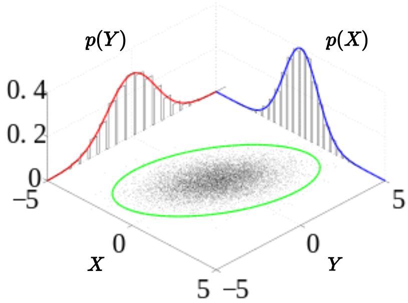
WIKIPEDIA

Joint probability distribution

Given random variables $X, Y, \ldots$, that are defined on a probability space, the **joint probability distribution** for $X, Y, \ldots$ is a probability distribution that gives the probability that each of $X, Y, \ldots$ falls in any particular range or discrete set of values specified for that variable. In the case of only two random variables, this is called a **bivariate distribution**, but the concept generalizes to any number of random variables, giving a **multivariate distribution**.



Many sample observations (black) are shown from a joint probability distribution. The marginal densities are shown as well.

The joint probability distribution can be expressed either in terms of a joint cumulative distribution function or in terms of a joint probability density function (in the case of continuous variables) or joint probability mass function (in the case of discrete variables). These in turn can be used to find two other types of distributions: the marginal distribution giving the probabilities for any one of the variables with no reference to any specific ranges of values for the other variables, and the conditional probability distribution giving the probabilities for any subset of the variables conditional on particular values of the remaining variables.

Contents

Examples

Draws from an urn

Suppose each of two urns contains twice as many red balls as blue balls, and no others, and suppose one ball is randomly selected from each urn, with the two draws independent of each other. Let A and B be discrete random variables associated with the outcomes of the draw from the first urn and second urn respectively. The probability of drawing a red ball from either of the urns is 2/3, and the probability of drawing a blue ball is 1/3. We can present the joint probability distribution as the following table:

	A=Red	A=Blue	P(B)
B=Red	(2/3)(2/3)=4/9	(1/3)(2/3)=2/9	4/9+2/9=2/3
B=Blue	(2/3)(1/3)=2/9	(1/3)(1/3)=1/9	2/9+1/9=1/3
P(A)	4/9+2/9=2/3	2/9+1/9=1/3	

Each of the four inner cells shows the probability of a particular combination of results from the two draws; these probabilities are the joint distribution. In any one cell the probability of a particular combination occurring is (since the draws are independent) the product of the probability of the specified result for A and the probability of the specified result for B. The probabilities in these four cells sum to 1, as it is always true for probability distributions.

Moreover, the final row and the final column give the marginal probability distribution for A and the marginal probability distribution for B respectively. For example, for A the first of these cells gives the sum of the probabilities for A being red, regardless of which possibility for B in the column above the cell occurs, as 2/3. Thus the marginal probability distribution for A gives A's probabilities *unconditional* on B, in a margin of the table.

Coin flips

Consider the flip of two fair coins; let A and B be discrete random variables associated with the outcomes of the first and second coin flips respectively. Each coin flip is a Bernoulli trial and has a Bernoulli distribution. If a coin displays "heads" then the associated random variable takes the value 1, and it takes the value 0 otherwise. The probability of each of these outcomes is 1/2, so the marginal (unconditional) density functions are

$$P(A) = 1/2 \quad \text{for} \quad A \in \{0, 1\};$$
$$P(B) = 1/2 \quad \text{for} \quad B \in \{0, 1\}.$$

The joint probability mass function of A and B defines probabilities for each pair of outcomes. All possible outcomes are

$$(A=0,B=0),(A=0,B=1),(A=1,B=0),(A=1,B=1).$$

Since each outcome is equally likely the joint probability mass function becomes

$$P(A,B)=1/4 \quad \text{for} \quad A,B\in\{0,1\}.$$

Since the coin flips are independent, the joint probability mass function is the product of the marginals:

$$P(A,B)=P(A)P(B) \quad \text{for} \quad A,B\in\{0,1\}.$$

Rolling a dice

Consider the roll of a fair die and let $A=1$ if the number is even (i.e. 2, 4, or 6) and $A=0$ otherwise. Furthermore, let $B=1$ if the number is prime (i.e. 2, 3, or 5) and $B=0$ otherwise.

	1	2	3	4	5	6
A	0	1	0	1	0	1
B	0	1	1	0	1	0

Then, the joint distribution of A and B, expressed as a probability mass function, is

$$\mathrm{P}(A=0,B=0)=P\{1\}=\frac{1}{6}, \qquad \mathrm{P}(A=1,B=0)=P\{4,6\}=\frac{2}{6},$$
$$\mathrm{P}(A=0,B=1)=P\{3,5\}=\frac{2}{6}, \qquad \mathrm{P}(A=1,B=1)=P\{2\}=\frac{1}{6}.$$

These probabilities necessarily sum to 1, since the probability of *some* combination of A and B occurring is 1.

