Lectures 15-16

Let us recall some notations.

$$\begin{aligned} F(a-,b-) &:= \lim_{(x,y)\uparrow(a,b)} F(x,y) \\ F(a-,b) &:= \lim_{x\uparrow a} F(x,b) \\ F(a,b-) &:= \lim_{y\uparrow a} F(a,y), \end{aligned}$$

where $(x,y)\uparrow(a,b)$ means $(-\infty,\ x]\times(-\infty,y]\uparrow(-\infty,\ a)\times(-\infty,\ b)$.

Lemma 0.1 *Let $F:\mathbb{R}^2\to\mathbb{R}$ be a distribution function and (X,Y) be a corresponding random vector and let $x,y\in\mathbb{R}$.*
(i)

$$P(X=x,Y=y) \ = \ F(x,y)-F(x-,y)-F(x,y-)+F(x-,y-).$$

(ii) F is continuous at (x,y) iff $F(x,y)=F(x-,y-)$.

Proof: (Reading Exercise)
Proof of (i)
Note that

$$\bigcap_{n=1}^{\infty}\Big\{x-\frac{1}{n}<X\le x, y-\frac{1}{n}<Y\le y\Big\}=\{X=x,Y=y\}.$$

Hence

$$\begin{aligned} P(X=x,Y=y) &= \lim_{n\to\infty} P\Big(x-\frac{1}{n}<X\le x, y-\frac{1}{n}<Y\le y\Big) \\ &= \lim_{n\to\infty}\big(F(x,y)-F(x-\frac{1}{n},y)-F(x,y-\frac{1}{n})+F(x-\frac{1}{n},y-\frac{1}{n})\big) \\ &= F(x,y)-F(x-,y)-F(x,y-)+F(x-,y-). \end{aligned}$$

Proof of (ii) Follows from

$$F(x-,y-)\le F(x-,y)\le F(x,y),\ F(x-,y-)\le F(x,y-)\le F(x,y).$$

Definition 6.5. (i) A distribution function $F:\mathbb{R}^2\to\mathbb{R}$ is said to be of discrete type if there exists a countable set $\{(x_i,y_i)|i\in I\}$ such that

$$\sum_{i\in I}\big(F(x_i,y_i)-F(x_i-,y_i)-F(x_i,y_i-)+F(x_i-,y_i-)\big)=1.$$

$\Big($ in terms of the corresponding random vector, (X, Y) is discrete if

$$\sum_{i \in I} P\big(X = x_i, , Y = y_i\big) \; = \; 1. \Big)$$

(ii) (pmf of discrete random vector) Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,\, y) \;\; = \;\; F(x, y) - F(x-, y) - F(x, y-) + F(x-, y-), x, y \in \mathbb{R}\,.$$

Then f is called pmf of F.

Remark 0.1 *(X, Y) is discrete iff X and Y are discrete random variables.*

Let X, Y be discrete random variables with distribution functions F_1, F_2. Let D_i denote respectively the set of discontinuities of $F_i, i = 1, 2$. Then

$$\begin{aligned} \sum_{x \in D_1, y \in D_2} P(X = x, Y = y) \;\; &= \;\; \sum_{x \in D_1} P(X = x, Y \in D_2) \\ &= \;\; \sum_{x \in D_1} P(X = x) \;\; = \;\; 1. \end{aligned}$$

Hence (X, Y) is discrete. Converse follows by a similar argument. (exercise)

Example 0.1 *Consider the experiment of throwing a die twice independently. Let X_1 denote the number of 1's and X_2 denote the number of 3's. We will write down the joint pmf of X_1 and X_2.*

X_1/X_2	0	1	2
0	$\frac{4}{9}$	$\frac{2}{9}$	$\frac{1}{36}$
1	$\frac{2}{9}$	$\frac{1}{18}$	0
2	$\frac{1}{36}$	0	0

I will illustrate how the pmf is calculated. D_i denote the face shown by the die in the ith throw.

$$\begin{aligned} f(0, 0) \;\; &= \;\; P\big(X_1 = 0, X_2 = 0\big) \\ &= \;\; P\big(D_1 \neq 1, 3 \,\&\, D_2 \neq 1, 3\big) \\ &= \;\; \frac{4}{6} \times \frac{4}{6} \;\; = \;\; \frac{4}{9}. \end{aligned}$$

The above is an example of a lattice distribution, i.e. 'probability masses ' only at the 'lattice' points (i, j); $i, j \in \mathbb{Z}$.

