Note 1.267. A 1-dimensional random vector, by definition, is exactly an RV. A p-dimensional random vector is made up of p components, each of which are RVs. Keeping this connection in mind, we repeat the steps of our analysis as done for RVs.

Notation 1.268 (Pre-image of a set under an $\mathbb{R}^p$-valued function)**.** Let Ω be a non-empty set and let $X : \Omega \to \mathbb{R}^p$ be a function. Given any subset A of $\mathbb{R}^p$, we consider the subset $X^{-1}(A)$ of Ω defined by

$$X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\}.$$

The set $X^{-1}(A)$ shall be referred to as the pre-image of A under the function X. We shall suppress the symbols ω and use the following notation for convenience, viz.

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} = (X \in A).$$

Notation 1.269. As discussed for RVs, we now consider the following set function in relation to a given p-dimensional random vector. Given a random vector defined on $(\Omega, \mathcal{F}, \mathbb{P})$, consider the set function $\mathbb{P}_X(A) := \mathbb{P}(X^{-1}(A)) = \mathbb{P}(X \in A)$ for all subsets A of $\mathbb{R}^p$. We shall write $\mathbb{B}_p$ to denote the power set of $\mathbb{R}^p$.

Following arguments similar to Proposition 1.103, we get the next result. The proof is skipped for brevity.

Proposition 1.270. *Let X be a p-dimensional random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, the set function $\mathbb{P}_X$ is a probability function/measure defined on the collection $\mathbb{B}_p$, i.e. $(\mathbb{R}^p, \mathbb{B}_p, \mathbb{P}_X)$ is a probability space.*

Definition 1.271 (Induced Probability Space and Induced Probability Measure)**.** If X is a p-dimensional random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then the probability function/measure $\mathbb{P}_X$ on $\mathbb{B}_p$ is referred to as the induced probability function/measure induced by X. In this case, $(\mathbb{R}^p, \mathbb{B}_p, \mathbb{P}_X)$ is referred to as the induced probability space induced by X.

Notation 1.272. We shall call $\mathbb{P}_X$ as the joint law or joint distribution of the random vector X.

We have found that the DF of an RV identifies the law/distribution of the RV. Motivated by this fact, we now consider a similar function for random vectors.

Definition 1.273 (Joint Distribution function (Joint DF) and Marginal Distribution function (Marginal DF))**.** Let $X = (X_1, X_2, \cdots, X_p) : \Omega \to \mathbb{R}^p$ be a p-dimensional random vector.

(a) The joint DF of X is a function $F_X : \mathbb{R}^p \to [0, 1]$ defined by

$$\begin{aligned} F_X(x_1, x_2, \cdots, x_p) &:= \mathbb{P}_X((-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_p]) \\ &= \mathbb{P}(X \in \prod_{j=1}^{p} (-\infty, x_j]) \\ &= \mathbb{P}((X_1, X_2, \cdots, X_p) \in \prod_{j=1}^{p} (-\infty, x_j]) \\ &= \mathbb{P}(X_1 \le x_1, X_2 \le x_2, \cdots, X_p \le x_p), \forall x = (x_1, x_2, \cdots, x_p) \in \mathbb{R}^p. \end{aligned}$$

(b) The joint DF of any subset of the RVs $X_1, X_2, \cdots, X_p$ is called a marginal DF of the random vector X.

Note 1.274. Let $X = (X_1, X_2, X_3) : \Omega \to \mathbb{R}^3$ be a 3-dimensional random vector. Then the DF F_{X_2} of X_2 and the joint DF F_{X_1,X_3} of X_1 & X_3 are marginal DFs of the random vector X.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Unless stated otherwise, RVs and random vectors shall be defined on this probability space.

Note 1.275. Recall that for an RV Y, we have $F_Y(b) - F_Y(a) = \mathbb{P}(a < Y \le b) \ge 0$ for all $a, b \in \mathbb{R}$ with $a < b$.

