6.1. Joint distribution functions.

Definition 6.1. Given two real-valued random variables X, Y (defined on the same sample space), the **joint (cumulative) distribution function** of X and Y is defined by

$$F(a, b) = P(X \le a, Y \le b), \quad \forall a, b \in \mathbb{R}.$$

The distribution function of X (or Y) is called the **marginal** distribution.

Remark 6.1. Let F_X and F_Y be the distribution functions of X and Y and let F be the joint distribution function of X, Y. Then, for $a \in \mathbb{R}$,

$$F_X(a) = P(X \le a) = P(X \le a, Y < \infty) = \lim_{b\to\infty} P(X \le a, Y \le b) = \lim_{b\to\infty} F(a, b).$$

For convenience, we write $F(a, \infty)$ for the last limit. Similarly, one has

$$\begin{aligned} P(X > a, Y > b) &= 1 - P(X \le a \text{ or } Y \le b) \\ &= 1 - P(X \le a) - P(Y \le b) + P(X \le a, Y \le b) \\ &= 1 - F_X(a) - F_Y(b) + F(a, b). \end{aligned}$$

Definition 6.2. If X, Y are discrete, the **joint probability mass function** of X, Y are defined by

$$p(x, y) = P(X = x, Y = y).$$

The probability mass functions of X and Y are called **marginal** probability mass functions.

Remark 6.2. If X, Y are discrete with p.m.f. p_X and p_Y, then

$$p_X(x) = P(X = x, Y < \infty) = \sum_{y:p_Y(y)>0} P(X = x, Y = y) = \sum_{y:p(x,y)>0} p(x, y).$$

Example 6.1. Consider the experiment of rolling a fair dice twice independently. Let X be the point of the first roll and Y be the sum of two rolls. Let p be the joint p.m.f. of X and Y. Then, for $1 \le i, j \le 6$,

$$p(i, i + j) = P(X = i, Y = i + j) = P(\text{two rolls result in } (i, j)) = \frac{1}{36}.$$

Definition 6.3. Two random variables X, Y are **jointly continuous** if there is an integrable function $f : \mathbb{R}^2 \to [0, \infty)$ such that

$$P((X, Y) \in C) = \iint_C f(x, y)dxdy,$$

for any (Borel) set $C \subset \mathbb{R}^2$. Here, f is called a **joint probability density function**.

Remark 6.3. The above definition is unchanged if C is restricted to the set $(-\infty, a] \times (-\infty, b]$ with $a, b \in \mathbb{R}$. This means that X, Y are jointly continuous if and only if the joint distribution function F satisfies

$$F(a, b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y)dxdy \quad \forall a, b \in \mathbb{R}.$$

In particular, if f is uniformly continuous, then $F(a, b)$ is differentiable and $F_{ab}(a, b) = f(a, b) = F_{ba}(a, b)$.

Remark 6.4. If X, Y are jointly continuous, then X and Y are continuous. To see the details, let f be the joint p.d.f. of X, Y. Then, for any (Borel) set $A \subset \mathbb{R}$,

$$P(X \in A) = P(X \in A, Y \in \mathbb{R}) = \int_A \left(\int_{\mathbb{R}} f(x,y)dy \right) dx.$$

This implies that $f_X(x) := \int_{-\infty}^{\infty} f(x,y)dy$ is the p.d.f. of X. Similarly, one can show that $f_Y(y) := \int_{-\infty}^{\infty} f(x,y)dx$ is the p.d.f. of Y. Here, f_X and f_Y are also called the **marginal** probability density functions. Note that $f(x,y) \neq f_X(x)f_Y(y)$ in general.

Example 6.2. Let $R > 0$. Suppose that X, Y are jointly continuous with joint p.d.f.

$$f(x,y) = \begin{cases} c & \text{if } x^2 + y^2 < R^2 \\ 0 & \text{otherwise} \end{cases}.$$

To determine the value of c, consider the following equalities,

$$1 = \iint_{\mathbb{R}^2} f(x,y)dxdy = c \iint_D dxdy = c\pi R^2.$$

This implies $c = 1/(\pi R^2)$. To see the marginal density, let $|x| < R$. Then, one has

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{\pi R^2} \int_{\{y: y^2 < R^2 - x^2\}} dy = \frac{2\sqrt{R^2 - x^2}}{\pi R^2}.$$

For $|x| \geq R$, $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = 0$. Similarly, one may derive $f_Y = f_X$.

Remark 6.5. For random variables $X_1, X_2, ..., X_n$, their joint distribution function is defined by

$$F(a_1, a_2, ..., a_n) = P(X_i \leq a_i, 1 \leq i \leq n), \quad \forall (a_1, ..., a_n) \in \mathbb{R}^n.$$

If X_i's are discrete, then their joint p.m.f. is defined by

$$p(x_1, ..., x_n) = P(X_1 = x_1, ..., X_n = x_n).$$

We say that $X_1, ..., X_n$ are jointly continuous if there is a function $f : \mathbb{R}^n \to [0, \infty)$ such that

$$F(a_1, ..., a_n) = \int_{-\infty}^{a_n} \cdots \int_{-\infty}^{a_1} f(x_1, ..., x_n)dx_1 \cdots dx_n.$$

Here, f is called the joint p.d.f. of $X_1, ..., X_n$.

Example 6.3. Consider the experiment of n independent draws with replacement from a box of n balls. Suppose those balls are of r different colors, say $c_1, ..., c_r$, and the proportion of balls with color c_i is p_i. Let X_i denote the number of times that balls with color c_i are selected. Then, for $k_1 + \cdots + k_r = n$,

$$P(X_i = k_i, 1 \leq i \leq n) = \binom{n}{k_1, k_2, ..., k_r} p_1^{k_1} \cdots p_r^{k_r}.$$

6.2. **Independence of random variables.**

Definition 6.4. Two random variables X, Y are **independent** if, for any (Borel) sets A, B,

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

Remark 6.6. In the above definition, X and Y are independent if and only if, for any $a, b \in \mathbb{R}$,

$$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b),$$

which is equivalent to

$$F(a,b) = F_X(a)F_Y(b),$$

where F is the joint distribution function of X, Y.

