Section 3: Random vector and its distribution

Shruti Sawant and Animesh Kumar Email: animesh@ee.iitb.ac.in

1 Random Vectors

Random vectors are used to model uncertain *parameters* in engineering systems. An understanding of random vectors is also essential in studying probability based convergence laws. We illustrated a few examples involving random vectors.

Example 1.1 (ETA in two hop journey). *Consider an ETA problem, where a person goes from point A in city X to point B in city Y, while involving a long distance flight between the two cities. The total transit time will be $T_X + T_Y + T_{XY}$, where T_X is the transit time in city X, T_Y is the transit time in city Y, and T_{XY} is the long distance flight time.*

The distribution of random vector (T_X, T_Y) will govern the properties of ETA in this case and naturally motivates the use of random vectors.

Example 1.2 (Image Denoising). *Contemporary digital imaging works with image sensors with very high spatial density of pixels. For example, in about an inch-square, it is common to pack few to ten of (RGB) megapixels. In low-light or high ISO speed scenarios, it is common to get speckle noise in digital camera images. The grains in image are due to hardware limits. An example is illustrated in Figure 1.*

Figure 1: In a low light scenario (near sunset), observe speckle noise which hides the texture of waves as well as imparts grain on an image of clear sky. The picture was taken with a DSLR.

Let $W(x, y)$ be the speckle noise values where (x, y) are integer valued and vary over the pixel indices. A common technique in image processing is to denoise the image by utilizing the underlying 'signal structure' of images. This process is called as denoising. The denoising is accomplished by treating $W(x, y)$ as a random vector of independent noise values.

The definition of random vectors is given first. It is a direct extension of the definition of a random variable.

Definition 1.1 (Random vectors)**.** *Let* $\vec{X} := (X_1, X_2, \dots, X_n)$*. Then* $\vec{X}$ *is a random vector if it satisfies the following properties:*

1. $\mathbb{P}(\vec{X} \in (-\infty, \infty)^n) = 1$*, that is,* $\vec{X}$ *has finite elements with probability one.*
2. $\mathbb{P}((X_1, X_2, X_3, \dots, X_n) \in [a_1, b_1) \times [a_2, b_2) \dots \times [a_n, b_n))$ *is well defined for any* $-\infty < a_i < b_i < \infty, \quad i = 1, 2, \dots, n.$
3. *For any disjoint sets* $I_1, I_2, \dots$ *in* $\mathbb{R}^n$*,*

$$\mathbb{P}(\vec{X} \in \bigcup_{\infty}^{i=1} I_i) = \sum_{i=1}^{\infty} \mathbb{P}(\vec{X} \in I_i). \tag{1}$$

Similar to random variables, a random vector's probability calculations can be done using its CDF. The (joint) CDF of a random vector is defined next.

Definition 1.2 (Joint CDF of a random vector)**.** *Let* $\vec{X} := (X_1, X_2, \dots, X_n)$ *be a random vector. Its CDF is defined as follows:*

$$F_{\vec{X}}(x) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n). \tag{2}$$

that is, it is probability of the intersection of events $\{X_i \leq x_i\}, i = 1, \dots, n.$

Note that the random variables $X_1, \dots, X_n$ forming a random vector need not be independent. Thus, generally speaking, finding the joint CDF of a random vector is more difficult than finding the joint CDF of independent random variables put together as a random vector. We enlist the key properties of a CDF next.