Proposition 1.276. *Let $X = (X_1, X_2) : \Omega \to \mathbb{R}^2$ be a 2-dimensional random vector. Let $a_1 < b_1, a_2 < b_2$. Then,*

$$\begin{aligned} F_X(b_1, b_2) - F_X(a_1, b_2) - F_X(b_1, a_2) + F_X(a_1, a_2) &= \mathbb{P}(X \in (a_1, b_1] \times (a_2, b_2]) \\ &= \mathbb{P}(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2) \\ &\ge 0. \end{aligned}$$

Proof. Consider the events $A_1 := (X_1 \leq a_1, X_2 \leq b_2)$ and $A_2 := (X_1 \leq b_1, X_2 \leq a_2)$. Note that

$$(X_1 \leq b_1, X_2 \leq b_2) \cap (a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2)^c = A_1 \cup A_2.$$

Now, $(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) \subseteq (X_1 \leq b_1, X_2 \leq b_2)$ and hence

$$\begin{aligned}
&\mathbb{P}((X_1 \leq b_1, X_2 \leq b_2) \cap (a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2)^c) \\
&= \mathbb{P}(X_1 \leq b_1, X_2 \leq b_2) - \mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) \\
&= F_X(b_1, b_2) - \mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2).
\end{aligned}$$

By the inclusion-exclusion principle (see Proposition 1.61)

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) = F_X(a_1, b_2) + F_X(b_1, a_2) - F_X(a_1, a_2).$$

The result follows. $\square$

For higher dimensions, the above result has an appropriate extension. To state this, we first need some notations.

Notation 1.277. Let $\prod_{j=1}^p (a_j, b_j]$ be a rectangle in $\mathbb{R}^p$. Observe that the co-ordinates of the vertices are made up of either a_j or b_j for each $j = 1, 2, \cdots, p$. Let Δ_k^p denote the set of vertices where exactly k many a_j's appear. Then the complete set of vertices is $\cup_{k=0}^p \Delta_k^p$. For example,

$$\Delta_0^2 = \{(b_1, b_2)\}, \quad \Delta_1^2 = \{(a_1, b_2), (b_1, a_2)\}, \quad \Delta_2^2 = \{(a_1, a_2)\}.$$

Proposition 1.276 can now be generalized to higher dimensions as follows. We skip the details of the proof for brevity.

Proposition 1.278. *Let $X = (X_1, X_2, \cdots, X_p) : \Omega \to \mathbb{R}^p$ be a p-dimensional random vector. Let $a_1 < b_1, a_2 < b_2, \cdots, a_p < b_p$. Then,*

$$\mathbb{P}(X \in \prod_{j=1}^p (a_j, b_j]) = \sum_{k=0}^p (-1)^k \sum_{x \in \Delta_k^p} F_X(x) = \mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, \ldots, a_p < X_p \leq b_p) \geq 0.$$

Proposition 1.279 (Computation of Marginal DFs from Joint DF)**.** *Let* $X = (X_1, X_2, \cdots, X_p) : \Omega \to \mathbb{R}^p$ *be a* p*-dimensional random vector. Fix* $1 \leq j \leq p$*. Then, for all* $x \in \mathbb{R}$ *we have*

$$\begin{aligned}
F_{X_j}(x) &= \lim_{\substack{t_k \to \infty \\ k \in \{1, \cdots, j-1, j+1, \cdots, p\}}} F_X(t_1, \cdots, t_{j-1}, x, t_{j+1}, \cdots, t_p) \\
&= \lim_{t \to \infty} F_X(\underbrace{t, \cdots, t}_{j-1 \ times}, x, \underbrace{t, \cdots, t}_{p-j \ times}) \\
&=: F_X(\underbrace{\infty, \cdots, \infty}_{j-1 \ times}, x, \underbrace{\infty, \cdots, \infty}_{p-j \ times}).
\end{aligned}$$