If X, Y are discrete with joint p.m.f. p and marginal p.m.f.s p_X and p_Y. Then, X and Y are independent if and only if

$$p(x, y) = p_X(x)p_Y(y) \quad \forall x, y.$$

If X, Y are jointly continuous with joint p.d.f. f and marginal p.d.f.s f_X, f_Y, then X and Y are independent if and only if

$$f(x, y) = f_X(x)f_Y(y) \quad \forall x, y.$$

Theorem 6.1. *Let X, Y be random variables.*

(1) *If X, Y are discrete with joint p.m.f. p, then X, Y are independent if and only if $p(x, y) = g(x)h(y)$ for all x, y, where g and h are real-valued functions.*
(2) *If X, Y are jointly continuous with joint p.d.f., then X, Y are independent if and only if $f(x, y) = g(x)h(y)$ for all x, y, where g, h are real-valued functions.*

Proof. We show the continuous case, while the discrete case can be proved in a similar way. If X, Y are independent, then $f(x, y) = f_X(x)f_Y(y)$ and one may set $g = f_X$and $h = f_Y$. Suppose that $f(x, y) = g(x)h(y)$ and, without loss of generality, assume $g \geq 0$ and $h \geq 0$. Note that, for any (Borel) subsets A, B of $\mathbb{R}$,

$$P(X \in A, Y \in B) = \int_{A\times B} f(x, y)dxdy = \left(\int_A g(x)dx\right)\left(\int_B h(y)dy\right).$$

Letting $A = B = \mathbb{R}$ gives

$$1 = P(X \in \mathbb{R}, Y \in \mathbb{R}) = \left(\int_{-\infty}^{\infty} g(x)dx\right)\left(\int_{-\infty}^{\infty} h(y)dy\right),$$

which implies $\int_{-\infty}^{\infty} g(x)dx \neq 0$ and $\int_{-\infty}^{\infty} h(y)dy \neq 0$. Thus, we have

$$P(X \in A) = P(X \in A, Y \in \mathbb{R}) = \int_A \left(\frac{g(x)}{\int_{-\infty}^{\infty} g(x)dx}\right) dx$$

and

$$P(Y \in B) = P(X \in \mathbb{R}, Y \in B) = \int_B \left(\frac{h(y)}{\int_{-\infty}^{\infty} h(y)dy}\right) dy.$$

This means $f_X = g/\int_{-\infty}^{\infty} g(x)dx$ and $f_Y = h/\int_{-\infty}^{\infty} h(y)dy$ and, for all $x, y \in \mathbb{R}$,

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

□

Example 6.4. Suppose that the number of customers entering a store on a given day is a Poisson random variable with parameter λ. Assume that, given a customer entering the store, the condition probability that the customer is male is equal to p and is independent of other customers. Let X and Y denote the number of males and females entering the store on a given day. Then, $X + Y$ is of Poisson(λ) distribution. Let p be the joint p.m.f. of X and Y. Then, for nonnegative integers i, j,

$$\begin{aligned} p(i, j) &= P(X = i, Y = j) = P(X = i, X + Y = i + j) \\ &= P(X = i|X + Y = i + j)P(X + Y = i + j). \end{aligned}$$

Clearly, $P(X+Y=i+j) = e^{-\lambda}\lambda^{i+j}/(i+j)!$. Given $i+j$ persons entering the store, the number of males is exactly a binomial random variable with parameters $(i+j,p)$. This implies

$$P(X=i|X+Y=i+j) = \binom{i+j}{i} p^i(1-p)^j.$$

As a result, we obtain

$$p(i,j) = \binom{i+j}{i} p^i(1-p)^j \frac{e^{-\lambda}\lambda^{i+j}}{(i+j)!} = \left(\frac{e^{-p\lambda}(p\lambda)^i}{i!}\right)\left(\frac{e^{-(1-p)\lambda}[(1-p)\lambda]^j}{j!}\right).$$

By Theorem 6.1, X,Y are independent and X is of Poisson($p\lambda$) distribution and Y is of Poisson($(1-p)\lambda$) distribution.

Example 6.5. Let (X,Y) be the coordinate of a bullet fired at a target. Suppose that

(1) X and Y are continuous and independent, and have differentiable p.d.f.

(2) The joint p.d.f. $f(x,y)$ is spherically symmetric, that is,

$$f(x_1,y_1) = f(x_2,y_2) \quad \text{if} \quad x_1^2+y_1^2 = x_2^2+y_2^2.$$

(3) The marginal p.d.f.s of X and Y, f_X and f_Y, are positive everywhere.

By the second assumption, one may choose a real-value function g such that $f(x,y) = g(x^2+y^2)$. Since $f(x,y) = f_X(x)f_Y(y)$, it is easy to see that f_X and f_Y are even functions. Note that, for $x>0$,

$$g(x) = f(\sqrt{x},0) = f_X(\sqrt{x})f_Y(0).$$

This implies g is differentiable on $(0,\infty)$. Taking the partial derivative of f w.r.t. x yields

$$f_X'(x)f_Y(y) = 2xg'(x^2+y^2).$$

Dividing both sides with $2xf(x,y)$ gives

$$\frac{f_X'(x)}{2xf_X(x)} = \frac{g'(x^2+y^2)}{g(x^2+y^2)}.$$

Observe that if $x_1^2+y_1^2 = x_2^2+y_2^2 > 0$ with $x_1 \neq 0$ and $x_2 \neq 0$, then

$$\frac{f_X'(x_1)}{2x_1f_X(x_1)} = \frac{g'(x_1^2+y_1^2)}{g(x_1^2+y_1^2)} = \frac{g'(x_2^2+y_2^2)}{g(x_2^2+y_2^2)} = \frac{f_X'(x_2)}{2x_2f_X(x_2)}.$$

This means that the mapping $f_X'(x)/[xf_X(x)]$ is constant on $(0,\infty)$, say c, and one may solves $f_X(x) = ae^{cx^2/2}$ for $x\in\mathbb{R}$. Since f_X is a p.d.f., $c<0$. By writing $c = -1/\sigma^2$, we have

$$1 = \int_{-\infty}^{\infty} f(x)dx = a\int_{-\infty}^{\infty} e^{-x^2/(2\sigma^2)}dx = a\sigma\sqrt{2\pi},$$

which leads to $a = \frac{1}{\sqrt{2\pi}\sigma}$. Thus, X is a normal random variable with mean 0 and variance σ^2. Similarly, one can show that Y is normal with parameters $(0,\overline{\sigma}^2)$. In fact, the spherical symmetry of f implies $\sigma = \overline{\sigma}$.

