CST 3526 Stochastic Process Lecture 2 - 09/25/2025

Lecture 2: Review of Probability Theory II

Lecturer: Ziqiao Wang

In this lecture[1], we will introduce several ways of specifying a discrete or continuous random variable or distribution, including the cumulative distribution function, probability mass function, and probability density function. We then cover functions of a random variable, the expected value and variance. Next, we extend these concepts to pairs of random variables, discussing independence, correlation and covariance, as well as conditional probability and conditional expectation.

1 CDF, PMF and PDF

All events of interest on the real line can be expressed as sets of the form $\{\zeta : X(\zeta) \leq b\}, \quad b \in \mathbb{R}$.

Definition 1.1 (Cumulative distribution function)**.** The cumulative distribution function (CDF) of a random variable X is defined as the probability of the event $\{X \leq x\}$:

$$F_X(x) = P(X \leq x) \qquad \text{for } -\infty < x < +\infty.$$

Clearly, $F_X(x)$ is a function of the variable x, since its value changes as x varies.

We are now ready to state the fundamental properties of the CDF. From the axioms of probability and their corollaries, it follows that the CDF satisfies the following properties:

(i) $0 \leq F_X(x) \leq 1$.

(ii) $\lim_{x\to\infty} F_X(x) = 1$.

(iii) $\lim_{x\to-\infty} F_X(x) = 0$.

(iv) $F_X(x)$ is a nondecreasing function of x, that is, if $a < b$, then $F_X(a) \leq F_X(b)$.

(v) $F_X(x)$ is continuous from the right, that is, for $h > 0$,

$$F_X(b) = \lim_{h\to 0} F_X(b+h) = F_X(b^+).$$

(vi) $\mathbb{P}[a < X \leq b] = F_X(b) - F_X(a)$.

(vii) $\mathbb{P}[X = b] = F_X(b) - F_X(b^-)$.

(viii) $\mathbb{P}[X > x] = 1 - F_X(x)$.

Remark 1.1. *If the CDF is continuous at a point b, then* $P(X = b) = 0$.

We may loosely say that a R.V. X is discrete if it takes values from a countable set, i.e. $\Omega_X = \{x_1, x_2, \dots\}$. A more formal definition is given below.

Definition 1.2 (Discrete R.V. and PMF)**.** A discrete random variable X is defined as a random variable whose CDF is a right-continuous, staircase function of x, with jumps at a countable set of points $\{x_1, x_2, \dots\}$. The probability mass function (PMF) of a discrete random variable X is defined as:

$$p_X(x_k) \triangleq P(X = x_k) \qquad \text{for } x_k \in \Omega_X.$$

[1] *Reading:* Chapter 4-5 of Leon-Garcia.

Notably, the CDF of a discrete random variable is given by the cumulative probability of all outcomes less than x, and can be expressed as a weighted sum of unit step functions:

$$F_X(x) = \sum_k p_X(x_k) u(x - x_k),$$

where

$$u(x) = \begin{cases} 0, & \text{for } x < 0, \\ 1, & \text{for } x \geq 0, \end{cases}$$

is the unit step function.

Definition 1.3 (Continuous R.V. and PDF)**.** If $F_X(x)$ is continuous everywhere on $\mathbb{R}$, then we say the random variable is a continuous random variable. The probability density function (PDF) of a continuous random variable X is defined as, for all $x \in \mathbb{R}$,

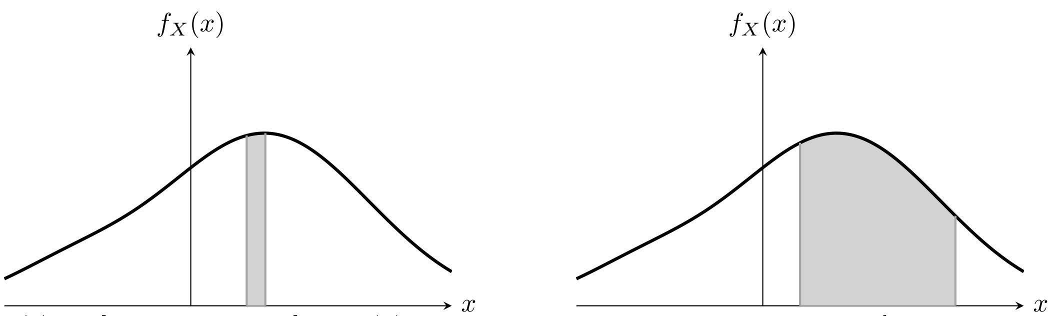
$$f_X(x) \triangleq \begin{cases} \dfrac{d}{dx} F_X(x), & \text{if } \dfrac{d}{dx} F_X(x) \text{ exists;} \\ 0, & \text{otherwise.} \end{cases}$$

Definition 1.4 (Mixed Typy R.V.)**.** A random variable which is neither continuous nor discrete is termed a random variable of mixed type.