1. The joint CDF is zero when any of the $x_i, 1 \leq i \leq n$ is reduced towards $-\infty$. That is,

$$\lim_{x_i \downarrow -\infty} F_{\vec{X}}(x_1, x_2, \dots, x_n) = 0. \tag{3}$$

   for all $x_j \in \mathbb{R}, \quad 1 \leq i \leq n, j \neq i$. Further,

$$\lim_{x_1, x_2, \dots, x_n \uparrow \infty} F_{\vec{X}}(x_1, x_2, \dots, x_n) = 1. \tag{4}$$

2. The distribution of a subset of random variables in $\vec{X}$ can be obtained by marginalization from the distribution $F_{\vec{X}}(\vec{x})$. The process is illustrated for a single random variable X_i below. For the distribution of X_i, taking limits $x_j \to \infty$ for all $j \neq i$ will result in the distribution for X_i. That is,

$$F_{X_i}(x_i) = \lim_{j \neq i, x_j \to \infty} F_{\vec{X}}(\vec{x}) = F_{\vec{X}}(\vec{x}). \tag{5}$$

3. For discrete random vector, we get an n dimensional PMF characterizing its CDF:

$$p_{\vec{X}}(\vec{x}) = \mathbb{P}(X_1 = x_1, X_1 = x_1, \dots, X_n = x_n). \tag{6}$$

   And, for continuous random vector, we get n dimensional PDF characterizing its CDF:

$$f_{\vec{X}}(\vec{x}) = \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F_{\vec{X}}(\vec{x}). \tag{7}$$

2 Independence of random variables

As discussed in the first section, convergence laws are the cornerstone of modern statistics. Convergence laws depend on repeated experiments, where independent random variables can be drawn from the same distribution *repeatedly*. Apart from repeated experiments, independent random variables show up in parameters of 'non-interacting' systems. The definition of independent random vector is discussed first.

Definition 2.1 (Independent random vector)**.** *A random vector* $\vec{X} := (X_1, X_2, \dots, X_n)$ *is (statistically) independent if*

$$F_{\vec{X}}(x) = F_{X_1}(x_1)F_{X_2}(x_2)\dots F_{X_n}(x_n) \tag{8}$$

for all $\vec{x} \in \mathbb{R}^n$*.*

For a continuous random vector $\vec{X}$, the independence condition is equivalent to

$$f_{\vec{X}}(\vec{x}) = f_{X_1}(x_1)f_{X_2}(x_2)\dots f_{X_n}(x_n), \quad \vec{x} \in \mathbb{R}^n. \tag{9}$$

For a discrete random vector $\vec{X}$, the independence condition is equivalent to

$$p_{\vec{X}}(\vec{x}) = p_{X_1}(x_1)p_{X_2}(x_2)\dots p_{X_n}(x_n), \quad \vec{x} \in \mathbb{R}^n. \tag{10}$$

Independent and identically distributed (IID) random vector is discussed next.

Definition 2.2 (IID random vector)**.** *The random variables* $X_1, X_2, \dots, X_n$ *are independent and identically distributed (IID) if*

1. *The random variables* $X_1, X_2, \dots, X_n$ *are independent.*
2. *Each random variable* $X_i, 1 \le i \le n$ *has the same distribution.*

As before, the IID condition can be specialized for the distributions of continuous IID random variables and discrete IID random variables separately as:

$$f_{\vec{X}}(\vec{x}) = \prod_{i=1}^{n} f_X(x_i), \quad \vec{x} \in \mathbb{R}^n \tag{11}$$

$$p_{\vec{X}}(\vec{x}) = \prod_{i=1}^{n} p_X(x_i) \quad \vec{x} \in \mathbb{R}^n. \tag{12}$$

3 Conditional distribution

The core idea behind conditional distribution is that in a pair of random variable (X, Y), after a random experiment, the value of $Y = y$ is revealed and an update is sought on the distribution of X. That is, initially the distribution of X is given by $F_X(x)$. And, (X, Y) have a joint distribution $F_{X,Y}(x, y)$. If we marginalize $F_{X,Y}(x, y)$, we would obtain $F_X(x)$. However, as an outcome of a random experiment, the value $Y = y$ is realized. Is there a more suitable CDF for X given that $Y = y$? The answer is in affirmative and it will be discussed next.