Definition 6.5. $X_1,...,X_n$ are **independent** if, for any (Borel) subsets $A_1,...,A_n$) of $\mathbb{R}$,

$$P(X_1\in A_1,...,X_n\in A_n) = P(X_1\in A_1)\cdots P(X_n\in A_n).$$

Remark 6.7. With a rigorous mathematical proof, one can show that $X_1,...,X_n$ are independent if and only if, for any $a_1,...,a_n$,

$$P(X_1\le a_1,...,X_n\le a_n) = P(X_1\le a_1)\cdots P(X_n\le a_n).$$

Remark 6.8. Using the identity,

$$P(X_1 \le a_1, ..., X_n \le a_n) = P(X_1 \le a_1)P(X_2 \le a_2 | X_1 \le a_1) \times \cdots \\ \times P(X_n \le a_n | X_1 \le a_1, ..., X_{n-1} \le a_{n-1}),$$

one may conclude $X_1, ..., X_n$ are independent if and only if

$$\begin{aligned} &X_2 \text{ is independent of } X_1 \\ &X_3 \text{ is independent of } X_1, X_2 \\ &\qquad\vdots \\ &X_n \text{ is independent of } X_1, ..., X_{n-1}. \end{aligned}$$

Remark 6.9. Similar to Theorem 6.1, if $X_1, ..., X_n$ are jointly continuous (discrete) with joint p.d.f. (p.m.f.) f, then $X_1, ..., X_n$ are independent if and only if $f(x_1, ..., x_n) = f_1(x_1) \cdots f_n(x_n)$ for all $x_1, ..., x_n$.

6.3. Sums of independent random variables.

Proposition 6.2. *Let X, Y be independent random variables. If X, Y are jointly continuous with marginal densities f_X, f_Y, then $X + Y$ is continuous and*

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy, \quad F_{X+Y}(a) = \int_{-\infty}^{\infty} F_X(a-y) f_Y(y) dy.$$

Proof. Let f be the joint density of X, Y. Since X, Y are independent, $f(x,y) = f_X(x) f_Y(y)$. For $a \in \mathbb{R}$, one has

$$\begin{aligned} F_{X+Y}(a) = P(X + Y \le a) &= \iint_{\{(x,y): x+y \le a\}} f(x,y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) f_Y(y) dx dy = \int_{-\infty}^{\infty} F_X(a-y) f_Y(y) dy. \end{aligned}$$

To see that $X + Y$ is continuous, observe that the change of variable yields

$$\int_{-\infty}^{a-y} f(x) dx = \int_{-\infty}^{a} f(x-y) dx.$$

By Fubini's theorem, we obtain

$$F_{X+Y}(a) = \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) f_Y(y) dx dy = \int_{-\infty}^{a} \left(\int_{-\infty}^{\infty} f_X(x-y) f_Y(y) dy \right) dx.$$

This implies $X + Y$ is continuous with p.d.f. $\int_{-\infty}^{\infty} f_X(x-y) f_Y(y) dy$. □

Remark 6.10. Let X, Y be discrete random variables and p_X, p_Y, p_{X+Y} be respectively the p.m.f.s of $X, Y, X + Y$. Then, for $a \in \mathbb{R}$,

$$\begin{aligned} p_{X+Y}(a) = P(X + Y = a) &= \sum_{x: p_X(x) > 0} P(X = x, Y = a - x) \\ &= \sum_{x: p_X(x) > 0} p_X(x) p_Y(a - x) = \sum_{y: p_Y(y) > 0} p_X(a - y) p_Y(y). \end{aligned}$$

Example 6.6. Let X, Y be independent binomial random variables with parameters (n, p) and (m, p). Then, for $0 \le k \le n+m$,

$$\begin{aligned} p_{X+Y}(k) &= \sum_{i=0}^{n} p_X(i) p_Y(k-i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \binom{m}{k-i} p^{k-i} (1-p)^{m-(k-i)} \\ &= p^k (1-p)^{n+m-k} \sum_{i=0}^{n} \binom{n}{i}\binom{m}{k-i} = \binom{n+m}{k} p^k (1-p)^{n+m-k}. \end{aligned}$$

This implies that $X+Y$ is a binomial random variable with parameters $(n+m, p)$.

Example 6.7. Let X, Y be independent Poisson random variables with parameters λ, δ. Then, for $k \ge 0$,

$$\begin{aligned} p_{X+Y}(k) &= \sum_{i=0}^{k} p_X(i) p_Y(k-i) = \sum_{i=0}^{k} e^{-\lambda} \frac{\lambda^i}{i!} e^{-\delta} \frac{\delta^{k-i}}{(k-i)!} \\ &= \frac{e^{-(\lambda+\delta)}}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda^i \delta^{k-i} = e^{-(\lambda+\delta)} \frac{(\lambda+\delta)^k}{k!}. \end{aligned}$$

This implies that $X+Y$ is Poisson with parameter $\lambda+\delta$.

Example 6.8. Let X, Y be independent geometric random variables with parameter p. Then, for $k \ge 2$,

$$p_{X+Y}(k) = \sum_{i=1}^{k-1} p(1-p)^{i-1} p(1-p)^{k-i-1} = (k-1)p^2(1-p)^{k-2}.$$

Inductively, if $X_1, ..., X_r$ are independent geometric random variables with parameter p and $X = X_1 + \cdots + X_r$, then, for $k \ge r$,

$$p_X(k) = \sum_{\substack{i_1+\cdots+i_r=k \\ i_1\ge 1,\ldots,i_r\ge 1}} p^r (1-p)^{k-r} = \binom{k-1}{r-1} p^r (1-p)^{k-r}.$$

Clearly, one can see that X is a negative binomial random variable with parameters (r, p).

Example 6.9. Let $X_1, ..., X_n$ be independent geometric random variables with parameters $p_1, ..., p_n$. Assume that $p_i \ne p_j$ if $i \ne j$. Set $S_i = X_1 + \cdots + X_i$ and $q_i = 1 - p_i$. Then, for $k \ge n = 2$,

$$\begin{aligned} P(S_2 = k) &= \sum_{i=1}^{k-1} p_1 q_1^{i-1} p_2 q_2^{k-i-1} = p_1 p_2 q_2^{k-2} \sum_{i=1}^{k-1} (q_1/q_2)^{i-1} \\ &= p_1 p_2 q_2^{k-2} \frac{(q_1/q_2)^{k-1} - 1}{q_1/q_2 - 1} = p_1 p_2 \frac{q_1^{k-1} - q_2^{k-1}}{q_1 - q_2} \\ &= p_1 q_1^{k-1} \frac{p_2}{p_2 - p_1} + p_2 q_2^{k-1} \frac{p_1}{p_1 - p_2}. \end{aligned}$$