Example 0.2 *Consider the function $F : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$F(x,y) \;=\; \begin{cases} 0 & \text{if} \quad x < 0 \text{ or } y < 0 \\ \frac{1}{3}\min\{x,1\} & \text{if} \quad (x,y) \in [0,\ \infty) \times [0,1) \\ \min\{x,1\} & \text{if} \quad (x,y) \in [0,\ \infty) \times [1,\ \infty). \end{cases}$$

$$F_1(x) \;:=\; \lim_{y\to\infty} F(x,y) \;=\; \begin{cases} 0 & \text{if} \quad x < 0 \\ x & \text{if} \quad 0 \le x < 1 \\ 1 & \text{if} \quad x \ge 1 \end{cases}$$

and

$$F_2(y) \;:=\; \lim_{x\to\infty} F(x,y) \;=\; \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{1}{3} & \text{if} \quad 0 \le x < 1 \\ 1 & \text{if} \quad x \ge 1. \end{cases}$$

Then F_1 is the distribution function of Uniform $(0,\ 1)$ random variable, say X and F_2 is the distribution function the Bernoulli $(\frac{2}{3})$ random variable, say Y.

Also observe that $F(x,y) = F_1(x)F_2(y)$, $x, y \in \mathbb{R}$. Hence F is the distribution of (X,Y) where X and Y are independent.

It is easy to verify that F is continuous on $B_1 \cup B_2 \cup B_2 \cup B_4 \cup B_5$ where

$$B_1 = \{(x,y) | x < 0 \text{ or } y < 0\},\ B_2 = \ (0,\ 1) \times (0,\ 1),$$

$$B_3 = \ (0,1) \times (1,\ \infty),\ B_4 = (1,\infty) \times (0,\ 1),\ B_5 = (1,\ \infty) \times (1,\infty).$$

Consider the boundary sets

$$C_1 = \{(0,y) \,|\, 0 < y < 1\}, C_2 \;=\; \{(x,0) \,|\, 0 < x < 1\},$$

$$C_3 = \{(1,y) \,|\, 0 < y < 1\}, C_4 \;=\; \{(x,1) \,|\, 0 < x < 1\},$$

$$D_1 = \{(0,y) \,|\, 1 < y < \infty\}, D_2 \;= \{(1,y) \,|\, 1 < y < \infty\},$$

$$D_3 = \{(x,0) \,|\, 1 < x < \infty\}, D_4 \;= \{(x,1) \,|\, 1 < x < \infty\},$$

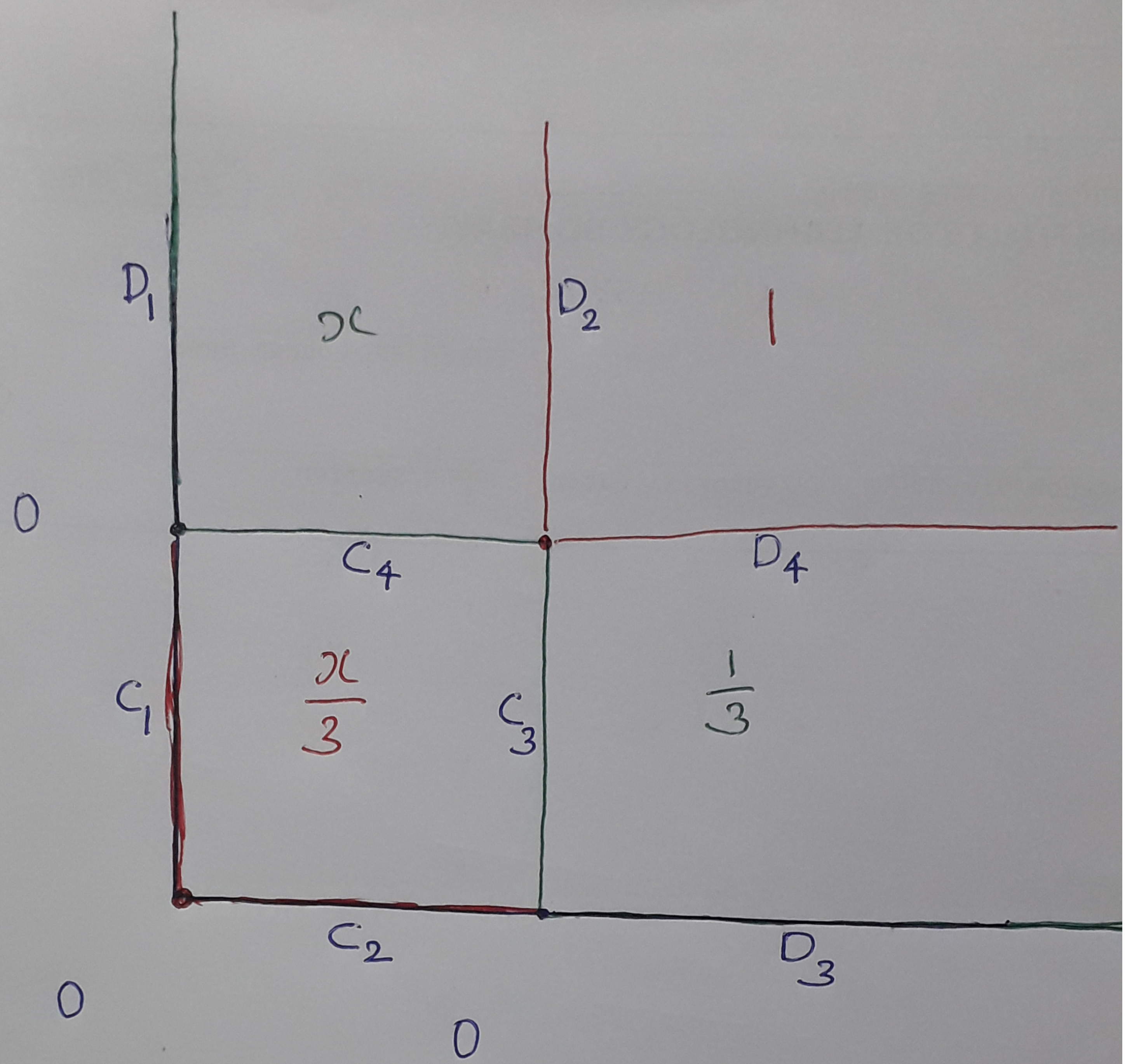
and the corner points $(0,0), (1,0), (0,1)$ and $(1,1)$.

