

4 Joint Distributions

4.1 Joint Probability Mass Function

Definition 1. The joint probability mass function of two discrete random variables X and Y is defined as

$$P_{XY}(x, y) = P(X = x, Y = y).$$

$$\sum_{(x_i, y_j) \in R_{XY}} P_{XY}(x_i, y_j) = 1$$

$$P((X, Y) \in A) = \sum_{(x_i, y_j) \in (A \cap R_{XY})} P_{XY}(x_i, y_j)$$

Example 2. An analyst is studying the traffic accidents in two adjacent towns. The random variable X represents the number of accidents in a day in town A, and the random variable Y represents the number of accidents in a day in town B. The joint probability function for X and Y is given by

$$p(x, y) = \frac{e^{-2}}{x!y!} \text{ for } x = 0, 1, 2, \dots \text{ and } y = 0, 1, 2, \dots$$

The probability that on a given day there will be 1 accident in town A and 2 accidents in town B is

$$p(1, 2) = \frac{e^{-2}}{1!2!} \approx 0.068.$$

Definition 3. Marginal PMFs of X and Y :

$$\begin{aligned} P_X(x) &= \sum_{y_j \in R_Y} P_{XY}(x, y_j), & \text{for any } x \in R_X \\ P_Y(y) &= \sum_{x_i \in R_X} P_{XY}(x_i, y), & \text{for any } y \in R_Y \end{aligned} \tag{1}$$

Example 4. The table of joint probabilities for the asset values is the following

y/x	90	100	110
0	.05	.27	.18
10	.15	.33	.02

The probability that X is 90 can be found by adding all joint probabilities in the first column of the table above.

$$P(X = 90) = P(X = 90, Y = 0) + P(X = 90, Y = 10) = .05 + .15 = .20$$

$$P(y = 0) = P(X = 90, Y = 0) + P(X = 100, Y = 0) + P(X = 110, Y = 0) = .05 + .27 + .18 = .50$$

4.1.1 Joint Cumulative Distributive Function (CDF)

Definition 5. The joint cumulative distribution function of two random variables X and Y is defined as

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) = P((X \leq x) \cap (Y \leq y)).$$

Marginal CDFs of X and Y :

$$\begin{aligned} F_X(x) &= F_{XY}(x, \infty) = \lim_{y \rightarrow \infty} F_{XY}(x, y), & \text{for any } x, \\ F_Y(y) &= F_{XY}(\infty, y) = \lim_{x \rightarrow \infty} F_{XY}(x, y), & \text{for any } y \end{aligned} \quad (2)$$

Remark 6.

$$\begin{aligned} F_{XY}(\infty, \infty) &= 1, \\ F_{XY}(-\infty, y) &= 0, & \text{for any } y, \\ F_{XY}(x, -\infty) &= 0, & \text{for any } x. \end{aligned}$$

Theorem 7. For two random variables X and Y , and real numbers $x_1 \geq x_2, y_1 \geq y_2$, we have

$$\begin{aligned} P(x_1 < X \leq x_2, y_1 < Y \leq y_2) &= \\ F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1). \end{aligned}$$

4.1.2 Conditioning and Independence

We have discussed conditional probability before, and you have already seen some problems regarding random variables and conditional probability. Here, we will discuss conditioning for random variables more in detail and introduce the conditional PMF, conditional CDF, and conditional expectation. We would like to emphasize that there is only one main formula regarding conditional probability which is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ when } P(B) > 0.$$

Any other formula regarding conditional probability can be derived from the above formula. Specifically, if you have two random variables X and Y , you can write

$$P(X \in C | Y \in D) = \frac{P(X \in C, Y \in D)}{P(Y \in D)}, \text{ where } C, D \subset \mathbb{R}.$$

Conditional PMF and CDF:

Remember that the PMF is by definition a probability measure, i.e., it is $P(X = x_k)$. Thus, we can talk about the conditional PMF. Specifically, the conditional PMF of X given event A , is defined as

$$\begin{aligned} P_{X|A}(x_i) &= P(X = x_i|A) \\ &= \frac{P(X = x_i \text{ and } A)}{P(A)}. \end{aligned}$$