A mixed-type random variable is one whose CDF exhibits jumps at a countable set of points $\{x_1, x_2, \ldots\}$ while also increasing continuously over at least one interval of x values. The CDF of such a random variable can be expressed in the following form:

$$F_X(x) = pF_1(x) + (1-p)F_2(x),$$

where $p \in (0,1)$, and $F_1(x)$ is the CDF of a discrete R.V. and $F_2(x)$ is the CDF of a continuous R.V.



(a) $P[x < X \leq x + dx] \approx f_X(x)\, dx$

(b) $P[a \leq X \leq b] = \int_a^b f_X(x)\, dx$

PDF. The PDF represents the "density" of probability at a point x in the following sense: for a sufficiently small $h > 0$, the probability that X falls within a small interval around x, namely $[x, x+h]$, is approximately proportional to the value of the PDF at x.

$$P(x < X \leq x + h) = F_X(x+h) - F_X(x) = \frac{F_X(x+h) - F_X(x)}{h} h \simeq f_X(x)h.$$

The PDF satisfies the following properties:

(i) $f_X(x) \geq 0$.

(ii) $P(a \leq X \leq b) = \int_a^b f_X(x)dx$.

(iii) $F_X(x) = \int_{-\infty}^{b} f_X(t)dt$.

(iv) $1 = \int_{-\infty}^{+\infty} f_X(t)dt$.

The PDF completely specifies the behavior of continuous random variables.

2 Functions of a Random Variable

Let X be a R.V. and g(x) a real-valued function defined on $\mathbb{R}$. Define $Y = g(X)$, then Y is itself a R.V., and the probabilities associated with the possible values of Y are determined by the mapping $g(x)$.

The CDF of Y is defined as the probability of the event $\{Y \le y\}$. In principle, it can be obtained by evaluating the probability of the equivalent event $\{g(X) \le y\}$.

Example 1 (A Linear Function)**.** Let the random variable Y be defined by

$$Y = aX + b,$$

where a is a nonzero constant. Suppose that X has CDF $F_X(x)$. Find $F_Y(y)$.

The event $\{Y \le y\}$ occurs when $A = \{aX + b \le y\}$ occurs. If $a > 0$, then $A = \{X \le (y-b)/a\}$, and thus

$$F_Y(y) = P\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right), \qquad a > 0.$$

On the other hand, if $a < 0$, then $A = \{X \ge (y-b)/a\}$, and

$$F_Y(y) = P\left(X \ge \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right), \qquad a < 0.$$

We can obtain the PDF of Y by differentiating with respect to y. Using the chain rule,

$$\frac{dF}{dy} = \frac{dF}{du}\frac{du}{dy},$$

where u is the argument of F. In this case, $u = (y-b)/a$. Hence, for $a > 0$,

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right), \qquad a > 0,$$

and for $a < 0$,

$$f_Y(y) = \frac{1}{-a} f_X\left(\frac{y-b}{a}\right), \qquad a < 0.$$

The above two results can be written compactly as

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right). \tag{1}$$

Example 2 (A Linear Function of a Gaussian Random Variable)**.** Let X be a R.V. with a Gaussian PDF with mean m and standard deviation σ:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-m)^2/2\sigma^2}, \qquad -\infty < x < \infty. \tag{2}$$

Let $Y = aX + b$, then find the PDF of Y. Substitution of Eq. (2) into Eq. (1) yields

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,|a\sigma|} e^{-(y-b-am)^2/2(a\sigma)^2}.$$

Note that Y also has a Gaussian distribution with mean $b + am$ and standard deviation $|a|\sigma$. Therefore, *a linear function of a Gaussian R.V. is also a Gaussian R.V.*.

3 Expectation and Variance

Definition 3.1 (Expectation)**.** The expected value or mean of a random variable X is defined by

$$\mathbb{E}[X] \triangleq \int_{-\infty}^{+\infty} t f_X(t) dt$$

Expectation is a linear operator:

(i) $\mathbb{E}[g(X) + h(X)] = \mathbb{E}[g(X)] + \mathbb{E}[h(X)]$.