For $(x,0) \in C_2$, $F(x-,0-) = 0 = F(x,0)$. Hence F is continuous on C_2. Similarly, for $(x,1) \in C_3$, $F(x-,1-) = \frac{1}{3} = F(x,1)$. So F is continuous on C_3. For $(0,y) \in C_1$, $F(0-,y-) = 0 = F(0,y)$, hence F is continuous on C_1.

For $(x,1) \in C_4$, $F(x-,1-) = \frac{x}{3} \ne x = F(x,1)$. Hence F is not continuous on $C4$. Similarly F is not continuous on D_1, D_2, D_3 and D_4.

Left limit at the corner points are $F(0-,0-) = 0 = F(0,0), F(1-,0-) = 0 \ne \frac{1}{3} = F(1,0), F(0-,1-) = 0 = F(0,1), F(1-,1-) = \frac{1}{3} \ne 1 = F(1,1)$.

Hence F is not left continuous at $(1,0)$ and $(1,1)$.



D_1
x
D_2
1
0
C_4
D_4
C_1
$\frac{x}{3}$
C_3
$\frac{1}{3}$
C_2
D_3
0
0

Example 0.3

$$G(x,y) \;=\; \begin{cases} 0 & \text{if} \quad x<0 \text{ or } y<0 \\ \frac{x}{4} & \text{if} \quad (x,y)\in[0,\ 1)\times[0,1) \\ x & \text{if} \quad (x,y)\in[0,\ 1)\times[1,\ \infty) \\ \frac{1}{3} & \text{if} \quad (x,y)\in[1,\ \infty)\times[0,1) \\ 1 & \text{if} \quad (x,y)\in[1,\ \infty)\times[1,\ \infty). \end{cases}$$

Verify that G is a distribution function by showing the properties (i), (ii) and (iii).

Let F_1, F_2 be respectively be the distribution function of uniform $(0,\ 1)$ and Bernoulli $(\frac{2}{3})$ random variables. Then

$$\lim_{y\to\infty} G(x,y) = F_1(x),\ \lim_{x\to\infty} G(x,y) = F_2(y).$$

(exercise)

Definition 6.6. (joint pdf of continuous random vector) Let $(X,\,Y)$ be a continuous random vector (i.e., distribution function F is continuous). If there exists a function $f:\mathbb{R}^2\to\mathbb{R}$ such that

$$F(x,\,y) \;=\; \int_{-\infty}^{y}\int_{-\infty}^{x} f(t,\,s)dtds,\quad \forall\ \ x,y\in\mathbb{R}$$

then f is called the pdf of (X,Y), in otherwords, the joint pdf of the random variables X and Y.

Theorem 0.1 *Let (X,Y) be a continuous random vector with pdf f. Then*

$$P\big((X,\,Y)\in B\big) \;=\; \iint_B f(x,\,y)dxdy,\ \ B\in\mathcal{B}_{\mathbb{R}^2}\,.$$

Proof. (Proof is not part of the syllabus and but the proof is given as a reading exercise which is optional)

Note that L.H.S of the equality corresponds to the law of $(X,\,Y)$.

Let $\mathcal{F}_0$ denote the set of all finite union of rectangles in $\mathbb{R}^2$. Then $\mathcal{F}_0$ is a field (exercise for the student).