Real life example:

Consider a production facility that fills plastic bottles with laundry detergent. The weight of each bottle (Y) and the volume of laundry detergent it contains (X) are measured.

Marginal probability distribution

If more than one random variable is defined in a random experiment, it is important to distinguish between the joint probability distribution of X and Y and the probability distribution of each variable individually. The individual probability distribution of a random variable is referred to as its marginal probability distribution. In general, the marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables.

If the joint probability density function of random variable X and Y is $f_{X,Y}(x,y)$, the marginal probability density function of X and Y are:

$$f_X(x)=\int f_{XY}(x,y)\;dy\,,\; f_Y(y)=\int f_{XY}(x,y)\;dx$$

where the first integral is over all points in the range of (X,Y) for which X=x and the second integral is over all points in the range of (X,Y) for which Y=y.[1]

Joint cumulative distribution function

For a pair of random variables X,Y, the joint cumulative distribution function (CDF) F_{XY} is given by[2]:p. 89

$$F_{X,Y}(x,y)=\mathrm{P}(X\leq x,Y\leq y) \qquad \text{(Eq. 1)}$$

where the right-hand side represents the probability that the random variable X takes on a value less than or equal to x **and** that Y takes on a value less than or equal to y.

For N random variables $X_1,\ldots,X_N$, the joint CDF $F_{X_1,\ldots,X_N}$ is given by

$$F_{X_1,\ldots,X_N}(x_1,\ldots,x_N)=\mathrm{P}(X_1\leq x_1,\ldots,X_N\leq x_N) \qquad \text{(Eq. 2)}$$

Interpreting the N random variables as a random vector $\mathbf{X}=(X_1,\ldots,X_N)^T$ yields a shorter notation:

$$F_{\mathbf{X}}(\mathbf{x})=\mathrm{P}(X_1\leq x_1,\ldots,X_N\leq x_N)$$

Joint density function or mass function

Discrete case

The joint probability mass function of two discrete random variables X,Y is:

$$p_{X,Y}(x,y) = \mathrm{P}(X = x \text{ and } Y = y) \qquad \text{(Eq. 3)}$$

or written in terms of conditional distributions

$$p_{X,Y}(x,y) = \mathrm{P}(Y = y \mid X = x) \cdot \mathrm{P}(X = x) = \mathrm{P}(X = x \mid Y = y) \cdot \mathrm{P}(Y = y)$$

where $\mathrm{P}(Y = y \mid X = x)$ is the probability of $Y = y$ given that $X = x$.

The generalization of the preceding two-variable case is the joint probability distribution of n discrete random variables $X_1, X_2, \ldots, X_n$ which is:

$$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \mathrm{P}(X_1 = x_1 \text{ and } \ldots \text{ and } X_n = x_n) \qquad \text{(Eq. 4)}$$

or equivalently

$$\begin{aligned} p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = {} & \mathrm{P}(X_1 = x_1) \cdot \mathrm{P}(X_2 = x_2 \mid X_1 = x_1) \\ & \cdot \mathrm{P}(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2) \\ & \ldots \\ & \cdot P(X_n = x_n \mid X_1 = x_1, X_2 = x_2, \ldots, X_{n-1} = x_{n-1}). \end{aligned}$$

This identity is known as the chain rule of probability.

Since these are probabilities, we have in the two-variable case

$$\sum_i \sum_j \mathrm{P}(X = x_i \text{ and } Y = y_j) = 1,$$

which generalizes for n discrete random variables $X_1, X_2, \ldots, X_n$ to

$$\sum_i \sum_j \ldots \sum_k \mathrm{P}(X_1 = x_{1i}, X_2 = x_{2j}, \ldots, X_n = x_{nk}) = 1.$$

Continuous case

The **joint probability density function** $f_{X,Y}(x,y)$ for two continuous random variables is defined as the derivative of the joint cumulative distribution function (see **Eq.1**):

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \qquad \text{(Eq. 5)}$$

This is equal to:

$$f_{X,Y}(x,y) = f_{Y|X}(y \mid x) f_X(x) = f_{X|Y}(x \mid y) f_Y(y)$$

where $f_{Y|X}(y \mid x)$ and $f_{X|Y}(x \mid y)$ are the conditional distributions of Y given $X = x$ and of X given $Y = y$ respectively, and $f_X(x)$ and $f_Y(y)$ are the marginal distributions for X and Y respectively.