(ii) $\mathbb{E}[aX] = a\mathbb{E}[X]$.

(iii) $\mathbb{E}[X + c] = \mathbb{E}[X] + c$.

Definition 3.2 (Variance)**.** The variance of the random variable X is defined by

$$\mathrm{Var}[X] \triangleq \mathbb{E}_X\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X].$$

The standard deviation of the R.V. X is defined by

$$\mathrm{Std}[X] \triangleq (\mathrm{Var}[X])^{1/2}.$$

(i) $\mathrm{Var}[X + c] = \mathrm{Var}[X]$.

(ii) $\mathrm{Var}[cX] = c^2\mathrm{Var}[X]$.

Definition 3.3 (Moments)**.** The nth moment of the R.V. X is defined by

$$\mathbb{E}[X^n] \triangleq \int_{-\infty}^{+\infty} x^n f_X(x) dx.$$

4 Two Probability Inequalities

Theorem 4.1 (Markov's Inequality)**.** *If X is a nonnegative R.V. with mean $\mathbb{E}[X]$, then for any $a \geq 0$*

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Theorem 4.2 (Chebyshev's Inequality)**.** *Let X be a R.V. with variance σ^2 and expected value m, then for any $a \geq 0$, we have*

$$P(|X - m| \geq a) \leq \frac{\sigma^2}{a^2}.$$

5 Pairs of R.V.'s

Consider a random experiment with sample space Ω and event class $\mathcal{F}$. We now define a function that assigns to each outcome $\omega \in \Omega$ an ordered pair of real numbers $(X(\omega), Y(\omega))$. In other words, we are dealing with a vector-valued R.V. that maps Ω into the real plane $\mathbb{R}^2$.

Suppose (X, Y) takes values from a countable set $\Omega_{X,Y} = \{(x_j, y_k) : j = 1, 2, \ldots; k = 1, 2, \ldots\}$. The **joint probability mass function** of (X, Y) specifies the probability of the event $X = x, , Y = y$ for each pair $(x, y) \in \Omega_{X,Y}$.

$$P_{X,Y}(x, y) = P[\{X = x\} \cap \{Y = y\}] \triangleq P[X = x, Y = y]$$

for $x, y \in \mathbb{R}^2$.

The joint PMF of (X, Y) characterizes the joint behavior of X and Y. We are also also interested in the probabilities of events involving each R.V. individually. These can be obtained from the marginal PMFs:

$$p_X(x_j) = P(X = x_j) = P(X = x_j, Y = \text{Anything}) = \sum_{k=1}^{\infty} P(X = x_j, Y = y_k),$$

similarly,

$$p_Y(y_k) = P(Y = y_k) = P(X = \text{Anything}, Y = y_k) = \sum_{j=1}^{\infty} P(X = x_j, Y = y_k),$$

The **joint cumulative distribution function** of X and Y is defined as the probability of the event $\{X \le x_1\} \cap \{Y \le y_1\}$:

$$F_{X,Y}(x_1, y_1) = P(X \le x_1, Y \le y_1).$$

The joint CDF satisfies the following properties.

(i) **Monotonicity.** The joint cdf is a nondecreasing function of x and y:

$$F_{X,Y}(x_1, y_1) \;\le\; F_{X,Y}(x_2, y_2) \quad \text{if } x_1 \le x_2 \text{ and } y_1 \le y_2. \tag{3}$$

(ii) **Boundary values.**

$$F_{X,Y}(x_1, -\infty) = 0, \qquad F_{X,Y}(-\infty, y_1) = 0, \qquad F_{X,Y}(\infty, \infty) = 1. \tag{4}$$

(iii) **Marginals (remove one constraint).**

$$F_X(x_1) = F_{X,Y}(x_1, \infty) \quad \text{and} \quad F_Y(y_1) = F_{X,Y}(\infty, y_1). \tag{5}$$

(iv) **Right–continuity in each coordinate ("north" and "east").**

$$\lim_{x \to a^+} F_{X,Y}(x, y) = F_{X,Y}(a, y) \quad \text{and} \quad \lim_{y \to b^+} F_{X,Y}(x, y) = F_{X,Y}(x, b). \tag{6}$$