Set

$$\begin{aligned} \mu_1(B') \;&=\; \iint_{B'} f(x,\,y)dxdy, \qquad B'\in\mathcal{B}_{\mathbb{R}^2} \quad \text{and} \\ \mu_2(B') \;&=\; P\big((X,\,Y)\in B'\big), \qquad B'\in\mathcal{B}_{\mathbb{R}^2} \end{aligned}$$

Then μ_1, μ_2 are probability measures on $\mathcal{B}_{\mathbb{R}^2}$ and $\mu_1 \;=\; \mu_2$ on $\mathcal{F}_0$. Hence, using extension theorem, we have

$$\mu_1 \;=\; \mu_2 \text{ on } \mathcal{B}_{\mathbb{R}^2}.$$

i.e.,

$$P\big((X,\,Y) \in B\big) \;=\; \iint_B f(x,\,y)dxdy,\;\; B \in \mathcal{B}_{\mathbb{R}^2}\,.$$

Remark 0.2 *The integral in Theorem 0.1 is in general not understood in the Riemann integral sense but for all our computations, the integral will be in the Riemann sense. Note mostly will be considering Riemann integrable functions in domains which are either rectangles or of the form*

$$B \;=\; \{(x,y)|a \le x \le b, \varphi_1(x) \le y \le \varphi_2(x)\},$$

for some continuous function $\varphi_1, \varphi_2 : [a,\;b] \to \mathbb{R}, \varphi_1 \le \varphi_2$*, or of the form*

$$B' \;=\; \{(x,y)|c \le y \le d, \psi_1(y) \le x \le \psi_2(y)\},$$

for some continuous function $\psi_1, \psi_2 : [c,\;d] \to \mathbb{R}, \psi_1 \le \psi_2$*. We call these domains as* **elementary domains***.*

For the elementary domains, using Theorem 0.1, we get

$$\begin{aligned} P\big((X,Y) \in B\big) &= \int_a^b \int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)dydx, \\ P\big((X,Y) \in B'\big) &= \int_c^d \int_{\psi_1(y)}^{\psi_2(y)} f(x,y)dxdy, \end{aligned}$$

where (X,Y) *is with pdf* f*.*

Now we will give some important class of multi valued distribution functions.

Example 0.4 *(Uniform distribution on disc of radius R) Consider the function*

$$f(x,y) \;=\; \begin{cases} \frac{1}{\pi R^2} & \text{if } \;x^2+y^2 \le R^2 \\ 0 & \text{if } \;\text{otherwise.} \end{cases}$$

Then f *is called joint density of uniform distribution on* $\{(x,y)|x^2+y^2 \le R^2\}$*. In fact student can try to construct a pair of random variables* (X,Y)

on some probability space $(\Omega, \mathcal{F}, P)$ such that its joint distribution function is given by

$$P\big(X \leq x, Y \leq y\big) \;=\; \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)dxdy.$$

Any Guess! Hint: see how uniform $(0,1)$ random variable is contructed.

Example 0.5 *Let X, Y be two random variables with joint pdf given by*

$$f(x,\,y) \;=\; \frac{\sqrt{3}}{4\pi}e^{-\frac{(x^2-xy+y^2)}{2}},\;\; x,\,y \in \mathbb{R}\,.$$

If $f_X(.)$, $f_Y(.)$ denote the marginal pdfs of X and Y respectively, then

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,\,y)dy \\
&= \frac{\sqrt{3}}{4\pi}\int_{-\infty}^{\infty} e^{-\frac{(x^2-xy+y^2)}{2}}dy \\
&= \frac{\sqrt{3}}{4\pi}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-\frac{x}{2})^2}e^{\frac{-3x^2}{8}}dy \\
&= \frac{\sqrt{3}}{4\pi}e^{-\frac{3x^2}{8}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-\frac{x}{2})^2}dy \\
&= \frac{\sqrt{3}}{2}\frac{1}{\sqrt{2\pi}}e^{-\frac{3x^2}{8}}\,.
\end{aligned}$$

Therefore

$$X \sim N(0,\,\frac{4}{3})\,.$$

Here $X \sim N(m, \sigma^2)$ means X is normally distributed with parameters m and σ^2. Similarly,

$$\begin{aligned}
f_y(y) &= \int_{-\infty}^{\infty} f(x,\,y)dy \\
&= \frac{\sqrt{3}}{2}\frac{1}{\sqrt{2\pi}}e^{-\frac{3y^2}{8}}\,.
\end{aligned}$$

Therefore

$$Y \sim N(0,\,\frac{4}{3})\,.$$

Also note that X and Y are dependent since,

$$f(x,\,y) \neq f_X(x)f_Y(y)\,,$$

see exercise.