For $k \geq n = 3$, one has

$$\begin{aligned}
P(S_3 = k) &= \sum_{i=1}^{k-1} P(S_2 = i, X_3 = k-i) \\
&= \sum_{i=1}^{k-1} \left(p_1 q_1^{i-1} \frac{p_2}{p_2 - p_1} + p_2 q_2^{i-1} \frac{p_1}{p_1 - p_2} \right) p_3 q_3^{k-i-1} \\
&= \left(\sum_{i=1}^{k-1} p_1 q_1^{i-1} p_3 q_3^{k-i-1} \right) \frac{p_2}{p_2 - p_1} + \left(\sum_{i=1}^{k-1} p_2 q_2^{i-1} p_3 q_3^{k-i-1} \right) \frac{p_1}{p_1 - p_2} \\
&= p_1 q_1^{k-1} \frac{p_2 p_3}{(p_2 - p_1)(p_3 - q_1)} + p_3 q_3^{k-1} \frac{p_1 p_2}{(p_1 - p_3)(p_2 - p_1)} \\
&\quad + p_3 q_3^{k-1} \frac{p_2 p_1}{(p_2 - p_3)(p_1 - q_2)} + p_2 q_2^{k-1} \frac{p_3 p_1}{(p_3 - p_2)(p_1 - p_2)} \\
&= p_1 q_1^{k-1} \frac{p_2 p_3}{(p_2 - p_1)(p_3 - q_1)} + p_2 q_2^{k-1} \frac{p_3 p_1}{(p_3 - p_2)(p_1 - p_2)} \\
&\quad + p_3 q_3^{k-1} \frac{p_1 p_2}{(p_1 - p_3)(p_2 - p_3)}
\end{aligned}$$

By induction, we obtain the following proposition.

Proposition 6.3. *Let $X_1, ..., X_n$ be independent geometric variable with parameters $p_1, ..., p_n$. Suppose that $p_i \neq p_i$ if $i \neq j$. Then, for $k \geq n$,*

$$P(X_1 + \cdots + X_n = k) = \sum_{i=1}^{n} p_i (1 - p_i)^{k-1} \prod_{j \neq i} \frac{p_j}{p_j - p_i}.$$

Proof. We prove this theorem by induction. Suppose that the identity holds for $n-1$. Set $S_m = X_1 + \cdots + X_m$. By the independence of S_{n-1} and X_n, one has that, for $k \geq n$,

$$P(X_n = k) = P(S_n = k, X_n = 1) + P(S_n = k, X_n > 1) = I + II,$$

where

$$I = P(S_{n-1} = k-1) P(X_n = 1) = \left(\sum_{i=1}^{n-1} p_i q_i^{k-2} \prod_{j \neq i, i < n} \frac{p_j}{p_j - p_i} \right) p_n$$

and

$$II = P(S_n = k, X_n > 1) = \sum_{i=1}^{k-1} P(S_{n-1} = k-i) P(X_n = i | X_n > 1) P(X_n > 1).$$

Note that

$$\begin{aligned}
P(X_n = i | X_n > 1) &= P(X_n > i | X_n > 1) - P(X_n > i+1 | X_n > 1) \\
&= P(X_n > i-1) - P(X_n > i) = P(X_n = i-1).
\end{aligned}$$

This implies

$$\begin{aligned}
II &= \sum_{i=1}^{k-1} P(S_{n-1} = k-i) P(X_n = i-1) P(X_n > 1) = P(S_n = k-1) q_n \\
&= \left(\sum_{i=1}^{n} p_i q_i^{k-2} \prod_{j \neq i, j \leq n} \frac{p_j}{p_j - p_i} \right) q_n.
\end{aligned}$$

Thus,

$$\begin{aligned}
P(S_n = k) &= \left(\sum_{i=1}^{n-1} p_i q_i^{k-2} \prod_{j\neq i, i<n} \frac{p_j}{p_j - p_i}\right) p_n + \left(\sum_{i=1}^{n} p_i q_i^{k-2} \prod_{j\neq i, j\leq n} \frac{p_j}{p_j - p_i}\right) q_n \\
&= \sum_{i=1}^{n-1} p_i q_i^{k-2}(p_n - p_i) \prod_{j\neq i, j\leq n} \frac{p_j}{p_j - p_i} + \sum_{i=1}^{n} p_i q_i^{k-2} q_n \prod_{j\neq i, j\leq n} \frac{p_j}{p_j - p_i} \\
&= \sum_{i=1}^{n-1} p_i q_i^{k-2}(p_n - p_i + q_n) \prod_{j\neq i, j\leq n} \frac{p_j}{p_j - p_i} + p_n q_n^{k-1} \prod_{j<n} \frac{p_j}{p_j - p_n} \\
&= \sum_{i=1}^{n} p_i q_i^{k-1} \prod_{j\neq i, j\leq n} \frac{p_j}{p_j - p_i}.
\end{aligned}$$

□

Example 6.10. Let X, Y be independent uniform random variables over $(0,1)$. Since X, Y are independent and continuous, $X+Y$ is continuous with p.d.f.

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy.$$

Clearly, $f_{X+Y}(a) = 0$ for $a \in (-\infty, 0) \cup (2, \infty)$. For $a \in (0,1)$,

$$f_{X+Y}(a) = \int_0^a dy = a$$

and, for $a \in [1,2)$,

$$f_{X+Y}(a) = \int_{a-1}^{1} dy = 2 - a.$$

Immediately, this implies

$$F_{X+Y}(a) = \int_0^a u du = a^2/2 \quad \forall a \in (0,1),$$

and

$$F_{X+Y}(a) = \int_0^1 u du + \int_1^a (2-u) du = -\frac{a^2}{2} + 2a - 1 \quad \forall a \in [1,2).$$

Example 6.11. Let X, Y be independent gamma random variables with parameters (α, λ) and (β, λ). Then, $X+Y$ is continuous with density function

$$\begin{aligned}
f_{X+Y}(a) &= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^a \lambda e^{-\lambda(a-y)} (\lambda(a-y))^{\alpha-1} \lambda e^{-\lambda y} (\lambda y)^{\beta-1} dy \\
&= \frac{\lambda e^{-\lambda a} (\lambda a)^{\alpha+\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^a \frac{1}{a} \left(1 - \frac{y}{a}\right)^{\alpha-1} \left(\frac{y}{a}\right)^{\beta-1} dy \\
&= \frac{\lambda e^{-\lambda a} (\lambda a)^{\alpha+\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 t^{\beta-1}(1-t)^{\alpha-1} = \frac{\lambda e^{-\lambda a} (\lambda a)^{\alpha+\beta-1} B(\beta, \alpha)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{1}{\Gamma(\alpha+\beta)} \lambda e^{-\lambda a} (\lambda a)^{\alpha+\beta-1}.
\end{aligned}$$

This shows that $X+Y$ is a gamma random variable with parameters $(\alpha+\beta, \lambda)$. In particular, if $X_1, ..., X_n$ are independent exponential random variables with parameter λ, then $X_1+\cdots+X_n$ has the gamma distribution with parameters (n, λ).

