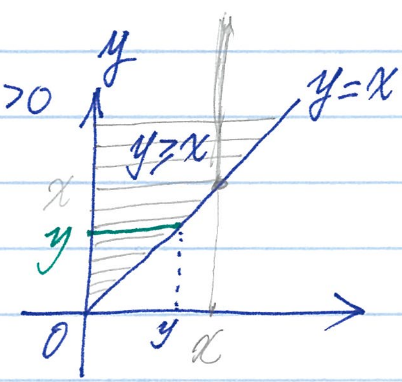
$$= \frac{12}{7}\left(x^2 \cdot 1 + x \cdot \frac{y^2}{2}\Big|_0^1\right)$$

$$= \frac{12}{7}\left(x^2 + x \cdot \frac{1}{2}\right) = \frac{12}{7}\left(x^2 + \frac{x}{2}\right).$$

i.e., $X \sim f_X(x) = \begin{cases} \frac{12}{7}\left(x^2 + \frac{x}{2}\right), & 0 \le x \le 1 \\ 0, & \text{else.} \end{cases}$

Expl D. $(X,Y) \sim f(x,y) = \begin{cases} \lambda^2 e^{-\lambda y}, & 0 \le x \le y,\ \lambda > 0 \\ 0, & \text{else.} \end{cases}$

find $f_X(x)$, $f_Y(y)$.



Ans: $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$,

consider $x \ge 0$, $f_X(x) = \int_x^{\infty} \lambda^2 e^{-\lambda y}\,dy$

$$= \lambda \cdot \left(-e^{-\lambda y}\right)\Big|_x^{\infty} = 0 - \lambda\left(-e^{-\lambda x}\right) = \lambda e^{-\lambda x},\ x \ge 0$$

i.e., $X \sim f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$, $X \sim \exp(\lambda)$.

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^y \lambda^2 e^{-\lambda y}\,dx = \lambda^2 \cdot y e^{-\lambda y},\ y \ge 0$$

$Y \sim \text{Gamm}(2,\lambda)$. $f(y) = \frac{\lambda^2}{\Gamma(2)} y^{2-1} e^{-\lambda y} = \lambda^2 y e^{-\lambda y},\ y \ge 0.$

Uniform dist. over a region R $\quad f(x,y) = \text{constant}$, over R.

Expl E. A point is chosen randomly in a disk of radius 1. In other words, the point is uniformly distributed within the circle. If we let the center of the circle denote the origin and define X and Y to the coordinates of the point chosen, it follows, since (X,Y) is equally likely to be near each point in the circle, that the j.p.d.f of (X,Y) is given $f(x,y)=\begin{cases} c, & x^2+y^2\le 1 \\ 0, & x^2+y^2>1. \end{cases}$ for some constant c.

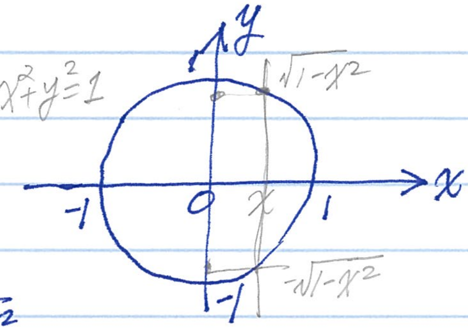
(a) Determine c. (b) $X\sim f_X(x)=?$ $Y\sim f_Y(y)=?$

(c) Calculate the distribution of R, the distance of the point from the origin.

Ans: (a) 1°. $f(x,y)\ge 0$, $c\ge 0$

2°. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy=1$, left side $=\iint_{x^2+y^2\le 1} c\,dx\,dy = c\cdot\pi\cdot 1^2 = c\pi$, $c=1/\pi$.

(b) $f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$,

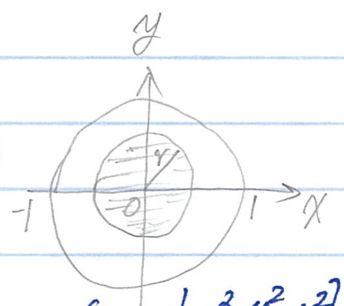


consider $x\in(-1,1)$, $f_X(x)=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,dy$

$$=\frac{2\sqrt{1-x^2}}{\pi}.$$

i.e., $X\sim f_X(x)=\begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, & x\in(-1,1) \\ 0, & \text{else}. \end{cases}$

(c). $R=\sqrt{X^2+Y^2}$. R takes value in $[0,1]$.



$0\le r\le 1$

$P(R\le r) = P(\sqrt{X^2+Y^2}\le r) = P(X^2+Y^2\le r^2)$

$=\iint_A f(x,y)\,dx\,dy = \iint_A \frac{1}{\pi}\,dx\,dy = \frac{1}{\pi}\cdot\text{area}(A)$ $\quad A=\{(x,y)\mid x^2+y^2\le r^2\}$

$=\frac{1}{\pi}\cdot\pi\cdot r^2 = r^2.$

thus $f_R(r)=F_R'(r)=(r^2)_r'=2r$, i.e. $R\sim f_R(r)=\begin{cases} 2r, & 0\le r\le 1 \\ 0, & \text{else}. \end{cases}$