Real valued transformations of multi random variables: In this subsection, we look at transformations of (X,Y) of the form $\varphi \circ (X,Y)$, i.e. $\varphi(X,Y)$ where $\varphi : \mathbb{R}^2 \to \mathbb{R}$, i.e. real valued functions of (X,Y), for example $X+Y, XY$ etc. One can write down the distribution function $Z = \varphi \circ (X,Y)$. Set

$$A_z = \{(x,y) | \varphi(x,y) \leq z\}, \; z \in \mathbb{R}.$$

Then

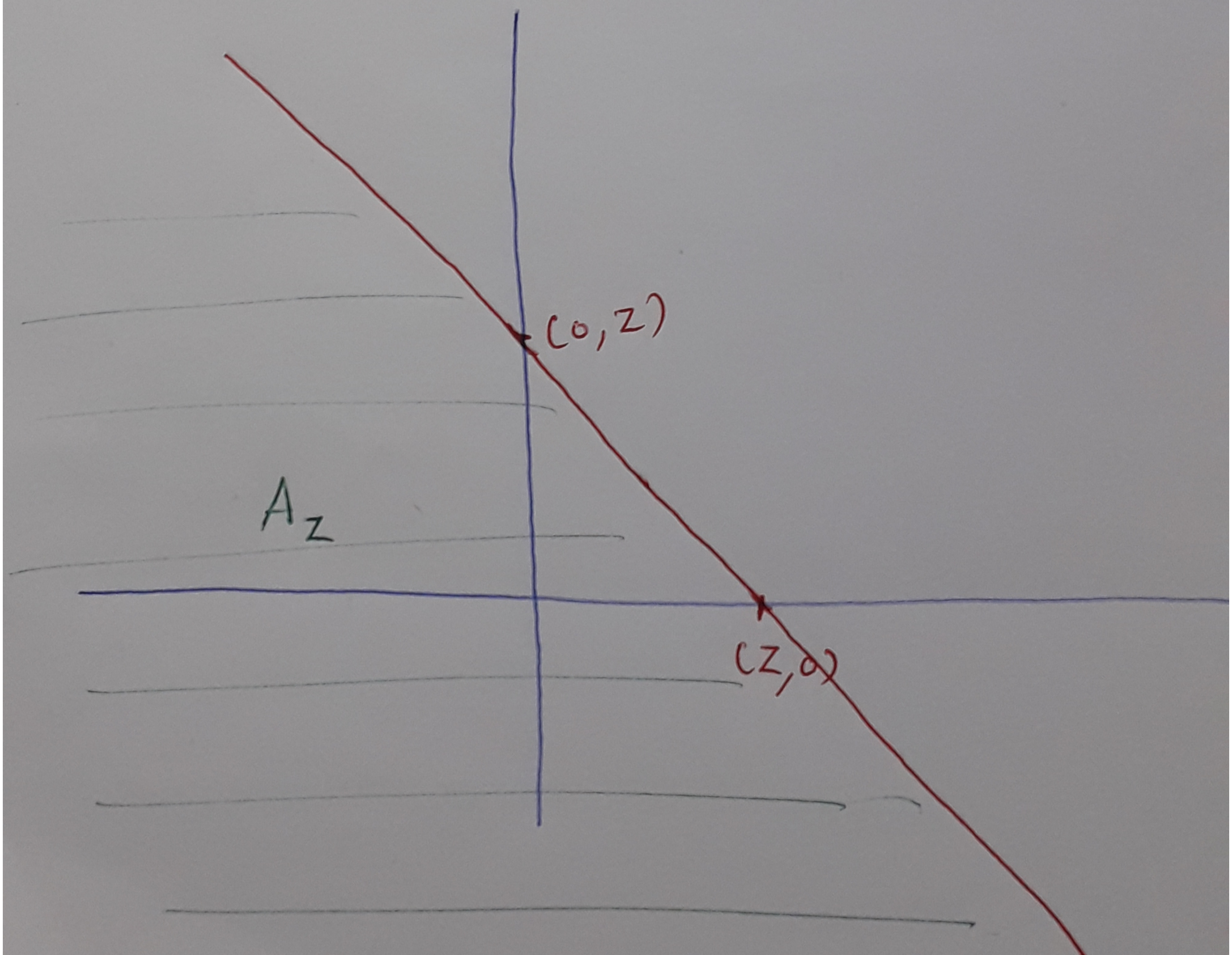
$$F_Z(z) = P\big(\varphi(X,Y) \in (-\infty, z]\big) = P\big((X,Y) \in A_z\big),$$

where F_Z denote the distribution function of $Z = \varphi(X,Y)$. Hence it is all about understanding A_z and then computing $P\big((X,Y) \in A_z\big)$.

Example 0.6 *(Distribution of sum) Let X, Y be with joint pdf f. Then one can compute the distribution of the sum $Z = X + Y$ as follows. Note*

$$A_z = \{(x,y) | -\infty < x < \infty, \; -\infty < y \leq z - x\}$$

is an elementary domain, in fact it is the region 'below' the line $x + y = z$.