Proposition 6.4. *Let $X_1, ..., X_n$ be independent exponential random variables with parameters $\lambda_1, ..., \lambda_n$, where $\lambda_i \neq \lambda_j$ for $i \neq j$. Set $Y = X_1 + \cdots + X_n$. Then, Y is continuous and*

$$f_Y(y) = \sum_{i=1}^n \lambda_i e^{-\lambda_i y} \prod_{j\neq i} \frac{\lambda_j}{\lambda_j - \lambda_i}.$$

Proof. The proof is similar to the case of geometric random variables. To get the desired p.d.f., it suffices to prove

$$1 - F_Y(a) = \sum_{i=1}^n e^{-\lambda_i a} \prod_{j\neq i} \frac{\lambda_j}{\lambda_j - \lambda_i} \quad \forall a > 0. \tag{6.1}$$

We prove the above identity by induction. For $i \geq 1$, set $S_i = X_1 + \cdots + X_i$. Assume that (6.1) holds for $Y = S_{n-1}$. For $a > 0$,

$$\begin{aligned} 1 - F_{S_n}(a) = P(S_n > a) &= \iint_{x+y>a} f_{S_{n-1}}(x) f_{X_n}(y)dxdy \\ &= \sum_{i=1}^{n-1} \left(\prod_{j\neq i, j<n} \frac{\lambda_j}{\lambda_j - \lambda_i} \right) \iint_{\substack{x>0,y>0 \\ x+y>a}} \lambda_i \lambda_n e^{-\lambda_i x} e^{-\lambda_n y} dxdy. \end{aligned}$$

Note that

$$\begin{aligned} &\iint_{\substack{x>0,y>0 \\ x+y>a}} \lambda_i \lambda_n e^{-\lambda_i x} e^{-\lambda_n y} dxdy \\ &\qquad = \int_0^a \int_{a-y}^\infty \lambda_i \lambda_n e^{-\lambda_i x} e^{-\lambda_n y} dxdy + \int_a^\infty \int_0^\infty \lambda_i \lambda_n e^{-\lambda_i x} e^{-\lambda_n y} dxdy \\ &\qquad = \frac{\lambda_n}{\lambda_n - \lambda_i} e^{-\lambda_i a} + \frac{\lambda_i}{\lambda_i - \lambda_n} e^{-\lambda_n a} \end{aligned}$$

As a result, this implies

$$1 - F_{S_n}(a) = \sum_{i=1}^{n-1} e^{-\lambda_i a} \prod_{j\neq i, j\leq n} \frac{\lambda_j}{\lambda_j - \lambda_i} + e^{-\lambda_n a} p(\lambda_n) \prod_{j=1}^{n-1} \frac{\lambda_j}{\lambda_j - \lambda_n},$$

where

$$p(t) = \sum_{i=1}^{n-1} \prod_{j\neq i, j<n} \frac{\lambda_j - t}{\lambda_j - \lambda_i}.$$

Observe that $p(t)$ is a polynomial of degree at most $n-2$. Since $p(\lambda_i) = 1$ for $i = 1, 2, ..., n-1$, $p(t) = 1$ for all t. Particularly, $p(\lambda_n) = 1$. □

Example 6.12. Let X, Y be normal with parameters $(0, 1)$ and $(0, \sigma^2)$. Suppose that X, Y are independent. Then,

$$f_{X+Y}(a) = \frac{1}{\sqrt{2\pi}\sqrt{2\pi\sigma^2}} \int_{-\infty}^\infty e^{-(a-y)^2/2} e^{-y^2/(2\sigma^2)} dy.$$

Note that

$$\frac{(a-y)^2}{2} + \frac{y^2}{2\sigma^2} = \frac{1}{2\sigma^2/(\sigma^2+1)} \left(y - \frac{a}{(\sigma^2+1)/\sigma^2} \right)^2 + \frac{a^2}{2(\sigma^2+1)}.$$

This implies

$$f_{X+Y}(a) = \frac{\sqrt{2\pi\sigma^2/(\sigma^2+1)}}{\sqrt{(2\pi)^2\sigma^2}} e^{-a^2/(2(\sigma^2+1))} = \frac{1}{\sqrt{2\pi(1+\sigma^2)}} e^{-a^2/(2(1+\sigma^2))},$$

which means that $X+Y$ is normal with parameters $(0, 1+\sigma^2)$.

For the general case, let X, Y are independent normal random variables with parameters (μ_1, σ_1) and (μ_2, σ_2). By writing $X = \mu_1 + \sigma_1 X'$ and $Y = \mu_2 + \sigma_2 Y'$, where X', Y' are independent standard normal random variables, one has

$$X + Y = \mu_1 + \mu_2 + \sigma_1(X' + (\sigma_2/\sigma_1)Y').$$

This implies that $X+Y$ is a normal random variable with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2[1 + (\sigma_2/\sigma_1)^2] = \sigma_1^2 + \sigma_2^2$. Inductively, if $X_1, ..., X_n$ are independent and X_i is a normal random variable with parameters (μ_i, σ_i^2), then $X_1 + \cdots + X_n$ is a normal random variable with mean $\mu_1 + \cdots + \mu_n$ and variance $\sigma_1^2 + \cdots + \sigma_n^2$.

6.4. **Conditional distributions.** Recall the definition of the condition probability.

$$P(F|E) = \frac{P(EF)}{P(E)} \quad \text{given } P(E) > 0.$$

Definition 6.6. Let X, Y be discrete random variables with probability mass functions p_X, p_Y and joint probability mass function p. Suppose $P(Y = y) > 0$.