Proof. As in the proof of Theorem 1.115, using Proposition 1.110, we have

$$\begin{aligned}
&\lim_{\substack{t_k \to \infty \\ k \in \{1, \cdots, j-1, j+1, \cdots, p\}}} F_X(t_1, \cdots, t_{j-1}, x, t_{j+1}, \cdots, t_p) \\
&= \lim_{\substack{t_k \to \infty \\ k \in \{1, \cdots, j-1, j+1, \cdots, p\}}} \mathbb{P}_X((-\infty, t_1] \times \cdots (-\infty, t_{j-1}] \times (-\infty, x] \times (-\infty, t_{j-1}] \times \cdots \times (-\infty, t_p]) \\
&= \lim_{n \to \infty} \mathbb{P}_X((-\infty, n] \times \cdots (-\infty, n] \times (-\infty, x] \times (-\infty, n] \times \cdots \times (-\infty, n]) \\
&= \mathbb{P}_X(\mathbb{R} \times \cdots \times \mathbb{R} \times (-\infty, x] \times \mathbb{R} \times \cdots \times \mathbb{R}) \\
&= \mathbb{P}(X_1 \in \mathbb{R}, \cdots, X_{j-1} \in \mathbb{R}, X_j \in (-\infty, x], X_{j+1} \in \mathbb{R}, \cdots, X_p \in \mathbb{R}) \\
&= \mathbb{P}(X_j \in (-\infty, x]) = F_{X_j}(x).
\end{aligned}$$

This completes the proof. □

Remark 1.280. Using Proposition 1.279, we can compute the DFs of each component RVs from the joint DF of a random vector. More generally, the higher dimensional marginal DFs can be computed from the joint DF in a similar manner. For example, if $X = (X_1, X_2, \cdots, X_p)$ is a p-dimensional random vector, then

$$F_{X_1, X_2}(x_1, x_2) = \lim_{t \to \infty} F_X(x_1, x_2, \underbrace{t, \cdots, t}_{p-2 \text{ times}}) =: F_X(x_1, x_2, \underbrace{\infty, \cdots, \infty}_{p-2 \text{ times}}).$$

The joint DF of a random vector has properties similar to the DF of an RV. Compare the next result with Theorem 1.115.

Theorem 1.281. *Let $X = (X_1, X_2, \cdots, X_p) : \Omega \to \mathbb{R}^p$ be a p-dimensional random vector with joint DF F_X. Then,*

(a) F_X is non-decreasing in the sense of Proposition 1.278, i.e. for $a_1 < b_1, a_2 < b_2, \cdots, a_p < b_p$ we have

$$\sum_{k=0}^{p} (-1)^k \sum_{x \in \Delta_k^p} F_X(x) \geq 0.$$

(b) F_X is jointly right continuous in the co-ordinates, i.e.

$$\lim_{\substack{h_k \downarrow 0 \\ k \in \{1,2,\cdots,p\}}} F_X(x_1 + h_1, x_2 + h_2, \cdots, x_p + h_p) = F_X(x_1, x_2, \cdots, x_p).$$

In particular, F_X is right continuous in each co-ordinate, keeping other co-ordinates fixed.

(c) We have

$$\lim_{\substack{x_k \to \infty \\ k \in \{1,2,\cdots,p\}}} F_X(x_1, x_2, \cdots, x_p) = 1.$$

(d) For any fixed $j \in \{1, 2, \cdots, p\}$ and $x_1, x_2, \cdots, x_{j-1}, x_{j+1}, \cdots, x_p \in \mathbb{R}$, we have

$$\lim_{x_j \to -\infty} F_X(x_1, x_2, \cdots, x_p) = 0.$$

Proof. Statement (a) is already mentioned in Proposition 1.278.
Proofs of (b), (c) and (d) follow from Proposition 1.110, similar to the proof of Theorem 1.115. We only prove (b) to illustrate the idea.

$$\begin{aligned}
&\lim_{\substack{h_k \downarrow 0 \\ k \in \{1,2,\cdots,p\}}} F_X(x_1 + h_1, x_2 + h_2, \cdots, x_p + h_p) \\
&= \lim_{\substack{h_k \downarrow 0 \\ k \in \{1,2,\cdots,p\}}} \mathbb{P}_X((-\infty, x_1 + h_1] \times (-\infty, x_2 + h_2] \times \cdots \times (-\infty, x_p + h_p]) \\
&= \lim_{n \to \infty} \mathbb{P}_X(\left(-\infty, x_1 + \frac{1}{n}\right] \times \left(-\infty, x_2 + \frac{1}{n}\right] \times \cdots \times \left(-\infty, x_p + \frac{1}{n}\right]) \\
&= \mathbb{P}_X((-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_p])
\end{aligned}$$

$$= F_X(x_1, x_2, \cdots, x_p).$$

□

The next theorem, an analogue of Theorem 1.116, is stated without proof. The arguments required to prove this statement is beyond the scope of this course.