(0,z)
A_z
(z,0)

Hence

$$
\begin{aligned}
F_Z(z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)dydx \\
(\text{put } t = y+x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{z} f(x,t-x)dtdx \\
(\text{change order of integration}) &= \int_{-\infty}^{z}\Big[\int_{-\infty}^{\infty} f(x,t-x)dx\Big]dt.
\end{aligned}
$$

Hence $X+Y$ has pdf given by

$$
f_Z(z) = \int_{-\infty}^{\infty} f(x,z-x)dx.
$$

Corollary 0.1 *Let X, Y be independent random variables with joint pdf f. Then the pdf of $X+Y$ is given by*

$$
f_{X+Y}(z) = f_X * f_Y(z),
$$

where $f_X * f_Y$ denote the convolution of f_X and f_Y and is defined as

$$
f_X * f_Y(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x)dx = \int_{\mathbb{R}} f_X(z-y) f_Y(y)dy\,.
$$

Proof. The proof (exercise) follows immediately from $f(x,y) = f_X(x)f_Y(y)$ and the above example.

Example 0.7 *Let X, Y be independent exponential random variables with parameters λ_1 and λ_2 respectively. Then*

$$
f_X(x) = \begin{cases} \lambda_1 e^{-\lambda_1 x} & \text{if } \quad x \geq 0 \\ 0 & \text{if } \quad x < 0\,. \end{cases}
$$

f_Y is given similarly. Now for $z \leq 0$, clearly $f_X * f_Y(z) = 0$. For $z > 0$,

$$
\begin{aligned}
f_X * f_Y(z) &= \int_0^z \lambda_1 e^{-\lambda_1 x}\lambda_2 e^{-\lambda_2(z-x)}dx \\
&= \begin{cases} \frac{\lambda_1\lambda_2}{\lambda_2-\lambda_1}(e^{-\lambda_1 z} - e^{-\lambda_2 z}) & \text{if } \quad \lambda_1 \neq \lambda_2 \\ \lambda^2 z e^{-\lambda z} & \text{if } \quad \lambda_1 = \lambda_2 = \lambda \end{cases}.
\end{aligned}
$$

The above gives the pdf of $X+Y$.

Linear transformations of multi random variables: In this section we look at when $\varphi(x,y) = (x,y)A$, where A is a 2×2 invertible (i.e. non singular) matrix. For example, when

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

then A rotates (X,Y) through an angle θ in the counter clockwise direction. To find the distribution of $(U,V) = (X,Y)A$, we use the following change of variable formula. Let $\varphi = (\varphi_1, \varphi_2) : \mathbb{R}^2 \to \mathbb{R}^2$ be a map (with continuously differentiable partial derivatives) such that the Jacobian at (u,v)

$$J(\varphi_1(u,v), \varphi_2(u,v)) = \det \begin{pmatrix} \frac{\partial \varphi_1}{\partial u} & \frac{\partial \varphi_1}{\partial v} \\ \frac{\partial \varphi_2}{\partial u} & \frac{\partial \varphi_2}{\partial v} \end{pmatrix} \neq 0$$

Then the area element under the mapping $(u,v) \mapsto (x = \varphi_1(u,v), y = \varphi_2(u,v))$ makes the transformation $dudv \mapsto dxdy = |J(\varphi_1(u,v), \varphi_2(u,v))| dudv$ [1] This leads to the following change of variable formula.

Theorem 0.2 *Let D be a region in $\mathbb{R}^2$ and $f : D \to \mathbb{R}$ be continuous. Let $\varphi = (\varphi_1, \varphi_2) : O \to \mathbb{R}^2$ be such that*

(i) φ is one to one

[1] Here note that infinitesemal (small) rectangle $[u,\ u + du] \times [v, v + dv]$ i.e., $dv \times dv$ is approximately mapped to the parallelogram $dx \times dy$. Now dx is the vector joining the points $(\varphi_1(u,v), \varphi_2(u,v))$ and $(\varphi_1(u + du, v), \varphi_2(u + du, v))$ and hence

$$\begin{aligned} dx &= (\varphi_1(u + du, v) - \varphi_1(u,v), \varphi_2(u + du, v) - \varphi_2(u,v)) \\ &\sim \Big(\frac{\partial \varphi_1(u,v)}{\partial u} du,\ \frac{\partial \varphi_2(u,v)}{\partial u} du\Big). \end{aligned}$$

Similarly

$$dy \sim \Big(\frac{\partial \varphi_1(u,v)}{\partial v} dv,\ \frac{\partial \varphi_2(u,v)}{\partial v} dv\Big).$$