(1) The **conditional probability mass function** of X given $Y = y$ is defined by

$$p_{X|Y}(x|y) = P(X = x|Y = y)\frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p(x, y)}{p_Y(y)}.$$

(2) The **conditional probability distribution function** of X given $Y = y$ is defined by

$$F_{X|Y}(a|y) = P(X \le a|Y = y) = \sum_{x \le a} p_{X|Y}(x|y).$$

Remark 6.11. Note that a conditional p.m.f. is a p.m.f. and a conditional distribution function is exactly a distribution function. That is, the mapping $x \mapsto p_{X|Y}(x, y)$ is nonzero on a countable set satisfying $\sum_{x: p_{X|Y}(x|y) > 0} p_{X|Y}(x|y) = 1$, and $x \mapsto F_{X|Y}(x|y)$ is non-decreasing, right-continuous and satisfies

$$\lim_{x \to -\infty} F_{X|Y}(x|y) = 0, \quad \lim_{x \to \infty} F_{X|Y}(x|y) = 1.$$

Remark 6.12. If X, Y are independent discrete random variables, then

$$p_{X|Y}(x|y) = \frac{p(x, y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x).$$

Example 6.13. Let X, Y be independent Poisson random variables with parameters λ, δ. The conditional p.m.f. of X given $X + Y = n$ is

$$p_{X|X+Y}(k|n) = P(X = k|X + Y = n) = \frac{P(X = k, X + Y = n)}{P(X + Y = n)} \quad \forall 0 \le k \le n.$$

It has been shown before that $X+Y$ is Poisson with parameter $\lambda+\delta$. This implies $P(X+Y = n) = e^{-\lambda-\delta}(\lambda + \delta)^n/n!$. By the independence of X, Y,

$$P(X = k, X+Y = n) = P(X = k, Y = n-k) = P(X = k)P(Y = n-k) = e^{-\lambda}\frac{\lambda^k}{k!}e^{-\delta}\frac{\delta^{n-k}}{(n-k)!}.$$

This implies

$$p_{X|X+Y}(k|n) = \binom{n}{k}\left(\frac{\lambda}{\lambda+\delta}\right)^k\left(\frac{\delta}{\lambda+\delta}\right)^{n-k} \quad \forall 0 \le k \le n.$$

This implies that, given $X + Y = n$, X is a binomial random variable with parameters $(n, \lambda/(\lambda+\delta))$.

Definition 6.7. Let X, Y be jointly continuous with joint p.d.f. f and marginal p.d.f.s f_X, f_Y. Suppose $f_Y(y) > 0$.

(1) The **conditional probability density function** of X given $Y = y$ is defined by

$$f_{X|Y}(x|y) := \frac{f(x,y)}{f_Y(y)}.$$

(2) The **conditional distribution function** of X given $Y = y$ is defined by

$$F_{X|Y}(a|y) := \int_{-\infty}^{a} f_{X|Y}(x|y)dx.$$

(3) The **conditional probability** of $\{X \in A\}$ given $Y = y$ is defined by

$$P(X \in A|Y = y) := \int_A f_{X|Y}(x|y)dx.$$

Remark 6.13. The conditional p.d.f. is a p.d.f.. That is, $f_{X|Y}(x|y) \ge 0$ and

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx = 1.$$

Remark 6.14. To see what the conditional p.d.f. means, we assume that $f(x,y)$ and $f_Y(y)$ are continuous. Then, for $\epsilon, \delta > 0$,

$$\frac{P(X \in (x, x+\epsilon), Y \in (y, y+\delta))}{\epsilon P(Y \in (y, y+\delta))} = \frac{f(x,y)\epsilon\delta(1+o_{\epsilon,\delta}(1))}{f_Y(y)\epsilon\delta(1+o_\epsilon(1))} \to \frac{f(x,y)}{f_Y(y)}$$

as $\epsilon, \delta \to 0$.

Remark 6.15. If X, Y are independent, then $f_{X|Y}(x|y) = f_X(x)$.

Example 6.14. Let X, Y be jointly continuous with p.d.f.

$$f(x,y) = \begin{cases} e^{-x/y}e^{-y}/y & \text{if } x > 0, y > 0 \\ 0 & \text{o.w.} \end{cases}.$$

This implies $f_Y(y) = 0$ for $y \le 0$ and

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^{\infty} \frac{e^{-x/y}e^{-y}}{y}dx = e^{-y},$$

for $y > 0$. Hence, we have

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{e^{-x/y}}{y} \quad \forall x > 0, y > 0.$$

This proves that, given $Y = y$, X is an exponential random variable with parameter y.

Example 6.15. Consider a bag containing infinitely many coins of which probability landing on heads is uniform over $(0,1)$. Select a coin and then flip it independently for $n+m$ times. Find the conditional distribution of the probability that a coin lands on heads given that $n+m$ flips results in n heads.

First, observe that if X, Y are jointly continuous with density function f, then

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{f_X(x)}{f_Y(y)} \times \frac{f(x,y)}{f_X(x)} = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}.$$

We may follow the above definition to define the conditional p.d.f. or p.m.f. given the observation of a discrete random variable or a continuous random variable respectively. For instance, let X denote the probability of heads of the selected coin and Y be the number of heads in the $n+m$ flips. Then, the setting of this example says that, given $X = p$, Y is a binomial random variable with parameters $(n+m, p)$. Then, the conditional p.d.f. of X given $Y = n$ is

$$f_{X|Y}(p|n) = \frac{p_{Y|X}(n|p) f_X(p)}{p_Y(n)} = \frac{\binom{n+m}{n}}{P(Y=n)} p^n (1-p)^m, \quad \forall p \in (0,1).$$

Since $f_{X|Y}(p|n)$ is a density function, we have

$$1 = \int_0^1 f_{X|Y}(p|n) dp = \frac{\binom{n+m}{n}}{P(Y=n)} \int_0^1 p^n (1-p)^m dp = \frac{\binom{n+m}{n} B(n+1, m+1)}{P(Y=n)}.$$

This implies

$$P(Y=n) = \binom{n+m}{n} B(n+1, m+1), \quad f_{X|Y}(p|n) = \frac{1}{B(n+1, m+1)} p^n (1-p)^m,$$

where $B(a,b)$ is the coefficient in the p.d.f. of a beta random variable. As a consequence, given $Y = n$, X has is a beta random variable with parameters $(n+1, m+1)$.

6.5. **Order statistics.**

Definition 6.8. Random variables, $X_1, ..., X_n$, are called **identically distributed** if their distribution functions are equal and called **i.i.d.** if they are independent and identically distributed.

Let $X_1, ..., X_n$ be i.i.d. random variables. Define $X_{(j)}$ be the jth smallest of $X_1, ..., X_n$ for $1 \le j \le n$. Mathematically, one has

$$X_{(j)} = \min_{k_1 < k_2 < \cdots < k_j} \max\{X_{k_1}, ..., X_{k_j}\}.$$

Here, $X_{(1)}, ..., X_{(n)}$ is called the **order statistics** corresponding to $X_1, ..., X_n$.