Theorem 1.282. *Any function $F : \mathbb{R}^p \to [0,1]$ satisfying the properties in Theorem 1.281 is the joint DF of some p-dimensional random vector.*

Note 1.283. Using arguments similar to above discussion, it is immediate that the joint DF of a random vector is non-decreasing in each co-ordinate, keeping other co-ordinates fixed.

Definition 1.284 (Mutually Independent RVs)**.** Let $\mathcal{I}$ be a non-empty indexing set (can be finite, countably infinite or uncountable). We say that a collection of RVs $\{X_\alpha : \alpha \in \mathcal{I}\}$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is mutually independent (or simply, independent) if for all finite subcollections $\{X_{\alpha_1}, X_{\alpha_2}, \cdots, X_{\alpha_n}\}$ we have

$$F_{X_{\alpha_1}, X_{\alpha_2}, \cdots, X_{\alpha_n}}(x_1, x_2, \cdots, x_n) = \prod_{j=1}^{n} F_{X_{\alpha_j}}(x_j), \forall x_1, x_2, \cdots, x_n \in \mathbb{R}.$$

Notation 1.285. If a collection of RVs $\{X_\alpha : \alpha \in \mathcal{I}\}$ is independent, we may also say that the RVs $X_\alpha, \alpha \in \mathcal{I}$ are independent.

Proposition 1.286. *The RVs $X_1, X_2, \cdots, X_p$, with $p \geq 2$, are independent if and only if*

$$F_{X_1, X_2, \cdots, X_p}(x_1, x_2, \cdots, x_p) = \prod_{j=1}^{p} F_{X_j}(x_j), \forall x_1, x_2, \cdots, x_p \in \mathbb{R}.$$

Proof. If the RVs $X_1, X_2, \cdots, X_p$ are independent, then the relation involving the joint DF follows from the definition.

Conversely, let $\mathcal{J} \subset \{1, 2, \cdots, p\}$. We would like to show that the subcollection $\{X_j : j \in \mathcal{J}\}$ is independent. Let Y be the $|\mathcal{J}|$-dimensional random vector with the component RVs $X_j, j \in \mathcal{J}$. Then F_Y is a joint DF of Y as well as a marginal DF of X. Then by Remark 1.280, for all $y \in \mathbb{R}^{|\mathcal{J}|}$,

$$F_Y(y) = \lim_{\substack{x_j \to \infty, j \notin \mathcal{J} \\ x_j = y_j, j \in \mathcal{J}}} F_X(x) = \lim_{\substack{x_j \to \infty, j \notin \mathcal{J} \\ x_j = y_j, j \in \mathcal{J}}} \prod_{j \notin \mathcal{J}} F_{X_j}(x_j) \prod_{j \in \mathcal{J}} F_{X_j}(x_j) = \prod_{j \in \mathcal{J}} F_{X_j}(y_j).$$

This shows that the subcollection $\{X_j : j \in \mathcal{J}\}$ is independent and the proof is complete. $\square$

Remark 1.287. It follows from the definition that if a collection of RVs $\{X_\alpha : \alpha \in \mathcal{I}\}$ is independent, then any subcollection of RVs $\{X_\alpha : \alpha \in \mathcal{J}\}$, with $\mathcal{J} \subset \mathcal{I}$ is also independent.

Definition 1.288 (Pairwise Independent RVs)**.** Let $\mathcal{I}$ be a non-empty indexing set (can be finite, countably infinite or uncountable). We say that a collection of RVs $\{X_\alpha : \alpha \in \mathcal{I}\}$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is pairwise independent if for all distinct indices $\alpha, \beta \in \mathcal{I}$, the subcollection $\{X_\alpha, X_\beta\}$ is independent, i.e.

$$F_{X_\alpha, X_\beta}(x_1, x_2) = F_{X_\alpha}(x_1) F_{X_\beta}(x_2), \forall x_1, x_2 \in \mathbb{R}.$$

Note 1.289. So far, we have not discussed examples of random vectors. In fact, as considered for RVs, we shall consider special classes of random vectors and explicit examples shall then be discussed.