Hence

$$\begin{aligned} dx &\sim \frac{\partial \varphi_1(u,v)}{\partial u} du\, \mathbf{i} + \frac{\partial \varphi_2(u,v)}{\partial u} du\, \mathbf{j} + 0\, \mathbf{k} \\ dy &\sim \frac{\partial \varphi_1(u,v)}{\partial v} dv\, \mathbf{i} + \frac{\partial \varphi_2(u,v)}{\partial v} dv\, \mathbf{j} + 0\, \mathbf{k}. \end{aligned}$$

$$\begin{aligned} \text{Area of the parallelogram} &= |dx \times dy| \\ &= \Big| \mathbf{k} \Big(\frac{\partial \varphi_1(u,v)}{\partial u} \frac{\partial \varphi_2(u,v)}{\partial v} dudv - \frac{\partial \varphi_2(u,v)}{\partial u} \frac{\partial \varphi_1(u,v)}{\partial v} dudv \Big) \Big| \\ &= |J(\varphi_1(u,v), \varphi_2(u,v))|\, dudv. \end{aligned}$$

(ii) φ_1, φ_2 *have continuous paritial derivatives in* O*,* O *is an open set in* $\mathbb{R}^2$

(iii) $J(\varphi_1(u,v), \varphi_2(u,v)) \neq 0$ *in* O

(iv) There exists $E \subseteq O$ *such that* E *is elementary and* $\varphi(E) = D$*.*

Then

$$\boxed{\iint_D f(x,y)dxdy = \iint_E f(\varphi_1(u,v), \varphi_2(u,v))|J(\varphi_1(u,v), \varphi_2(u,v))|dudv.}$$

As an application of the above change of variable formula, we have the following result.

Corollary 0.2 *Let* (X,Y) *be a random vector in* $\mathbb{R}^2$ *with joint pdf* f *and* A *be a non singular* 2×2 *matrix. Then* $(U,V) = (X,Y)A$ *is with pdf* g *given by*

$$g(u,v) = \frac{1}{|\text{det}A|} f((u,v)A^{-1})$$

Proof: For $u, v \in \mathbb{R}$, set $R_{uv} = (-\infty, u] \times (-\infty, v]$ and $I = \varphi(R_{uv})$, where $\varphi(u,v) = (u,v)A^{-1}$. Now the distribution function $F_{(U,V)}$ of (U,V) is given by

$$\begin{aligned}
F_{(U,V)}(u,v) &= P\big((U,V) \in R_{uv}\big) \\
[\text{using } (x,y) \mapsto (x,y)A, \text{ a bijection}] &= P\big((X,Y) \in I\big) \\
&= \iint_I f(x,y)dxdy \\
(\text{Theorem 0.2 to } (u,v) \mapsto (u,v)A^{-1}) &= \iint_{R_{uv}} f((u,v)A^{-1}) \frac{1}{|\text{det}A|} dudv.
\end{aligned}$$

Hence (U,V) has a pdf say g and is given by

$$g(u,v) = \frac{1}{|\text{det}A|} f((u,v)A^{-1}).$$

$\square$

Multinormal distribution. X is said to be a non degenerate multinormal with parameters $m = (m_1, m_2)$ and $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$ if the joint pdf f_X is given by

$$f_X(x_1, x_2) = \frac{1}{2\pi\sqrt{\text{det}\Sigma}} e^{-\frac{1}{2}(x-m)\Sigma^{-1}(x-m)^{\perp}},$$

where Σ is symmetric positive definite and $(x-m)^{\perp}$ is the column vector corresponding to the row vector $(x_1 - m_1, x_2 - m_2)$. Also Σ is positive definite implies that $|\sigma_{12}| \leq \sigma_1\sigma_2$. Rewrite Σ as

$$\Sigma \;=\; \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

then (exercise)

$$f_X(x_1, x_2) \;=\; \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-m_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right]}.$$

Now we will see the marginal pdfs of multinormal.

Essence of the calculation is 'completing the square'. I will do 'completing the square' separately. Consider

$$\begin{aligned}
&\tfrac{(x_1-m_1)^2}{\sigma_1^2} - \tfrac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \tfrac{(x_2-m_2)^2}{\sigma_2^2} \\
&= \left[\left(\tfrac{(x_2-m_2)}{\sigma_2}\right)^2 - \tfrac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \left(\tfrac{\rho(x_1-m_1)}{\sigma_1}\right)^2\right] + (1-\rho^2)\left(\tfrac{(x_1-m_1)}{\sigma_1}\right)^2 \\
&= \left(\tfrac{(x_2-m_2)}{\sigma_2} - \tfrac{\rho(x_1-m_1)}{\sigma_1}\right)^2 + (1-\rho^2)\left(\tfrac{(x_1-m_1)}{\sigma_1}\right)^2 \\
&= \left(\tfrac{(x_2-m_2)}{\sigma_2} - a\right)^2 + (1-\rho^2)\left(\tfrac{(x_1-m_1)}{\sigma_1}\right)^2,
\end{aligned}$$

where

$$a = \frac{\rho(x_1 - m_1)}{\sigma_1}.$$

Now

$$\begin{aligned}
\int_{-\infty}^{\infty} f_X(x_1, x_2)dx_2 \;&=\; \frac{e^{-\frac{1}{2}\left(\frac{(x_1-m_1)}{\sigma_1}\right)^2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_2-m_2)}{\sigma_2}-a\right)^2} dx_2 \\
\left[\text{put } x = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{x_2-m_2}{\sigma_2} - a\right)\right] \;&=\; \frac{e^{-\frac{1}{2}\left(\frac{(x_1-m_1)}{\sigma_1}\right)^2}}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \\
&=\; \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{(x_1-m_1)}{\sigma_1}\right)^2}.
\end{aligned}$$