Suppose that X_1 is continuous and has density function f. Then, $X_1, ..., X_n$ are jointly continuous with density function $g(x_1, ..., x_n) = f(x_1) \cdots f(x_n)$. This implies

$$\begin{aligned} &P(X_{(i)} = X_{(i+1)} \text{ for some } i) = P(X_i = X_j \text{ for some } i < j) \\ \le &\binom{n}{2} P(X_1 = X_2) = \binom{n}{2} \iint_{x=y} f(x) f(y) dx dy = 0. \end{aligned}$$

This implies

$$P(X_{(1)} < X_{(2)} < \cdots < X_{(n)}) = 1.$$

We now turn to determine the joint density of $X_{(i)}$. Assume that f is continuous. Let $x_1 < x_2 < \cdots < x_n$ and choose $\epsilon > 0$ such that $x_i - x_{i-1} > \epsilon$ for all i. Then,

$$\begin{aligned}
&\lim_{\epsilon\to 0^+} \frac{P(X_{(i)} \in (x_i - \epsilon/2, x_i + \epsilon/2), \forall i)}{\epsilon^n} \\
=&\lim_{\epsilon\to 0^+} \frac{n!P(X_i \in (x_i - \epsilon/2, x_i + \epsilon/2))}{\epsilon^n} \\
=&n!\prod_{i=1}^n \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\int_{x_i-\epsilon/2}^{x_i+\epsilon/2} f(x)dx = n!\prod_{i=1}^n f(x_i).
\end{aligned}$$

This gives a candidate for the joint density function of the order statistics.

Proposition 6.5. *Let $X_1, ..., X_n$ be i.i.d. and $X_{(1)}, ..., X_{(n)}$ be their corresponding order statistics. Suppose that X_i is continuous with common density function f. Then,*

$$f_{X_{(1)},...,X_{(n)}}(x_1, ..., x_n) = \begin{cases} n!f(x_1)\cdots f(x_n) & \text{if } x_1 < \cdots < x_n \\ 0 & \text{otherwise} \end{cases}.$$

Proof. We prove this proposition by induction. The case $n = 1$ is obvious. Suppose that the proposition holds for $n - 1$. Let g be the function at the right side of the above identity and $a_1, a_2, ..., a_n \in \mathbb{R}$. Note that

$$P(X_{(i)} \le a_i\, \forall 1 \le i \le n) = \sum_{j=1}^n P(X_j = X_{(n)} \le a_n, X_{(i)} \le a_i\, \forall 1 \le i < n).$$

Let $Y_{(1)}, ..., Y_{(n-1)}$ be the order statistics of $X_1, ..., X_{j-1}, X_{j+1}, ..., X_n$. Then, one has

$$\begin{aligned}
&P(X_j = X_{(n)} \le a_n, X_{(i)} \le a_i\ \forall 1 \le i < n) \\
=&P(Y_{(n-1)} < X_j \le a_n, Y_{(i)} \le a_i\ \forall 1 \le i < n) \\
=&\int\cdots\int_{\substack{y_i < a_i, 1 \le i \le n \\ y_{n-1} < y_n}} f_{Y_{(1)},...,Y_{(n-1)}}(y_1, ..., y_{n-1}) f(y_n) dy_1\cdots dy_n \\
=&\int\cdots\int_{\substack{y_i \le a_i, 1 \le i \le n \\ y_1 < \cdots < y_n}} (n-1)!f(y_1)\cdots f(y_n)dy_1\cdots dy_n \\
=&\frac{1}{n}\int\cdots\int_{y_i \le a_i, 1 \le i \le n} g(y_1, ..., y_n)dy_1\cdots dy_n.
\end{aligned}$$

Note that the last term is independent of j. This implies

$$F_{X_{(1)},...,X_{(n)}}(a_1, ..., a_n) = \int\cdots\int_{y_i \le a_i, 1 \le i \le n} g(y_1, ..., y_n)dy_1\cdots dy_n.$$

This proves that g is the joint density of $X_{(1)}, ..., X_{(n)}$. $\square$

Corollary 6.6. *Let $X_1, ..., X_n$ be i.i.d. continuous random variables with common density function f. Then,*

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!}[F(x)]^{j-1}[1 - F(x)]^{n-j} f(x).$$

Proof. Using the above proposition, one has

$$\begin{aligned} F_{X_{(j)}}(a) &= n! \int_{-\infty}^{a} \left(\int \cdots \int_{y_1 < \cdots < y_j < \cdots < y_n} f(y_1) \cdots f(y_n) dy_1 \cdots dy_{j-1} dy_{j+1} \cdots dy_n \right) f(y_j) dy_j \\ &= n! \int_{-\infty}^{a} g(y_j) h(y_j) f(y_j) dy_j, \end{aligned}$$

where

$$g(x) = \int \cdots \int_{y_1 < \cdots < y_{j-1} < x} f(y_1) \cdots f(y_{j-1}) dy_1 \cdots dy_{j-1}$$

and

$$h(x) = \int \cdots \int_{x < y_{j+1} < \cdots < y_n} f(y_{j+1}) \cdots f(y_n) dy_{j+1} \cdots dy_n.$$

Clearly, $f_{X_{(j)}}(x) = f(x)g(x)h(x)$. By the continuity of distribution functions, one has

$$\begin{aligned} g(x) &= \int \cdots \int_{y_2 < \cdots < y_{j-1} < x} F(y_2) f(y_2) \cdots f(y_{j-1}) dy_2 \cdots dy_{j-1} \\ &= \int \cdots \int_{y_3 < \cdots < y_{j-1} < x} \frac{F^2(y_3)}{2} f(y_3) \cdots f(y_{j-1}) dy_3 \cdots dy_{j-1} \\ &= \cdots = \int_{-\infty}^{x} \frac{F^{j-2}(y_{j-1})}{(j-2)!} f(y_{j-1}) dy_{j-1} = \frac{F^{j-1}(x)}{(j-1)!} \end{aligned}$$

Similarly, $h(x) = [1 - F(x)]^{n-j}/(n-j)!$. □

Remark 6.16. One may prove inductively that

$$F_{X_{(j)}}(a) = \sum_{k=j}^{n} \binom{n}{k} [F(a)]^k [1 - F(a)]^{n-k}.$$

Example 6.16. Let $X_1, ..., X_n$ be i.i.d. uniform random variables over $(0, 1)$. Then,

$$f_{X_{(1)},...,X_{(n)}}(x_1, ..., x_n) = \begin{cases} n! & \text{if } x_1 < \cdots < x_n \\ 0 & \text{o.w.} \end{cases}$$

and

$$f_{X_{(j)}}(x) = \frac{n!}{(n-j)!(j-1)!} x^{j-1} (1-x)^{n-j} \quad \forall x \in (0, 1).$$

This implies that $X_{(j)}$ is a beta random variable with parameters $(n-j+1, j)$. In particular, if $X_1, ..., X_{2n+1}$ are i.i.d. uniform over $(0, 1)$, then the **median** $X_{(n+1)}$ is a beta random variable with parameters $(n + 1, n + 1)$.