Example 8. *I roll a fair die. Let X be the observed number. Find the conditional PMF of X given that we know the observed number was less than 5. Here, we condition on the event $A = \{X < 5\}$, where $P(A) = \frac{4}{6}$. Thus,*

$$\begin{aligned} P_{X|A}(1) &= P(X = 1|X < 5) \\ &= \frac{P(X = 1 \text{ and } X < 5)}{P(X < 5)} \\ &= \frac{P(X = 1)}{P(X < 5)} = \frac{1}{4}. \end{aligned}$$

$$P_{X|A}(2) = P_{X|A}(3) = P_{X|A}(4) = \frac{1}{4}.$$

$$P_{X|A}(5) = P_{X|A}(6) = 0.$$

Definition 9. *For a discrete random variable X and event A , the conditional PMF of X given A is defined as*

$$\begin{aligned} P_{X|A}(x_i) &= P(X = x_i|A) \\ &= \frac{P(X = x_i \text{ and } A)}{P(A)}, \quad \text{for any } x_i \in R_X. \end{aligned}$$

Similarly, we define the conditional CDF of X given A as

$$F_{X|A}(x) = P(X \leq x|A).$$

Definition 10. *For discrete random variables X and Y , the conditional PMFs of X given Y and vice versa are defined as*

$$\begin{aligned} P_{X|Y}(x_i|y_j) &= \frac{P_{XY}(x_i, y_j)}{P_Y(y_j)}, \\ P_{Y|X}(y_j|x_i) &= \frac{P_{XY}(x_i, y_j)}{P_X(x_i)} \end{aligned}$$

for any $x_i \in R_X$ and $y_j \in R_Y$.

Example 11. The joint probability function for the two assets is given below

y/x	90	100	110	$p_Y(y)$
0	.05	.27	.18	.50
10	.15	.33	.02	.50
$p_X(x)$.20	.60	.20	

Suppose we are given that $Y=0$. Then we can compute conditional probabilities for X based on this information.

$$P(X = 90|Y = 0) = \frac{P(X = 90, Y = 0)}{P(Y = 0)} = \frac{.05}{.50} = .10$$

$$P(X = 100|Y = 0) = \frac{P(X = 100, Y = 0)}{P(Y = 0)} = \frac{.27}{.50} = .54$$

$$P(X = 110|Y = 0) = \frac{P(X = 110, Y = 0)}{P(Y = 0)} = \frac{.18}{.50} = .36$$

These values give a complete probability function $p(x|Y = 0)$ for X , given the information that $Y=0$

x	90	100	110
$p(x 0)$.10	.54	.36

Definition 12. Two discrete random variables X and Y are independent if

$$P_{XY}(x, y) = P_X(x)P_Y(y), \quad \text{for all } x, y.$$

Equivalently, X and Y are independent if

$$F_{XY}(x, y) = F_X(x)F_Y(y), \quad \text{for all } x, y.$$

Example 13. The joint probability function for the two assets is given below

y/x	90	100	110	$p_Y(y)$
0	.05	.27	.18	.50
10	.15	.33	.02	.50
$p_X(x)$.20	.60	.20	

Note that $p(90, 0) = .05$ and $p_X(90) \cdot p_Y(0) = (.10)$. The random variables X and Y are not independent.

Definition 14 (The Multinomial Distribution: Counting partition). The number of partitions of n objects into k distinct groups of size $n_1, n_2, n_3, \dots, n_k$ is given by

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1!n_2! \cdots n_k!}$$

Suppose that a random experiment has k mutually exclusive outcomes E_1, \dots, E_k , with $P(E_i) = p_i$. Suppose that you repeat this experiment in n independent trials. Let X_i be the number of times that the outcome E_i occurs in the trials. Then

$$P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = \binom{n}{n_1, n_2, \dots, n_k} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

Example 15. You are spinning a spinner that can land on three colors-red, blue and yellow. For this spinner $P(\text{red}) = .4$, $P(\text{blue}) = .35$ and $P(\text{yellow}) = .25$, you spin the spinner 10 times. What is the probability that you spin red five times, blue three times and yellow two times?