(v) **Probability of a rectangle.**

$$P\left(x_1 < X \le x_2,\ y_1 < Y \le y_2\right) = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1). \tag{7}$$

We say that the R.V.'s X and Y are jointly continuous if the probabilities of events involving (X, Y) can be expressed as an integral of a PDF. Specifically, there exists a nonnegative function $f_{X,Y}(x, y)$ called the joint PDF, defined on the real plane such that for every event B, a subset of the plane,

$$P((X, Y) \in B) = \int_B \int f_{X,Y}(x', y') dx' dy'.$$

When B is the entire plane, the integral must equal one:

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

The joint CDF can be obtained in terms of the joint PDF of jointly continuous R.V.'s by integrating over the semi-infinite rectangle defined by (x, y):

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x', y') dx' dy'.$$

It then follows that if X and Y are jointly continuous R.V.'s, then the PDF can be obtained from the CDF by differentiation:

$$f_{X,Y}(x',y') = \frac{\partial F_{X,Y}(x,y)}{\partial x \partial y}.$$

The probability of a rectangular region is obtained by letting $B = \{(x,y) : a_1 < x \le b_1, a_2 < y \le b_2\}$:

$$P(a_1 < X \le b_1, a_2 < Y \le b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{X,Y}(x',y') dx' dy'.$$

The marginal PDF's $f_X(x)$ and $f_Y(y)$ are obtained by taking the derivative of the corresponding marginal CDF's, $F_X(x) = F_{X,Y}(x,\infty)$ and $F_Y(y) = F_{X,Y}(\infty, y)$. Thus,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y') dy',$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x',y) dx'.$$

6 Independence, Conditional Probability and Conditional Expectation

Discrete R.V.'s X and Y are independent if and only if the joint PMF is equal to the product of the marginal PMF's

$$p_{X,Y}(x_i, y_k) = P(X = x_j, Y = y_k) = P(X = x_j)P(Y = y_k) = p_X(x_j)p_Y(y_k).$$

In general, **two R.V.'s X and Y are independent if and only if their joint CDF equals the product of their marginal CDFs**.

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \quad \text{for } \forall x, y$$

Similarly, **if X and Y are jointly continuous, then they are independent if and only if their joint PDF equals the product of their marginal PDFs**:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \text{for } \forall x, y$$

Remark 6.1. *If X and Y are independent R.V.'s, then the R.V.'s defined by any pair of functions $g(X)$ and $h(Y)$ are also independent.*

The joint moments of two R.V. X and Y summarize information about their joint behavior. The jkth joint moment of X and Y is defined by

$$\mathbb{E}\left[X^j Y^k\right] = \begin{cases} \iint_{-\infty}^{\infty} x^j y^k \, f_{X,Y}(x,y) \, dx \, dy, & X, Y \text{ jointly continuous,} \\ \sum_i \sum_n x_i^j y_n^k \, p_{X,Y}(x_i, y_n), & X, Y \text{ discrete.} \end{cases}$$

It is customary to refer to the $j = 1, k = 1$ moment, $\mathbb{E}[XY]$, as the correlation of X and Y. When $\mathbb{E}[XY] = 0$, we say that X and Y are orthogonal.

The jkth central moment of X and Y is defined as the joint moment of the centered R.V.'s $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$,

$$\mathbb{E}\left[(X - \mathbb{E}[X])^i (Y - \mathbb{E}[Y])^k\right].$$

The covariance of X and Y is defined as the $j = k = 1$ central moment:

$$\mathrm{Cov}(X,Y) \triangleq \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Note that $\mathrm{Cov}(X,Y) = \mathbb{E}[XY]$ if either of the random variables has mean zero.

Example 3. Let X and Y be independent random variables. Find their covariance.

$$\mathrm{Cov}(X,Y) = \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big] = \mathbb{E}[X - \mathbb{E}[X]]\,\mathbb{E}[Y - \mathbb{E}[Y]] = 0,$$

where the second equality follows from the fact that X and Y are independent, and the third equality follows from

$$\mathbb{E}[X - \mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[X] = 0.$$

Therefore *pairs of independent random variables have covariance zero.*

Multiplying either X or Y by a large constant increases the covariance, so we normalize it in order to measure correlation on an absolute scale.

Definition 6.1 (Correlation Coefficient)**.** The correlation coefficient of X and Y is defined by

$$\rho_{X,Y} \;=\; \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sigma_X \sigma_Y},$$

where $\sigma_X = \sqrt{\mathrm{Var}(X)}$ and $\sigma_Y = \sqrt{\mathrm{Var}(Y)}$ are the standard deviations of X and Y, respectively.