Definition 1.290 (Discrete Random Vector)**.** A random vector $X = (X_1, X_2, \cdots, X_p)$ is said to be a discrete random vector if there exists a finite or countably infinite set $S \subset \mathbb{R}^p$ such that

$$1 = \mathbb{P}_X(S) = \mathbb{P}(X \in S) = \sum_{x \in S} \mathbb{P}_X(\{x\}) = \sum_{x \in S} \mathbb{P}(X = x)$$

and $\mathbb{P}(X = x) > 0, \forall x \in S$. In this situation, we refer to the set S as the support of the discrete random vector X.

Definition 1.291 (Joint Probability Mass Function for a discrete random vector)**.** Let $X = (X_1, X_2, \cdots, X_p)$ be a discrete random vector with support S_X. Consider the function $f_X : \mathbb{R}^p \to \mathbb{R}$ defined by

$$f_X(x) := \begin{cases} \mathbb{P}(X = x), \text{ if } x \in S_X, \\ 0, \text{ if } x \in S_X^c. \end{cases}$$

This function f_X is called the joint probability mass function (joint p.m.f.) of the random vector X.

Remark 1.292. Let $X = (X_1, X_2, \cdots, X_p)$ be a discrete random vector with joint DF F_X, joint p.m.f. f_X and support S_X. Then, similar to the p.m.f. for RVs, we have the following observations.

(a) The joint p.m.f. $f_X : \mathbb{R}^p \to \mathbb{R}$ is a function such that

$$f_X(x) = 0, \forall x \in S_X^c, \quad f_X(x) > 0, \forall x \in S_X, \quad \sum_{x \in S_X} f_X(x) = 1.$$

(b) $\mathbb{P}_X(S_X^c) = 1 - \mathbb{P}_X(S_X) = 0$. In particular, $\mathbb{P}(X = x) = f_X(x) = 0, \forall x \in S_X^c$.

(c) Since $\mathbb{P}_X(S_X) = 1$, for any $A \subseteq \mathbb{R}^p$ we have,

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}_X(A \cap S_X) = \sum_{x \in A \cap S_X} \mathbb{P}(X = x) = \sum_{x \in A \cap S_X} f_X(x).$$

Since S_X is finite or countably infinite, the set $A \cap S_X$ is also finite or countably infinite.

(d) By (c), for any $x = (x_1, x_2, \cdots, x_p) \in \mathbb{R}^p$, we consider $A = \prod_{j=1}^p (-\infty, x_j]$, we obtain

$$\begin{aligned} F_X(x) &= \mathbb{P}_X\left(\prod_{j=1}^p (-\infty, x_j]\right) \\ &= \mathbb{P}(X_1 \le x_1, X_2 \le x_2, \cdots, X_p \le x_p) \\ &= \sum_{y \in S_X \cap \prod_{j=1}^p (-\infty, x_j]} f_X(y). \end{aligned}$$

Therefore, the joint p.m.f. f_X is uniquely determined by the joint DF F_X and vice versa.

(e) To study a discrete random vector X, we may study any one of the following three quantities, viz. the joint law/distribution $\mathbb{P}_X$, the joint DF F_X or the joint p.m.f. f_X.

(f) For any $j \in \{1, 2, \cdots, p\}$, for $x_j \in \mathbb{R}$

$$\begin{aligned} F_{X_j}(x_j) &= \mathbb{P}(X_j \in (-\infty, x_j]) \\ &= \mathbb{P}(X_1 \in \mathbb{R}, \cdots, X_{j-1} \in \mathbb{R}, X_j \in (-\infty, x_j], X_{j+1} \in \mathbb{R}, \cdots, X_p \in \mathbb{R}) \\ &= \mathbb{P}_X(\mathbb{R} \times \cdots \times \mathbb{R} \times (-\infty, x_j] \times \mathbb{R} \times \cdots \times \mathbb{R}) \\ &= \sum_{y \in S_X \cap \mathbb{R} \times \cdots \times \mathbb{R} \times (-\infty, x_j] \times \mathbb{R} \times \cdots \times \mathbb{R}} f_X(y) \end{aligned}$$

$$= \sum_{\substack{y \in S_X \\ y_j \leq x_j}} f_X(y).$$

Consider $g_j : \mathbb{R} \to \mathbb{R}$ defined by $g_j(x) := \sum_{\substack{y \in S_X \\ y_j = x}} f_X(y)$. It is immediate that g_j satisfies the properties of a p.m.f. and $F_{X_j}(x_j) = \sum_{z \leq x_j} g_j(z)$ and $g_j(x) > 0$ if and only if $x \in \{t \in \mathbb{R} : y_j = t \text{ for some } y \in S_X\}$. Therefore, X_j is a discrete RV with p.m.f. g_j. More generally, all marginal distributions of X are also discrete. The function g_j is usually referred to as the marginal p.m.f. of X_j.