Solution: There are $k=3$ mutually exclusive outcomes. Let X_1, X_2 and X_3 be the number of times the spinner comes up red, blue and yellow respectively. Then $p_1 = P(X_1) = .4, p_2 = P(X_2) = .35$ and $p_3 = P(X_3) = .25$. We need to find

$$P(X_1 = 5, X_2 = 3, X_3 = 2) = \binom{10}{5, 3, 2} (0.4)^5 (.35)^3 (.25)^2 = .069$$

Example 16 (The sum of two discrete random variables). We look at the two asset random variables X and Y

y/x	90	100	110
0	.05	.27	.18
10	.15	.33	.02

Probabilities for the sum $S = X + Y$ can be found by direct inspection.

$$P(X + Y = 90) = p(90, 0) = 0.05$$

$$P(X + Y = 100) = p(90, 10) + p(100, 0) = .27 + .15 = 0.42$$

$$P(X + Y = 110) = p(110, 0) + p(100, 10) = .18 + .33 = 0.51$$

$$P(X + Y = 120) = p(110, 10) = .2$$

we found the entire distribution of $S=X+Y$.

s	90	100	110	120
$p(s)$	0.05	.42	.51	.02

so

$$p(s) = \sum_x p(x, s - x)$$

Definition 17 (Probability function for $S=X+Y$ (X and Y are independent)).

$$p(s) = \sum_x p(x, s - x) = p_X(x) \cdot p_Y(s - x)$$

Definition 18.

$$E[X|A] = \sum_{x_i \in R_X} x_i P_{X|A}(x_i),$$

$$E[X|Y = y_j] = \sum_{x_i \in R_X} x_i P_{X|Y}(x_i|y_j)$$

Definition 19. Law of Total Probability:

$$P(X \in A) = \sum_{y_j \in R_Y} P(X \in A|Y = y_j)P_Y(y_j), \quad \text{for any set } A.$$

Law of Total Expectation:

1. If B_1, B_2, B_3, \dots is a partition of the sample space S ,

$$E[X] = \sum_i E[X|B_i]P(B_i) \quad (3)$$

2. For a random variable X and a discrete random variable Y ,

$$E[X] = \sum_{y_j \in R_Y} E[X|Y = y_j]P_Y(y_j) \quad (4)$$

4.1.3 Functions of Two Random Variables

Definition 20. *Law of the unconscious statistician (LOTUS) for two discrete random variables:*

$$E[g(X, Y)] = \sum_{(x_i, y_j) \in R_{XY}} g(x_i, y_j)P_{XY}(x_i, y_j) \quad (5.5) \quad (5)$$

4.1.4 Conditional Expectation (Revisited) and Conditional Variance

Remember that the conditional expectation of X given that $Y = y$ is given by

Definition 21.

$$E[X|Y = y] = \sum_{x_i \in R_X} x_i P_{X|Y}(x_i|y).$$

Note that $E[X|Y = y]$ depends on the value of y . In other words, by changing y , $E[X|Y = y]$ can also change. Thus, we can say $E[X|Y = y]$ is a function of y , so let's write

$$g(y) = E[X|Y = y].$$

$$E[g(X)h(Y)|X] = g(X)E[h(Y)|X] \quad (6)$$

Law of Iterated Expectations: $E[X] = E[E[X|Y]]$

Remark 22. If X and Y are independent random variables, then

1. $E[X|Y] = E[X]$
2. $E[g(X)|Y] = E[g(X)]$
3. $E[XY] = E[X]E[Y]$
4. $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

Definition 23.

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y]) \quad (7)$$

To describe the law of total variance intuitively, it is often useful to look at a population divided into several groups. In particular, suppose that we have this random experiment: We pick a person in the world at random and look at his/her height. Let's call the resulting value X . Define another random variable Y whose value depends on the country of the chosen person, where $Y=1,2,3,\dots,n$, and n is the number of countries in the world. Then, let's look at the two terms in the law of total variance.

$$\text{Var}(X) = E(\text{Var}(X|Y)) + \text{Var}(E[X|Y]).$$

Note that $\text{Var}(X|Y = i)$ is the variance of X in country i . Thus, $E(\text{Var}(X|Y))$ is the average of variances in each country. On the other hand, $E[X|Y = i]$ is the average height in country i . Thus, $\text{Var}(E[X|Y])$ is the variance between countries. So, we can interpret the law of total variance in the following way. Variance of X can be decomposed into two parts: the first is the average of variances in each individual country, while the second is the variance between height averages in each country.