The correlation coefficient is a number that is at most 1 in magnitude:

$$-1 \;\le\; \rho_{X,Y} \;\le\; 1. \tag{8}$$

To establish the inequalities, we start from the fact that the expected square of a R.V. is nonnegative, which yields the following inequality:

$$\begin{aligned} 0 \;&\le \mathbb{E}\left\{\left(\frac{X - \mathbb{E}[X]}{\sigma_X} \;\pm\; \frac{Y - \mathbb{E}[Y]}{\sigma_Y}\right)^2\right\} \\ &= 1 \;\pm\; 2\rho_{X,Y} \;+\; 1 \\ &= 2(1 \pm \rho_{X,Y}). \end{aligned}$$

The last equation implies Eq. (8).

The extreme values of $\rho_{X,Y}$ are achieved when X and Y are related linearly, $Y = aX + b$; $\rho_{X,Y} = 1$ if $a > 0$ and $\rho_{X,Y} = -1$ if $a < 0$.

R.V.'s X and Y are said to be **uncorrelated** if $\rho_{X,Y} = 0$. If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$, so $\rho_{X,Y} = 0$. Thus if X and Y are independent, then X and Y are uncorrelated. The following example shows that **it is possible for X and Y to be uncorrelated but not independent**.

Example 4. Let Θ be uniformly distributed in the interval $(0, 2\pi)$. Let

$$X = \cos\Theta \qquad \text{and} \qquad Y = \sin\Theta.$$

The point (X,Y) then corresponds to the point on the unit circle specified by the angle Θ. The marginal PDF's of X and Y are arcsine PDF's, which are nonzero in the interval $(-1,1)$. The product of the marginals is nonzero in the square defined by $-1 \le x \le 1$ and $-1 \le y \le 1$, so if X and Y were independent the point (X,Y) would assume all values in this square. This is not the case, so X and Y are dependent.

We now show that X and Y are uncorrelated:

$$\mathbb{E}[XY] \;=\; \mathbb{E}[\sin\Theta\cos\Theta] = \frac{1}{2\pi}\int_0^{2\pi} \sin\phi\cos\phi\, d\phi$$

$$= \frac{1}{4\pi} \int_0^{2\pi} \sin 2\phi \, d\phi = 0.$$

Since $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, then it implies that X and Y are uncorrelated.

Note that if X and Y are jointly Gaussian and $\rho_{X,Y} = 0$ then X and Y are independent random variables.

Definition 6.2 (Conditional PMF)**.** For X and Y discrete random variables, the conditional PMF of Y given $X = x$ is defined by:

$$p_Y(y \mid x) \;=\; P\left(Y = y \mid X = x\right) = \frac{P\left(X = x,\, Y = y\right)}{P\left(X = x\right)} = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

for x such that $P(X = x) > 0$.

We define $p_Y(y \mid x) = 0$ for x s.t. $P(X = x) = 0$. Note that $p_Y(y \mid x)$ is a function of y over the real line, and that $p_Y(y \mid x) > 0$ only for y in a discrete set $\{y_1, y_2, \dots\}$.

Definition 6.3 (Conditional CPF)**.** Suppose Y is a continuous random variable and X is a discrete R.V., then the conditional CDF of Y given $X = x_k$ is defined as

$$F_Y(y \mid x_k) \;=\; \frac{P\left(Y \leq y,\, X = x_k\right)}{P(X = x_k)}, \qquad \text{for } P(X = x_k) > 0.$$

If X is a continuous R.V., then the conditional CDF of Y given $X = x$ is defined as

$$F_Y(y \mid x) = \lim_{h \to 0} F_Y(y \mid x < X \leq x + h).$$

Notice that

$$\begin{aligned} F_Y(y \mid x < X \leq x + h) &= \frac{P[Y \leq y,\, x < X \leq x + h]}{P[x < X \leq x + h]} \\ &= \frac{\int_{-\infty}^{y} \int_{x}^{x+h} f_{X,Y}(x', y')\, dx'\, dy'}{\int_{x}^{x+h} f_X(x')\, dx'} \\ &= \frac{\int_{-\infty}^{y} f_{X,Y}(x, y')\, dy'\, h}{f_X(x) h}. \end{aligned}$$