Remark 1.293. Let $\emptyset \neq S \subset \mathbb{R}^p$ be a finite or countably infinite set and let $f : \mathbb{R}^p \to \mathbb{R}$ be such that

$$f(x) = 0, \forall x \in S^c, \quad f(x) > 0, \forall x \in S, \quad \sum_{x \in S} f(x) = 1.$$

Then f is the joint p.m.f. of some p-dimensional discrete random vector X with support S. We are not going to discuss the proof of this statement in this course.

Theorem 1.294. *Let $X = (X_1, X_2, \cdots, X_p)$ be a discrete random vector with joint DF F_X, joint p.m.f. f_X and support S_X. Let f_{X_j} denote the marginal p.m.f. of X_j. Then $X_1, X_2, \cdots, X_p$ are independent if and only if*

$$f_{X_1, X_2, \cdots, X_p}(x_1, x_2, \cdots, x_p) = \prod_{j=1}^{p} f_{X_j}(x_j), \forall x_1, x_2, \cdots, x_p \in \mathbb{R}.$$

In this case, we have $S_X = S_{X_1} \times S_{X_2} \times \cdots \times S_{X_p}$, where S_{X_j} denotes the support of X_j.

Proof. By Proposition 1.286, the RVs $X_1, X_2, \cdots, X_p$ are independent if and only if

$$F_{X_1, X_2, \cdots, X_p}(x_1, x_2, \cdots, x_p) = \prod_{j=1}^{p} F_{X_j}(x_j), \forall x_1, x_2, \cdots, x_p \in \mathbb{R}.$$

If the condition for the joint p.m.f. holds as per the statement above, then the above condition for the joint DF holds and hence the required independence follows.
The proof of the converse statement is left as an exercise in Problem set 7.

To prove the statement for the support, observe that

$$\begin{aligned}
S_X &= \{x \in \mathbb{R}^p : f_X(x) > 0\} \\
&= \{x = (x_1, x_2, \cdots, x_p) \in \mathbb{R}^p : \prod_{j=1}^{p} f_{X_j}(x_j) > 0\} \\
&= \{x = (x_1, x_2, \cdots, x_p) \in \mathbb{R}^p : f_{X_j}(x_j) > 0, \forall j = 1, 2, \cdots, p\} \\
&= \prod_{j=1}^{p} \{x_j \in \mathbb{R} : f_{X_j}(x_j) > 0\} \\
&= S_{X_1} \times S_{X_2} \times \cdots \times S_{X_p}
\end{aligned}$$

This completes the proof. $\square$

Example 1.295. Given p.m.f.s $f_1, f_2, \cdots, f_p : \mathbb{R} \to [0,1]$ and corresponding support sets $S_1, S_2, \cdots, S_p$, consider the function $f : \mathbb{R}^p \to [0,1]$ defined by

$$f(x) := \prod_{j=1}^{p} f_j(x_j), \forall x = (x_1, x_2, \cdots, x_p) \in \mathbb{R}^p.$$

Then the set $S = S_1 \times S_2 \times \cdots \times S_p \subset \mathbb{R}^p$ is also finite or countably infinite and

$$f(x) = 0, \forall x \in S^c, \quad f(x) > 0, \forall x \in S, \quad \sum_{x \in S} f(x) = 1.$$

By Remark 1.293, we have that f is the joint p.m.f. of a p-dimensional discrete random vector such that the component RVs are independent, by Theorem 1.294. Using this method, we can construct many examples of discrete random vectors.