4.1.5 Joint Probability Density Function (PDF)

Definition 24. Two random variables X and Y are jointly continuous if there exists a nonnegative function $f_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$, such that, for any set $A \in \mathbb{R}^{\neq}$, we have

$$P((X,Y) \in A) = \iint_A f_{XY}(x,y) dx dy \quad (8)$$

The function $f_{XY}(x,y)$ is called the joint probability density function (PDF) of X and Y . We may define the range of (X,Y) as

$$R_{XY} = \{(x,y) | f_{X,Y}(x,y) > 0\}.$$

If we choose $A = \mathbb{R}^{\neq}$, then the probability of $(X,Y) \in A$ must be one, so we must have

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y) dx dy = 1$$

Definition 25. Marginal PDFs:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y) dy, \quad \text{for all } x, \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y) dx, \quad \text{for all } y. \end{aligned}$$

4.1.6 Joint Cumulative Distribution Function

Remark 26.

$$F_{XY}(x, y) = P(X \leq x, Y \leq y).$$

The joint CDF satisfies the following properties:

1. $F_X(x) = F_{XY}(x, \infty)$, for any x (marginal CDF of X);
2. $F_Y(y) = F_{XY}(\infty, y)$, for any y (marginal CDF of Y);
3. $F_{XY}(\infty, \infty) = 1$
4. $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0$
5. $P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1)$
6. if X and Y are independent, then $F_{XY}(x, y) = F_X(x)F_Y(y)$.

4.1.7 Conditioning and Independence

Remark 27. If X is a continuous random variable, and A is the event that $a < X < b$ (where possibly $b = \infty$ or $a = -\infty$), then

$$F_{X|A}(x) = \begin{cases} 1 & x > b \\ \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} & a \leq x < b \\ 0 & x < a \end{cases}$$

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{P(A)} & a \leq x < b \\ 0 & \text{otherwise} \end{cases}$$

$$E[X|A] = \int_{-\infty}^{\infty} x f_{X|A}(x) dx,$$

$$E[g(X)|A] = \int_{-\infty}^{\infty} g(x) f_{X|A}(x) dx,$$

$$\text{Var}(X|A) = E[X^2|A] - (E[X|A])^2$$

Conditioning by Another Random Variable:

Definition 28. For two jointly continuous random variables X and Y , we can define the following conditional concepts:

1. The conditional PDF of X given $Y=y$:

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

2. The conditional probability that $X \in A$ given $Y = y$:

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

3. The conditional CDF of X given $Y = y$:

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \int_{-\infty}^x f_{X|Y}(x|y)dx$$

Definition 29. For two jointly continuous random variables X and Y , we have:

1. Expected value of X given $Y = y$:

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

2. Conditional LOTUS:

$$E[g(X)|Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)dx$$

3. Conditional variance of X given $Y = y$:

$$\text{Var}(X|Y = y) = E[X^2|Y = y] - (E[X|Y = y])^2$$

Independent Random Variables

Remark 30. Two continuous random variables X and Y are independent if

1. Two continuous random variables X and Y are independent if

$$f_{XY}(x, y) = f_X(x)f_Y(y), \quad \text{for all } x, y.$$

2. Equivalently, X and Y are independent if

$$F_{XY}(x, y) = F_X(x)F_Y(y), \quad \text{for all } x, y.$$

3. If X and Y are independent, we have

$$\begin{aligned} E[XY] &= EXEY, \\ E[g(X)h(Y)] &= E[g(X)]E[h(Y)]. \end{aligned}$$

Definition 31. 1. Law of Total Probability:

$$P(A) = \int_{-\infty}^{\infty} P(A|X = x)f_X(x) dx \quad (5.16) \quad (9)$$

2. Law of Total Expectation:

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} E[Y|X = x]f_X(x) dx \\ &= E[E[Y|X]] \end{aligned} \quad (10)$$

3. Law of Total Variance:

$$\text{Var}(Y) = E[\text{Var}(Y|X)] + \text{Var}(E[Y|X]) \quad (11)$$

4.1.8 Functions of Two Continuous Random Variables

Definition 32. *LOTUS for two continuous random variables:*

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) \, dx dy \quad (5.19) \quad (12)$$

The Method of Transformations

Theorem 33. *Let X and Y be two jointly continuous random variables. Let $(Z, W) = g(X, Y) = (g_1(X, Y), g_2(X, Y))$, where $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a continuous one-to-one (invertible) function with continuous partial derivatives. Let $h = g^{-1}$, i.e., $(X, Y) = h(Z, W) = (h_1(Z, W), h_2(Z, W))$. Then Z and W are jointly continuous and their joint PDF, $f_{ZW}(z, w)$, for $(z, w) \in R_{ZW}$ is given by*

$$f_{ZW}(z, w) = f_{XY}(h_1(z, w), h_2(z, w)) |J|,$$

where J is the Jacobian of h defined by

$$J = \det \begin{bmatrix} \frac{\partial h_1}{\partial z} & \frac{\partial h_1}{\partial w} \\ \frac{\partial h_2}{\partial z} & \frac{\partial h_2}{\partial w} \end{bmatrix} = \frac{\partial h_1}{\partial z} \cdot \frac{\partial h_2}{\partial w} - \frac{\partial h_2}{\partial z} \frac{\partial h_1}{\partial w}.$$