As we let h approach zero,

$$F_Y(y \mid x) = \frac{\int_{-\infty}^{y} f_{X,Y}(x, y')\, dy'}{f_X(x)}. \tag{9}$$

Definition 6.4 (Conditional PDF)**.** The conditional PDF of Y given $X = x_k$, if the derivative exists, is given by

$$f_Y(y \mid x_k) \;=\; \frac{d}{dy} F_Y(y \mid x_k),$$

and the conditional PDF of Y given continuous $X = x$ is then:

$$f_Y(y \mid x) = \frac{d}{dy} F_Y(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}. \tag{10}$$

Definition 6.5 (Conditional Expectation)**.** The conditional expectation of Y given $X = x$ is defined by

$$\mathbb{E}[Y \mid x] = \int_{-\infty}^{\infty} y f_Y(y \mid x)\, dy.$$

In the special case where X and Y are both discrete random variables we have:

$$\mathbb{E}[Y \mid x_k] = \sum_{y_j} y_j \, p_Y(y_j \mid x_k). \tag{11}$$

Clearly, $\mathbb{E}[Y \mid x]$ is simply the center of mass associated with the conditional PDF or PMF.

Remark 6.2. *The conditional expectation $\mathbb{E}[Y \mid x]$ can be viewed as defining a function of x: $g(x) = \mathbb{E}[Y \mid x]$. It therefore makes sense to talk about the random variable $g(X) = \mathbb{E}[Y \mid X]$. We can imagine that a random experiment is performed and a value for X is obtained, say $X = x_0$, and then the value $g(x_0) = \mathbb{E}[Y \mid x_0]$ is produced. We are interested in $\mathbb{E}[g(X)] = \mathbb{E}[\mathbb{E}[Y \mid X]]$. In particular, we have*

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]], \tag{12}$$

where the RHS is

$$\mathbb{E}[\mathbb{E}[Y \mid X]] = \int_{-\infty}^{\infty} \mathbb{E}[Y \mid x] f_X(x) \, dx, \tag{13}$$

$$\mathbb{E}[\mathbb{E}[Y \mid X]] = \sum_{x_k} \mathbb{E}[Y \mid x_k] p_X(x_k) \tag{14}$$

If X and Y are jointly continuous random variables. Then

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[Y \mid X]] &= \int_{-\infty}^{\infty} \mathbb{E}[Y \mid x] f_X(x) \, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_Y(y \mid x) \, dy \, f_X(x) \, dx \\
&= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx \, dy \\
&= \int_{-\infty}^{\infty} y f_Y(y) \, dy \\
&= \mathbb{E}[Y].
\end{aligned}$$

The above result also holds for the expected value of a function of Y:

$$\mathbb{E}[h(Y)] = \mathbb{E}[\mathbb{E}[h(Y) \mid X]].$$

In particular, the kth moment of Y is given by

$$\mathbb{E}[Y^k] = \mathbb{E}[\mathbb{E}[Y^k \mid X]].$$

7 Multiple R.V.'s

Random vector $\mathbf{X} = [X_1, \ldots, X_n]$

Definition 7.1 (Joint CDF)**.** The joint CDF of $X_1, \ldots, X_n$ is defined as the probability of an n-dimensional semi-infinite rectangle associated with the point $(x_1, \ldots, x_n)$:

$$F_{\mathbf{X}}(\mathbf{x}) \triangleq F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P\left(X_1 \le x_1,\ X_2 \le x_2,\ \ldots,\ X_n \le x_n\right).$$

Definition 7.2 (Joint PMF)**.** The joint PMF of n discrete random variables is defined by

$$p_{\mathbf{X}}(\mathbf{x}) \ \triangleq\ p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P\left(X_1 = x_1,\ X_2 = x_2,\ \ldots,\ X_n = x_n\right).$$

A family of **conditional PMF's** is obtained from the joint PMF by conditioning on different subcollections of the random variables. For example, if $p_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1}) > 0$, then

$$p_{X_n}(x_n \mid x_1,\ldots,x_{n-1}) = \frac{p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{p_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1})}.$$

Repeated applications yield the following very useful expression (chain rule):

$$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = p_{X_n}(x_n \mid x_1,\ldots,x_{n-1})\, p_{X_{n-1}}(x_{n-1} \mid x_1,\ldots,x_{n-2}) \cdots p_{X_2}(x_2 \mid x_1)\, p_{X_1}(x_1).$$

Random variables $X_1, X_2, \ldots, X_n$ are **jointly continuous random variables** if the probability of any n-dimensional event A is given by an n-dimensional integral of a probability density function:

$$P(\mathbf{X} \in A) = \int_{\mathbf{x}\in A}\int \cdots \int f_{X_1,\ldots,X_n}(x'_1,\ldots,x'_n)\, dx'_1 \cdots dx'_n,$$

where $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ is the **joint PDF**.