For discrete events it is known that

$$\mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}, \text{ for } \mathbb{P}(Y = y) \neq 0. \tag{13}$$

For discrete case such formula can be written for any y such that $\mathbb{P}(Y = y) > 0$. However, this does not work for continuous random variables since if Y were continuous then $\mathbb{P}(Y = y) = 0$ for all $y \in \mathbb{R}$.

For the continuous case, the conditional PDF is used to define a conditional distribution. For $y : f_Y(y) > 0$,

$$f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}. \tag{14}$$

Using this conditional PDF, any probability over the random variable $X|Y = y$ can be computed. For instance, the conditional CDF is obtained from the conditional PDF as:

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(x|y)dt. \tag{15}$$

One way to interpret conditional density is using approximate limits. This is done as follows. By the definition of joint PDF,

$$\mathbb{P}\left((X,Y) \in [[x, x+\delta x) \times [y, y+\delta y)]\right) \approx f_{X,Y}(x,y)\delta x \delta y. \tag{16}$$

With the same approach on the random variable Y,

$$\mathbb{P}(Y \in [y, y+\delta y)) \approx f_Y(y)\delta y. \tag{17}$$

So,

$$\begin{align}\mathbb{P}(X \in [x, x+\delta x)|Y \in [y, y+\delta y)) &\approx \frac{f_{X,Y}(x,y)\delta x\delta y}{f_Y(y)\delta y} \tag{18}\\ &= \frac{f_{X,Y}(x,y)\delta x}{f_Y(y)}. \tag{19}\end{align}$$

With $\delta y \downarrow 0$,

$$\mathbb{P}(X \in [x, x+\delta x)|Y = y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}\delta x \tag{20}$$

or $f_{X,Y}(x,y)/f_Y(y)$ can be understood as the PDF of $X|Y = y$ whenever $f_Y(y) > 0$.

We illustrate an example involving random vectors next.

Example 3.1 (Clocked Arrivals)**.** *Consider a discrete time (clock based) task arrival setup. It is assumed that more than one tasks do not arrive at a given discrete-time (clock) instant. Let Z_i be the random variable that models the number of tasks arriving at the time i, where i is a positive integer. Let*

$$Z_i = \begin{cases} 1 & \textit{if a task arrives at time } i \\ 0 & \textit{otherwise.} \end{cases} \tag{21}$$

Let $Z_i \sim Ber(p)$, where $0 < p < 1$ is a parameter. Further, assume that $Z_1, Z_2, \dots$ are IID.

For analysis purpose we are interested in inter arrival time distribution as well as the number of arrivals till time n. Let S_n be the number of tasks arrived till time n. Let $X_1, X_2, \dots$ be the inter-arrival times. Then $X_1, X_2, \dots$ are IID (verify!). To find the distribution of X, note that:

$$\begin{aligned} p_X(1) &= p \\ p_X(2) &= \mathbb{P}(Z_1 = 0, Z_2 = 1) \\ &= \mathbb{P}(Z_1 = 0)\mathbb{P}(Z_2 = 1). \\ &= (1-p)p \\ p_X(3) &= \mathbb{P}(Z_1 = 0, Z_2 = 0, Z_3 = 1) \\ &= (1-p)^2 p \\ \textit{and in general, } p_X(k) &= (1-p)^k p. \end{aligned} \tag{22}$$

This distribution is geometric with parameter $(1-p)$. Observe that $p_X(k)$ decreases exponentially in k.

The distribution of S_n can be obtained through either X_i's or Z_i's. Observe that:

$$S_n = \sum_{i=1}^{n} Z_i \tag{23}$$

$$\text{and } p_{S_n}(k) = \begin{cases} 1 & k < 0, k > n \\ 0 & 0 \le k \le n. \end{cases} \tag{24}$$

We say that S_n is binomial random variable. By previously discussed weak law of large numbers $\frac{S_n}{n}$ will converge to p in probability.

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \le \delta_n \tag{25}$$

where δ_n vanishes as n becomes large.