The definition extends naturally to more than two random variables:

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{\partial^n F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{\partial x_1 \ldots \partial x_n} \qquad \text{(Eq. 6)}$$

Again, since these are probability distributions, one has

$$\int_x \int_y f_{X,Y}(x,y) \; dy \; dx = 1$$

respectively

$$\int_{x_1} \ldots \int_{x_n} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) \; dx_n \ldots \; dx_1 = 1$$

Mixed case

The "mixed joint density" may be defined where one or more random variables are continuous and the other random variables are discrete. With one variable of each type we have

One example of a situation in which one may wish to find the cumulative distribution of one random variable which is continuous and another random variable which is discrete arises when one wishes to use a logistic regression in predicting the probability of a binary outcome Y conditional on the value of a continuously distributed outcome X. One *must* use the "mixed" joint density when finding the cumulative distribution of this binary outcome because the input variables were initially defined in such a way that one could not collectively assign it either a probability density function or a probability mass function. Formally, $f_{X,Y}(x,y)$ is the probability density function of with respect to the product measure on the respective supports of X and Y. Either of these two decompositions can then be used to recover the joint cumulative distribution function:

The definition generalizes to a mixture of arbitrary numbers of discrete and continuous random variables.

Additional properties

Joint distribution for independent variables

In general two random variables X and Y are independent if and only if the joint cumulative distribution function satisfies

Two discrete random variables X and Y are independent if and only if the joint probability mass function satisfies

for all x and y.

While the number of independent random events grows, the related joint probability value decreases rapidly to zero, according to a negative exponential law.

Similarly, two absolutely continuous random variables are independent if and only if

for all x and y. This means that acquiring any information about the value of one or more of the random variables leads to a conditional distribution of any other variable that is identical to its unconditional (marginal) distribution; thus no variable provides any information about any other variable.

Joint distribution for conditionally dependent variables

If a subset A of the variables is conditionally dependent given another subset B of these variables, then the probability mass function of the joint distribution is . is equal to . Therefore, it can be efficiently represented by the lower-dimensional probability distributions and . Such conditional independence relations can be represented with a Bayesian network or copula functions.

Covariance

When two or more random variables are defined on a probability space, it is useful to describe how they vary together; that is, it is useful to measure the relationship between the variables. A common measure of the relationship between two random variables is the covariance. Covariance is a measure of linear relationship between the random variables. If the relationship between the random variables is nonlinear, the covariance might not be sensitive to the relationship.

The covariance between the random variable X and Y, denoted as cov(X,Y), is :

[3]

Correlation

There is another measure of the relationship between two random variables that is often easier to interpret than the covariance.

The correlation just scales the covariance by the product of the standard deviation of each variable. Consequently, the correlation is a dimensionless quantity that can be used to compare the linear relationships between pairs of variables in different units. If the points in the joint probability distribution of X and Y that receive positive probability tend to fall along a line of positive (or negative) slope, ρ_{XY} is near +1 (or −1). If ρ_{XY} equals +1 or −1, it can be shown that the points in the joint probability distribution that receive positive probability fall exactly along a straight line. Two random variables with nonzero correlation are said to be correlated. Similar to covariance, the correlation is a measure of the linear relationship between random variables.

The correlation between random variable X and Y, denoted as

Important named distributions

Named joint distributions that arise frequently in statistics include the multivariate normal distribution, the multivariate stable distribution, the multinomial distribution, the negative multinomial distribution, the multivariate hypergeometric distribution, and the elliptical distribution.

See also

- Bayesian programming
- Chow–Liu tree
- Conditional probability
- Copula (probability theory)
- Disintegration theorem
- Multivariate statistics

- Statistical interference
- Pairwise independent distribution

References

1. Montgomery, Douglas C. (19 November 2013). Applied statistics and probability for engineers. Runger, George C. (Sixth ed.). Hoboken, NJ. ISBN 978-1-118-53971-2. OCLC 861273897 (https://www.worldcat.org/oclc/861273897).
2. Park,Kun Il (2018). Fundamentals of Probability and Stochastic Processes with Applications to Communications. Springer. ISBN 978-3-319-68074-3.
3. Montgomery, Douglas C. (19 November 2013). Applied statistics and probability for engineers. Runger, George C. (Sixth ed.). Hoboken, NJ. ISBN 978-1-118-53971-2. OCLC 861273897 (https://www.worldcat.org/oclc/861273897).

External links

- "Joint distribution" (https://www.encyclopediaofmath.org/index.php?title=Joint_distribution), