8. Week 8

Remark 8.1 (Moments of a Standard Normal RV). Let $X \sim N(0,1)$. Then X is symmetric about 0 and using Proposition 7.17, we conclude $\mathbb{E}X^n = 0$ for all odd positive integers n. If n is an even positive integer, then $n = 2m$ for some positive integer m and

$$\begin{aligned}
\mathbb{E}X^n &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^n \exp\left(-\frac{x^2}{2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2m} \exp\left(-\frac{x^2}{2}\right) dx \\
&= \frac{\sqrt{2}}{\sqrt{\pi}} \int_{0}^{\infty} x^{2m} \exp\left(-\frac{x^2}{2}\right) dx \\
&= \frac{2^m}{\sqrt{\pi}} \int_{0}^{\infty} y^{m-\frac{1}{2}} \exp(-y)\, dy, \text{ (putting } y = \frac{x^2}{2}) \\
&= \frac{2^m}{\sqrt{\pi}} \Gamma(m + \frac{1}{2}) \\
&= 2^m \left(m - \frac{1}{2}\right) \times \cdots \times \frac{3}{2} \times \frac{1}{2} \\
&= (2m-1) \times \cdots \times 3 \times 1 =: (2m-1)!!,
\end{aligned}$$

where we have used the properties of the Gamma function. In particular, $\mathbb{E}X^4 = 3$.

Remark 8.2. If $Z \sim N(0,1)$, it can be checked that $\mathbb{P}(|Z| \leq 3) \approx 0.997$ and $\mathbb{P}(|Z| \leq 6) \approx 0.9997$. More generally, for $X \sim N(\mu, \sigma^2)$, we have $\mathbb{P}(|X - \mu| \leq 3\sigma) \approx 0.997$ and $\mathbb{P}(|X - \mu| \leq 6\sigma) \approx 0.9997$. This shows that the values of a normal RV is quite concentrated near its mean.

Definition 8.3 (Beta function)**.** Recall that the integral $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx$ exists if and only if $\alpha > 0$ and $\beta > 0$. On $(0,\infty) \times (0,\infty)$, consider the function $(\alpha,\beta) \mapsto \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx$. It is called the Beta function and the value at any (α,β) is denoted by $B(\alpha,\beta)$.

Remark 8.4. Note that for $\alpha > 0, \beta > 0$, we have $B(\alpha,\beta) > 0$ and $B(\alpha,\beta) = B(\beta,\alpha)$. Moreover,

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Example 8.5 (Beta(α, β) RV)**.** Fix $\alpha > 0, \beta > 0$. By the properties of the Beta function described above, the function $f : \mathbb{R} \to [0, \infty)$ defined by

$$f(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \text{ if } x \in (0,1) \\ 0, \text{ otherwise.} \end{cases}$$

is a p.d.f.. An RV X is said to follow Beta(α, β) distribution or equivalently, X is a Beta(α, β) RV if its distribution is given by the above p.d.f.. If $\alpha = \beta$, then $f(1-x) = f(x), \forall x \in \mathbb{R}$ and hence $X \stackrel{d}{=} 1 - X$. Then, $X - \frac{1}{2} \stackrel{d}{=} \frac{1}{2} - X$, i.e., X is symmetric about $\frac{1}{2}$. For all $\alpha, \beta, r > 0$, we have

$$\mathbb{E}X^r = \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha+r-1}(1-x)^{\beta-1}\, dx = \frac{B(\alpha + r, \beta)}{B(\alpha, \beta)}$$

and in particular,

$$\mathbb{E}X = \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha}{\alpha+\beta}$$

and

$$\mathbb{E}X^2 = \frac{B(\alpha+2, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}.$$

Then

$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}.$$

We now study important inequalities in connection with moments of RVs and probabilities of events involving the RVs. Given any RV X, we shall always assume that it is either discrete with p.m.f. f_X or continuous with p.d.f. f_X, if not stated otherwise.

Note 8.6. At times, it is possible to compute the moments of an RV, but the computation of probability of certain events involving the RV may be difficult. The inequalities, that we are going to study, give us estimates of the probabilities in question.

Theorem 8.7. *Let X be an RV such that X is non-negative (i.e. $\mathbb{P}(X \geq 0) = 1$). Suppose that $\mathbb{E}X$ exists. Then for any $c > 0$, we have*

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}X}{c}.$$

Proof. We discuss the proof when X is a continuous RV with p.d.f. f_X. The case when X is discrete can be proved using similar arguments.
For $x < 0$, we have $F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X < 0) = 1 - \mathbb{P}(X \geq 0) = 0$ and hence $f_X(x) = 0, \forall x < 0$. Then,

$$\mathbb{E}X = \int_0^\infty x f_X(x)\, dx \geq \int_c^\infty x f_X(x)\, dx \geq c \int_c^\infty f_X(x)\, dx = c\,\mathbb{P}(X \geq c).$$

This completes the proof. □

Note 8.8. Under the assumptions of Theorem 8.7, we have $\mathbb{E}X \geq 0$.

The following special cases of Theorem 8.7 are quite useful in practice.

Corollary 8.9. *(a) Let X be an RV and let $h : \mathbb{R} \to [0, \infty)$ be a function such that $\mathbb{E}h(X)$ exists. Then for any $c > 0$, we have*

$$\mathbb{P}(h(X) \geq c) \leq \frac{\mathbb{E}h(X)}{c}.$$

(b) Let X be an RV and let $h : \mathbb{R} \to [0, \infty)$ be a strictly increasing function such that $\mathbb{E}h(X)$ exists. Then for any $c > 0$, we have

$$\mathbb{P}(X \geq c) = \mathbb{P}(h(X) \geq h(c)) \leq \frac{\mathbb{E}h(X)}{h(c)}.$$

(c) Let X be an RV such that $\mathbb{E}X$ exists, i.e. $\mathbb{E}|X| < \infty$. Considering the RV $|X|$, for any $c > 0$ we have

$$\mathbb{P}(|X| \geq c) \leq \frac{\mathbb{E}|X|}{c}.$$

(d) (Markov's inequality) Let $r > 0$ and let X be an RV such that $\mathbb{E}|X|^r < \infty$. Then for any $c > 0$, we have

$$\mathbb{P}(|X| \geq c) = \mathbb{P}(|X|^r \geq c^r) \leq c^{-r}\, \mathbb{E}|X|^r.$$

(e) (Chernoff's inequality) Let X be an RV with $\mathbb{E}e^{\lambda X} < \infty$ for some $\lambda > 0$. Then for any $c > 0$, we have

$$\mathbb{P}\{X \geq c\} = \mathbb{P}\{e^{\lambda X} \geq e^{\lambda c}\} \leq e^{-\lambda c}\, \mathbb{E}e^{\lambda X}.$$

Note 8.10. Let X be an RV with finite second moment, i.e. $\mu_2' = \mathbb{E}X^2 < \infty$. By Remark 6.30, the first moment $\mu_1' = \mathbb{E}X$ exists. Hence

$$\mathbb{E}(X-c)^2 = \mathbb{E}[X^2 + c^2 - 2cX] = \mathbb{E}X^2 + c^2 - 2c\,\mathbb{E}X = \mu_2' + c^2 - 2c\,\mu_1' < \infty$$

Therefore, all second moments of X about any point $c \in \mathbb{R}$ exists. In particular, $Var(X) = \mathbb{E}(X - \mu_1')^2 < \infty$. By a similar argument, for any RV X with finite variance, we have $\mathbb{E}X^2 < \infty$.

The next result is a special case of Markov's inequality.

Corollary 8.11 (Chebyshev's inequality)**.** *Let X be an RV with finite second moment (equivalently, finite variance). Then*

$$\mathbb{P}[|X - \mu_1'| \geq c] \leq \frac{1}{c^2}\mathbb{E}(X - \mu_1')^2 = \frac{1}{c^2}Var(X).$$

Remark 8.12. Another form of the above result is also useful. Under the same assumptions, for any $\epsilon > 0$ we have

$$\mathbb{P}[|X - \mu_1'| \geq \epsilon\,\sigma(X)] \leq \frac{1}{\epsilon^2},$$

where $\sigma(X)$ is the standard deviation of X. This measures the spread/deviation of the distribution (of X) about the mean in multiples of the standard deviation. The smaller the variance, lesser the spread.