Remark 1.296. Let $X = (X_1, X_2, \cdots, X_p)$ be a discrete random vector with joint p.m.f. f_X and support S_X. Then $X_1, X_2, \cdots, X_p$ are independent if and only if

$$f_{X_1, X_2, \cdots, X_p}(x_1, x_2, \cdots, x_p) = \prod_{j=1}^{p} g_j(x_j), \forall x_1, x_2, \cdots, x_p \in \mathbb{R}$$

for some functions $g_1, g_2, \cdots, g_p : \mathbb{R} \to [0, \infty)$ with $S_j := \{x \in \mathbb{R} : g_j(x) > 0\}$ being finite or countably infinite and $S_X = S_1 \times S_2 \times \cdots \times S_p$. In this case, the marginal p.m.fs f_{X_j} have the form $c_j g_j$, where the number c_j can be determined from the relation $c_j = \left(\sum_{x \in S_j} g_j(x)\right)^{-1}$.

Example 1.297. Let $Z = (X, Y)$ be a 2-dimensional discrete random vector with the joint p.m.f. of the form

$$f_Z(x, y) = \begin{cases} \alpha(x + y), \text{ if } x, y \in \{1, 2, 3, 4\} \\ 0, \text{ otherwise} \end{cases}$$

for some constant $\alpha \in \mathbb{R}$. For f_Z to take non-negative values, we must have $\alpha > 0$. Now, $\sum_{x,y \in \{1,2,3,4\}} \alpha(x + y) = 1$ simplifies to $80\alpha = 1$ and hence $\alpha = \frac{1}{80}$. Also note that for this value of α, f_Z takes non-negative values. The support of Z is $\{(x, y) : x, y \in \{1, 2, 3, 4\}\} = \{1, 2, 3, 4\} \times \{1, 2, 3, 4\}$. The support of X is $\{1, 2, 3, 4\}$ and the marginal p.m.f. f_X can now be computed as

$$f_X(x) = \begin{cases} \sum_{y \in \{1,2,3,4\}} \frac{1}{80}(x + y), \text{ if } x \in \{1, 2, 3, 4\} \\ 0, \text{ otherwise} \end{cases} = \begin{cases} \frac{1}{40}(2x + 5), \text{ if } x \in \{1, 2, 3, 4\} \\ 0, \text{ otherwise} \end{cases}$$

By the symmetry of $f_Z(x, y)$ in the variables x and y, we conclude that $X \stackrel{d}{=} Y$. Note that $f_Z(1, 1) = \frac{1}{40}$ and $f_X(1)f_Y(1) = \frac{49}{1600}$. Hence X and Y are not independent.

Example 1.298. Let $U = (X, Y, Z)$ be a 3-dimensional discrete random vector with the joint p.m.f. of the form

$$f_U(x, y, z) = \begin{cases} \alpha xyz, \text{ if } x = 1, y \in \{1, 2\}, z \in \{1, 2, 3\} \\ 0, \text{ otherwise} \end{cases}$$

for some constant $\alpha \in \mathbb{R}$. For f_U to take non-negative values, we must have $\alpha > 0$. Now, $\sum_{x=1, y \in \{1,2\}, z \in \{1,2,3\}} \alpha xyz = 1$ simplifies to $18\alpha = 1$ and hence $\alpha = \frac{1}{18}$. Also note that for this value of α, f_U takes non-negative values. The support of U is $\{(x, y, z) : x = 1, y \in \{1, 2\}, z \in \{1, 2, 3\}\} = \{1\} \times \{1, 2\} \times \{1, 2, 3\}$. The support of X is $\{1\}$ and the marginal p.m.f. f_X can now be computed as

$$f_X(x) = \begin{cases} \sum_{y\in\{1,2\},z\in\{1,2,3\}} \frac{1}{18}yz, & \text{if } x = 1 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} 1, & \text{if } x = 1 \\ 0, & \text{otherwise} \end{cases}$$

as expected. Similar computation yields

$$f_Y(y) = \begin{cases} \frac{1}{3}, & \text{if } y = 1 \\ \frac{2}{3}, & \text{if } y = 2 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \frac{y}{3}, & \text{if } y \in \{1,2\} \\ 0, & \text{otherwise} \end{cases}$$

and

$$f_Z(z) = \begin{cases} \frac{1}{6}, & \text{if } z = 1 \\ \frac{1}{3}, & \text{if } z = 2 \\ \frac{1}{2}, & \text{if } z = 3 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \frac{z}{6}, & \text{if } z \in \{1,2,3\} \\ 0, & \text{otherwise} \end{cases}$$

Observe that $f_{X,Y,Z}(x,y,z) = f_X(x)f_Y(y)f_Z(z), \forall x,y,z$ and hence the RVs X, Y, Z are independent.