Theorem 34. *If X and Y are two jointly continuous random variables and $Z = X + Y$, then*

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(w, z - w) dw = \int_{-\infty}^{\infty} f_{XY}(z - w, w) dw.$$

If X and Y are also independent, then

$$\begin{aligned} f_Z(z) &= f_X(z) * f_Y(z) \\ &= \int_{-\infty}^{\infty} f_X(w) f_Y(z - w) dw = \int_{-\infty}^{\infty} f_Y(w) f_X(z - w) dw. \end{aligned}$$

Theorem 35. *If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are independent, then*

$$X + Y \sim N\left(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2\right).$$

4.1.9 Covariance and Correlation

Definition 36. *The covariance between X and Y is defined as*

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)] = E[XY] - (EX)(EY).$$

Remark 37. *The covariance has the following properties:*

1. $\text{Cov}(X, X) = \text{Var}(X)$;

2. if X and Y are independent then $Cov(X, Y) = 0$;
3. $Cov(X, Y) = Cov(Y, X)$;
4. $Cov(aX, Y) = aCov(X, Y)$;
5. $Cov(X + c, Y) = Cov(X, Y)$;
6. $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$;
7. more generally,

$$Cov\left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j\right) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j Cov(X_i, Y_j).$$

Definition 38. *Variance of a sum:*

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y) \quad (13)$$

Definition 39. *The correlation coefficient, denoted by ρ_{XY} or $\rho(X, Y)$, is obtained by normalizing the covariance. In particular, we define the correlation coefficient of two random variables X and Y as the covariance of the standardized versions of X and Y . Define the standardized versions of X and Y as*

$$U = \frac{X - EX}{\sigma_X}, \quad V = \frac{Y - EY}{\sigma_Y} \quad (5.22) \quad (14)$$

$$\rho_{XY} = \rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) Var(Y)}} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Remark 40. 1. $1 \leq \rho(X, Y) \leq 1$;

2. if $\rho(X, Y) = 1$, then $Y = aX + b$, where $a > 0$;
3. if $\rho(X, Y) = -1$, then $Y = aX + b$, where $a < 0$;
4. $\rho(aX + b, cY + d) = \rho(X, Y)$ for $a, c > 0$.
5. If $\rho(X, Y) = 0$, we say that X and Y are uncorrelated.
6. If $\rho(X, Y) > 0$, we say that X and Y are positively correlated.
7. If $\rho(X, Y) < 0$, we say that X and Y are negatively correlated.

Theorem 41. *If X and Y are uncorrelated, then*

$$Var(X + Y) = Var(X) + Var(Y).$$

More generally, if X_1, X_2, \dots, X_n are pairwise uncorrelated, i.e., $\rho(X_i, X_j) = 0$ when $i \neq j$, then

$$Var(X_1 + X_2 + \dots + X_n) = Var(X_1) + Var(X_2) + \dots + Var(X_n).$$

4.1.10 Bivariate Normal Distribution

Definition 42. Two random variables X and Y are said to be bivariate normal, or jointly normal, if $aX + bY$ has a normal distribution for all $a, b \in R$.

Definition 43. Two random variables X and Y are said to have the standard bivariate normal distribution with correlation coefficient ρ if their joint PDF is given by

$$f_{XY}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}[x^2 - 2\rho xy + y^2]\right\},$$

where $\rho \in (-1, 1)$. If $\rho = 0$, then we just say X and Y have the standard bivariate normal distribution.

Definition 44. Suppose X and Y are jointly normal random variables with parameters μ_X , σ_X^2 , μ_Y , σ_Y^2 , and ρ . Then, given $X=x$, Y is normally distributed with

$$\begin{aligned} E[Y|X = x] &= \mu_Y + \rho\sigma_Y \frac{x - \mu_X}{\sigma_X}, \\ \text{Var}(Y|X = x) &= (1 - \rho^2)\sigma_Y^2. \end{aligned}$$

Theorem 45. If X and Y are bivariate normal and uncorrelated, then they are independent.