The joint CDF of $\mathbf{X}$ is obtained from the joint PDF by integration:

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1,\ldots,X_n}(x'_1,\ldots,x'_n)\, dx'_1 \cdots dx'_n.$$

The joint PDF (if the derivative exists) is given by

$$f_{\mathbf{X}}(\mathbf{x}) \triangleq f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n).$$

The marginal PDF for a subset of the random variables is obtained by integrating the other variables out. For example, the marginal pdf of X_1 is

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, x'_2,\ldots,x'_n)\, dx'_2 \cdots dx'_n.$$

As another example, the marginal PDF for $X_1,\ldots,X_{n-1}$ is given by

$$f_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1}) = \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1,\ldots,x_{n-1},x'_n)\, dx'_n.$$

A family of **conditional PDF's** is also associated with the joint pdf. For example, the PDF of X_n given the values of $X_1,\ldots,X_{n-1}$ is

$$f_{X_n}(x_n \mid x_1,\ldots,x_{n-1}) = \frac{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{f_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1})}.$$

if $f_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1}) > 0$.

Repeated such applications yield an expression:

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_n}(x_n \mid x_1,\ldots,x_{n-1}) f_{X_{n-1}}(x_{n-1} \mid x_1,\ldots,x_{n-2}) \cdots f_{X_2}(x_2 \mid x_1)\, f_{X_1}(x_1).$$

Independence As shown in Table 1, independence means the joint distribution factorizes into the product of the marginals.

Case	Independence Condition
General (CDF)	$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = F_{X_1}(x_1)\cdots F_{X_n}(x_n)$
Discrete (PMF)	$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = p_{X_1}(x_1)\cdots p_{X_n}(x_n)$
Continuous (PDF)	$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n)$

Table 1: Equivalent formulations of independence for random variables

8 Functions of Several R.V.'s and Their Expected Value

For $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, the **mean vector** is defined as the column vector of expected values of the components X_k:

$$\mathbf{m}_X = \mathbb{E}[\mathbf{X}] = \mathbb{E}\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \triangleq \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix}.$$

The **correlation matrix** has the second moments of $\mathbf{X}$ as its entries:

$$\mathbf{R}_{\mathbf{X}} = \begin{bmatrix} \mathbb{E}\left[X_1^2\right] & \mathbb{E}[X_1X_2] & \cdots & \mathbb{E}[X_1X_n] \\ \mathbb{E}[X_2X_1] & \mathbb{E}\left[X_2^2\right] & \cdots & \mathbb{E}[X_2X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[X_nX_1] & \mathbb{E}[X_nX_2] & \cdots & \mathbb{E}\left[X_n^2\right] \end{bmatrix}.$$

The **covariance matrix** of $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is defined as

$$\mathrm{Cov}(\mathbf{X}) = \mathbb{E}\left[(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^\top\right].$$

In expanded form,

$$\mathbf{K}_{\mathbf{X}} = \mathrm{Cov}(\mathbf{X}) = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Var}(X_n) \end{bmatrix}.$$

Both $\mathbf{R}_{\mathbf{X}}$ and $\mathbf{K}_{\mathbf{X}}$ are $n \times n$ symmetric matrices. The diagonal elements of $\mathbf{K}_{\mathbf{X}}$ are given by the variances $\mathrm{Var}[X_k] = \mathbb{E}\left[(X_k - m_k)^2\right]$ of the components of $\mathbf{X}$. If these elements are uncorrelated, then $\mathrm{Cov}(X_j, X_k) = 0$ for $j \neq k$, and $\mathbf{K}_{\mathbf{X}}$ is diagonal. If the random variables $X_1, \ldots, X_n$ are independent, then they are uncorrelated and $\mathbf{K}_{\mathbf{X}}$ is diagonal. Finally, if the vector of expected values is $\mathbf{0}$, that is, $m_k = \mathbb{E}[X_k] = 0$ for all k, then

$$\mathbf{R}_{\mathbf{X}} = \mathbf{K}_{\mathbf{X}}.$$