Remark 1.299 (Conditional Distribution for discrete random vectors). Let $X = (X_1, X_2, \cdots, X_{p+q})$ be a discrete random vector with support S_X and joint p.m.f. f_X. Let $Y = (X_1, X_2, \cdots, X_p)$ and $Z = (X_{p+1}, X_{p+2}, \cdots, X_{p+q})$. Then Y and Z both are discrete random vectors. Let f_Y and S_Y denote the joint p.m.f. and support of Y, respectively. Let f_Z and S_Z denote the joint p.m.f. and support of Z, respectively. For $z \in S_Z$, consider the set

$$T_z := \{y \in \mathbb{R}^p : (y,z) \in S_X\}.$$

The conditional p.m.f. of Y given $Z = z \in S_Z$ is defined by

$$f_{Y|Z}(y \mid z) := \mathbb{P}(Y = y \mid Z = z) = \frac{\mathbb{P}(Y = y, Z = z)}{\mathbb{P}(Z = z)} = \begin{cases} \frac{f_X(y,z)}{f_Z(z)}, & \text{if } y \in T_z \\ 0, & \text{otherwise.} \end{cases}$$

By definition, $f_{Y|Z}(y \mid z) \geq 0, \forall y \in \mathbb{R}^p$ and $\sum_{y\in\mathbb{R}^p} f_{Y|Z}(y \mid z) = \sum_{y\in T_z} f_{Y|Z}(y \mid z) = 1$. Therefore, for every $z \in S_Z$, the function $y \in \mathbb{R}^p \mapsto f_{Y|Z}(y \mid z)$ is a joint p.m.f. with support T_z. We refer to the probability law/distribution described by this p.m.f. as the conditional distribution of Y given $Z = z \in S_Z$. The conditional DF of Y given $Z = z \in S_Z$ is given by

$$F_{Y|Z}(y \mid z) := \mathbb{P}(Y \leq y \mid Z = z) = \frac{\mathbb{P}(Y \leq y, Z = z)}{\mathbb{P}(Z = z)} = \sum_{\substack{t\leq y\\ t\in T_z}} \frac{f_X(t,z)}{f_Z(z)} = \sum_{\substack{t\leq y\\ t\in T_z}} f_{Y|Z}(t \mid z),$$

where $t \leq y$ refers to component-wise inequalities $t_j \leq y_j$ for all $j = 1, 2, \cdots, p$.

Note 1.300. For notational convenience, we have discussed the conditional distribution of first p component RVs with respect to the final q component RVs. However, as long as the $(p+q)$-dimensional joint distribution is known, we can discuss the conditional distribution of any of the k-component RVs with respect to the other $(p+q-k)$-component RVs.

Note 1.301. When values for some of the components RVs are given, the conditional distribution provides an updated probability distribution for the rest of the component RVs.

Note 1.302. Let (X, Y) be a 2-dimensional discrete random vector such that X and Y are independent. Then $f_{X,Y}(x, y) = f_X(x)f_Y(y), \forall x, y \in \mathbb{R}$. Then

$$f_{Y|X}(y \mid x) = f_Y(y), \forall x \in S_X, y \in S_Y.$$

This statement can be generalized to higher dimensions with appropriate changes in the notation.

Example 1.303. In Example 1.297, we have, for fixed $x \in \{1, 2, 3, 4\}$,

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)}, & \text{if } y \in \{1,2,3,4\} \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} \frac{x+y}{2(2x+5)}, & \text{if } y \in \{1,2,3,4\} \\ 0, & \text{otherwise.} \end{cases}$$

Definition 1.304 (Continuous Random Vector and its Joint Probability Density Function (Joint p.d.f.))**.** A random vector $X = (X_1, X_2, \cdots, X_p)$ is said to be a continuous random vector if there exists an integrable function $f : \mathbb{R}^p \to [0, \infty)$ such that

$$F_X(x) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_p \leq x_p)$$