Hence $X_1 \sim N(m_1, \sigma_1)$. Similarly $X_2 \sim N(m_2, \sigma_2)$.

Theorem 0.3 *Let $X = (X_1, X_2)$ be a multinormal non degenerete random variable with parameters m and Σ. Then for any $\alpha, \beta \in \mathbb{R}$ $\alpha X_1 + \beta X_2$ is a normal random variable.*

Proof: Since $X - m$ is normal with paramenters 0 and Σ, it is enough to prove the theorem when $m = 0$. (exercise)

Let A be a 2×2 symmetric matrix such that $AA^{\perp} = \Sigma$ [Here the choice of A is $\Sigma^{\frac{1}{2}}$ and $\Sigma^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}}P^{-1}$ where Λ is the diagonal matrix (of eigen values of Σ and $\Sigma = P\Lambda P^{-1}$) and hence $\Lambda^{\frac{1}{2}}$ is the diagonal matrix with diagonal entries as the square root of the eigen values of Λ]. Define $Y = X\Sigma^{-\frac{1}{2}}$, then using Theorem 0.2, the pdf g of Y exists and is given by

$$\begin{aligned} g(y_1, y_2) &= |\det A| f(yA) \\ &= \sqrt{\det\Sigma}\frac{1}{2\pi\sqrt{\det\Sigma}}e^{-\frac{1}{2}yA\Sigma^{-1}Ay^{\perp}} \\ &= \frac{1}{2\pi}e^{-\frac{1}{2}\|y\|^2},\ y = (y_1, y_2). \end{aligned}$$

Hence Y is multi normal with parameters 0 and I. Therefore $g(y_1, y_2) = g_{Y_1}(y_1)g_{Y_2}(y_2)$. This implies that Y_1 and Y_2 are independent normal random variables.

Now

$$\alpha X_1 + \beta X_2 = X\begin{pmatrix}\alpha \\ \beta\end{pmatrix} = Y\Sigma^{\frac{1}{2}}\begin{pmatrix}\alpha \\ \beta\end{pmatrix} = aY_1 + bY_2$$

for some $a, b \in \mathbb{R}$ and hence is a normal random variable. (exercise)
This completes the proof. $\square$

Remark 0.3 *The proof of Theorem 0.3 tells us some thing more. Let $X \sim N(m,\ \Sigma)$, i.e. X is a multinormal random variable with paramenters m and Σ. Set $\bar{X} = X - m$, then $\bar{X} \sim N(0, \Sigma)$.*

Then $\bar{Y} = \bar{X}\Sigma^{-\frac{1}{2}} \sim N(0, I)$. Hence

$$Y := X\Sigma^{-\frac{1}{2}} = m\Sigma^{-\frac{1}{2}} + \bar{X}\Sigma^{-\frac{1}{2}} \sim N(\mu\Sigma^{-\frac{1}{2}}, I).$$

Note the above is a generalization to the multidimentional case of the following result for the normal random variables.

$$X \sim N(m, \sigma^2),\ \text{then}\ aX \sim N(a\,m, a^2\sigma^2).$$

Theorem 0.3 leads to a more general definition multinormal which includes multinormal random variables which are degenerate also.

Definition 0.1 *A random vector $X = (X_1, X_2)$ is said to be multinormal if $\alpha X_1 + \beta X_2$ is normal for all $\alpha, \beta \in \mathbb{R}$.*

Example 0.8 *Let $X_1 \sim N(0,1)$ and $X = (X_1, -X_1)$. Then any linear combination of the components of X is normally distributed. Also X doesnot have a joint density function. To see this, let $L = \{(x,y)|x+y=0\}$, the graph of $x+y=0$. Then*

$$P\{X \in L\} = P(\Omega) = 1.$$

Now suppose X has a joint pdf f, then for $L_n \;=\; L \cap [-n,\; n]$,

$$P\{X \in L_n\} \;=\; \iint_{L_n} f(x,y)dxdy = 0, \text{ for all } n \geq 1.$$

Now

$$P\{X \in L\} \;=\; \lim_{n\to\infty} P\{X \in L_n\} = 0$$

a conradiction to $P\{X \in L\} = 1$. Hence X doesn't have a density. i.e., X is an example of a degenerate multinormal distribution.