Remark 8.13. In general, bounds in Theorem 8.7 or in Markov/Chebyshev's inequalities are very conservative. However, they can not be improved further. To see this, consider a discrete RV X with p.m.f. given by

$$f_X(x) := \begin{cases} \frac{3}{4}, & \text{if } x = 0, \\ \frac{1}{4}, & \text{if } x = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then $\mathbb{P}(X \geq 1) = \frac{1}{4} = \mathbb{E}X$, which is sharp. If we consider

$$f_X(x) := \begin{cases} \frac{3}{4}, & \text{if } x = 0, \\ \frac{1}{4}, & \text{if } x = 2, \\ 0, & \text{otherwise,} \end{cases}$$

then, $\mathbb{P}(X \geq 1) = \frac{1}{4} < \frac{1}{2} = \mathbb{E}X$.

Definition 8.14 (Convex functions)**.** Let I be an open interval in $\mathbb{R}$. We say that a function $h : I \to \mathbb{R}$ is convex on I if

$$h(\alpha x + (1 - \alpha)y) \leq \alpha h(x) + (1 - \alpha)h(y), \forall \alpha \in (0, 1), \forall x, y \in I.$$

We say that h is strictly convex on I if the above inequality is strict for all x, y and α.

We state the following result from Real Analysis without proof.

Theorem 8.15. *Let I be an open interval in $\mathbb{R}$ and let $h : I \to \mathbb{R}$ be a function.*

(a) If h is convex on $\mathbb{R}$, then h is continuous on $\mathbb{R}$.

(b) Let h be twice differentiable on I. Then,

  (i) h is convex if and only if $h''(x) \geq 0, \forall x \in I$.

  (ii) h is strictly convex if and only if $h''(x) > 0, \forall x \in I$.

The following result is stated without proof.

Theorem 8.16 (Jensen's Inequality)**.** *Let I be an interval in $\mathbb{R}$ and let $h : I \to \mathbb{R}$ be a convex function. Let X be an RV with support $S_X \subseteq I$. Then,*

$$h(\mathbb{E}X) \leq \mathbb{E}h(X),$$

provided the expectations exist. If h is strictly convex, then the inequality above is strict unless X is a degenerate RV.

Remark 8.17. Some special cases of Jensen's inequality are of interest.

(a) Consider $h(x) = x^2, \forall x \in \mathbb{R}$. Here, $h''(x) = 2 > 0, \forall x$ and hence h is convex on $\mathbb{R}$. Then $(\mathbb{E}X)^2 \leq \mathbb{E}X^2$, provided the expectations exist. We had seen this inequality earlier in Remark 6.34.

(b) For any integer $n \geq 2$, consider the function $h(x) = x^n$ on $[0, \infty)$. Here, $h''(x) = n(n-1)x^{n-2} \geq 0, \forall x \in (0, \infty)$ and hence h is convex. Then $(\mathbb{E}|X|)^n \leq \mathbb{E}|X|^n$, provided the expectations exist.

(c) Consider $h(x) = e^x, \forall x \in \mathbb{R}$. Here, $h''(x) = e^x > 0, \forall x$ and hence h is convex on $\mathbb{R}$. Then $e^{\mathbb{E}X} \leq \mathbb{E}e^X$, provided the expectations exist.

(d) Consider any RV X with $\mathbb{P}(X > 0) = 1$ and look at $h(x) := -\ln x, \forall x \in (0, \infty)$. Then $h''(x) = \frac{1}{x^2} > 0, \forall x \in (0, \infty)$ and hence h is convex. Then $-\ln(\mathbb{E}X) \leq \mathbb{E}(-\ln X)$, i.e. $\ln(\mathbb{E}X) \geq \mathbb{E}(\ln X)$, provided the expectations exist.

(e) Consider any RV X with $\mathbb{P}(X > 0) = 1$. Then $\mathbb{P}(\frac{1}{X} > 0) = 1$ and hence by (d), $-\ln(\mathbb{E}\frac{1}{X}) \leq \mathbb{E}(-\ln \frac{1}{X}) = \mathbb{E}(\ln X)$. Then $(\mathbb{E}\frac{1}{X})^{-1} = e^{-\ln(\mathbb{E}\frac{1}{X})} \leq e^{\mathbb{E}(\ln X)} \leq \mathbb{E}X$, by (c). This inequality holds, provided all the expectations exist. We may think of $\mathbb{E}X$ as the arithmetic mean (A.M.) of X, $e^{\mathbb{E}(\ln X)}$ as the geometric mean (G.M.) of X, and $\frac{1}{\mathbb{E}[\frac{1}{X}]}$ as the harmonic mean (H.M.) of X. The inequality obtained here is related to the classical A.M.-G.M.-H.M. inequality (see problem set 8).