Remark 6.17. Using the same reasoning as before, one can show that

$$\begin{aligned} f_{X_{(i)},X_{(j)}}(x_i, x_j) =& \frac{n!}{(i-1)!(j-i-1)!(n-j)!} [F(x_j)]^{i-1} [F(x_j) - F(x_i)]^{j-i-1} \\ & \times [1 - F(x_j)]^{n-j} f(x_i) f(x_j) \quad \forall x_i < x_j, \end{aligned}$$

and $f_{X_{(i)},X_{(j)}}(x_i, x_j) = 0$ if $x_i \geq x_j$.

Example 6.17. The **range** of i.i.d. random variables $X_1, ..., X_n$ is defined to be

$$R = X_{(n)} - X_{(1)}.$$

By the above remark, if f is the common density of X_i, then

$$f_{X_{(1)},X_{(n)}}(x,y) = \begin{cases} n(n-1)[F(y)-F(x)]^{n-2}f(x)f(y) & \text{if } x<y \\ 0 & \text{o.w.} \end{cases}.$$

This implies

$$\begin{aligned} P(R\le a) &= n(n-1)\int_{0<y-x\le a}[F(y)-F(x)]^{n-2}f(x)f(y)dxdy \\ &= n\int_{-\infty}^{\infty}\left((n-1)\int_x^{x+a}[F(y)-F(x)]^{n-2}f(y)dy\right)f(x)dx \\ &= n\int_{-\infty}^{\infty}\left([F(y)-F(x)]^{n-1}\Big|_{y=x}^{y=x+a}\right)f(x)dx \\ &= n\int_{-\infty}^{\infty}[F(x+a)-F(x)]^{n-1}f(x)dx. \end{aligned}$$

For the case that X_i is uniform over $(0,1)$ for $1\le i\le n$,

$$P(R\le a) = n\int_0^{1-a}a^{n-1}dx + n\int_{1-a}^1(1-x)^{n-1}dx = n(1-a)a^{n-1}+a^n \quad \forall a\in(0,1).$$

This implies

$$\frac{d}{da}P(R\le a) = \begin{cases} n(n-1)a^{n-2}(1-a) & \text{if } a\in(0,1) \\ 0 & \text{o.w.} \end{cases},$$

which says that R is a beta random variable with parameters $(n-1,2)$.

6.6. **Joint probability distribution of functions of random variables.** Let X_1, X_2 be jointly continuous random variables with joint p.d.f. f_{X_1,X_2} and $Y_i = g_i(X_1,X_2)$ with $i=1,2$. Let D be the support of f_{X_1,X_2}, i.e. $D=\{(x,y)|f_{X_1,X_2}(x,y)\neq 0\}$, and $E=(g_1,g_2)(D)$. Assume that $(g_1,g_2): D\to E$ is a bijection and write $(h_1,h_2)=(g_1,g_2)^{-1}$ Then, for $a_1,a_2\in\mathbb{R}$ and $R=(-\infty,a_1]\times(-\infty,a_2]$,

$$P(Y_1\le a_1, Y_2\le a_2) = \iint_{D'} f_{X_1,X_2}(x_1,x_2)dx_1dx_2,$$

where $D' = D\cap\{(x_1,x_2)|(g_1,g_2)(x_1,x_2)\in R\} = (h_1,h_2)(E\cap R)$. Assume further that g_1,g_2 have continuously differentiable partial derivatives with

$$J(x_1,x_2) = \det\begin{pmatrix} (g_1)_{x_1} & (g_1)_{x_2} \\ (g_2)_{x_1} & (g_2)_{x_2} \end{pmatrix} = (g_1)_{x_1}(g_2)_{x_2} - (g_1)_{x_2}(g_2)_{x_2} \neq 0 \quad \text{on } D.$$

By the inverse function theorem, one has

$$A(y_1,y_2) := \det\begin{pmatrix} (h_1)_{y_1} & (h_1)_{y_2} \\ (h_2)_{y_1} & (h_2)_{y_2} \end{pmatrix} = J(h_1(y_1,y_2),h_2(y_1,y_2))^{-1},$$

and then

$$\iint_{D'} f_{X_1,X_2}(x_1,x_2)dx_1dx_2 = \iint_{E\cap R} f_{X_1,X_2}(h_1(y_1,y_2),h_2(y_1,y_2))|A(y_1,y_2)|dy_1dy_2.$$

This implies that

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} f_{X_1,X_2}(h_1(y_1,y_2),h_2(y_1,y_2))J(h_1(y_1,y_2),h_2(y_1,y_2))^{-1} & \text{on } E \\ 0 & \text{otherwise.} \end{cases}$$

Example 6.18. Let X_1, X_2 be independent exponential random variables with parameters λ_1, λ_2. Set $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Then, $Y_i = g_1(X_1, X_2)$ for $i = 1, 2$, where $g_1(x_1, x_2) = x_1 + x_2$ and $g_2(x_1, x_2) = x_1 - x_2$. Clearly, (g_1, g_2) is a bijection from $(0, \infty) \times (0, \infty)$ onto $R = \{(y_1, y_2) | y_1 > |y_2|\}$ and the inverse function of (g_1, g_2), denoted by (h_1, h_2) satisfies $h_1(y_1, y_2) = (y_1 + y_2)/2$ and $h_2(y_1, y_2) = (y_1 - y_2)/2$. As $f_{X_1,X_2}(x_1, x_2) = \lambda_1 \lambda_2 e^{-\lambda_1 x_1 - \lambda_2 x_2}$, one has

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \frac{\lambda_1 \lambda_2}{2} \exp\left\{-\left(\frac{\lambda_1 + \lambda_2}{2}\right) y_1 - \left(\frac{\lambda_1 - \lambda_2}{2}\right) y_2\right\} & \text{for } y_1 > |y_2| \\ 0 & \text{otherwise.} \end{cases}$$

It's worthwhile to note that Y_1, Y_2 are not independent.