Note 8.18 (Why should we look at multiple RVs together?)**.** Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ associated with a random experiment $\mathcal{E}$. As motivated earlier, an RV associates some numerical quantity to each of the outcomes of the experiment. Such numerical quantities help us in the understanding of characteristics of the outcomes. However, it is important to note that, in practice, we may be interested in looking at these characteristics of the outcomes at the same time. This also allows us to see if the characteristics in question may be related. If we perform the random experiment separately for each of these characteristics, then there is also the issue of cost and time associated with the repeated performance of the experiment. Keeping this in mind, we now choose to consider multiple characteristics of the outcomes at the same time. This leads us to the concept of Random Vectors, which allows us to look at multiple RVs at the same time.

Example 8.19. Consider the random experiment of rolling a standard six-sided die three times. Here, the sample space is

$$\Omega = \{(i, j, k) : i, j, k \in \{1, 2, 3, 4, 5, 6\}\}.$$

Suppose we are interested in the sum of the first two rolls and the sum of all rolls. These characteristics of the outcomes can be captured by the RVs $X, Y : \Omega \to \mathbb{R}$ defined by $X((i, j, k)) := i + j$ and $Y((i, j, k)) := i + j + k$ for all $(i, j, k) \in \Omega$. If we look at X and Y simultaneously, we may comment on whether a 'large' value for X implies a 'large' Y and vice versa.

Definition 8.20 (Random Vector)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X = (X_1, X_2, \cdots, X_p) : \Omega \to \mathbb{R}^p$ is called a p-dimensional random vector (or simply, a random vector, if the dimension p is clear from the context). Here, the component functions are denoted by $X_1, X_2, \cdots, X_p$ and each of these are real valued functions defined on the sample space Ω and hence are RVs.

Note 8.21. 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 8.22 (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 8.23. 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 3.10, we get the next result. The proof is skipped for brevity.

Proposition 8.24. *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 8.25 (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 8.26. 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 8.27 (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 8.28. 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 8.29. Recall that for an RV Y, we have $F_Y(b) - F_Y(a) = \mathbb{P}(a < Y \leq b) \geq 0$ for all $a, b \in \mathbb{R}$ with $a < b$.

Proposition 8.30. *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,*

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

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 2.8)

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

The result follows. $\square$

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

Notation 8.31. 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 8.30 can now be generalized to higher dimensions as follows. We skip the details of the proof for brevity.

Proposition 8.32. *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 8.33 (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*

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

Proof. As in the proof of Theorem 4.1, using Proposition 3.17, we have

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

$$= \mathbb{P}(X_1 \in \mathbb{R}, \cdots, X_{j-1} \in \mathbb{R}, X_j \in (-\infty, x], X_{j+1} \in \mathbb{R}, \cdots, X_p \in \mathbb{R})$$
$$= \mathbb{P}(X_j \in (-\infty, x]) = F_{X_j}(x).$$

This completes the proof. □

Remark 8.34. Using Proposition 8.33, 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 4.1.

Theorem 8.35. *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 8.32, 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 8.32.
Proofs of (b), (c) and (d) follow from Proposition 3.17, similar to the proof of Theorem 4.1. We only prove (b) to illustrate the idea.

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

□

The next theorem, an analogue of Theorem 4.2, is stated without proof. The arguments required to prove this statement is beyond the scope of this course.

Theorem 8.36. *Any function $F : \mathbb{R}^p \to [0,1]$ satisfying the properties in Theorem 8.35 is the joint DF of some p-dimensional random vector.*

Note 8.37. 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 8.38 (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 8.39. 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 8.40. *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 8.34, 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 8.41. 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 8.42 (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 8.43. 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 8.44 (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 8.45 (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 8.46. 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 $x_j \in \mathbb{R}$

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

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 8.47. 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 8.48. *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 8.40, 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. □

Example 8.49. 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 8.47, 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 8.48. Using this method, we can construct many examples of discrete random vectors.

Remark 8.50. 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 8.51. 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 8.52. 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.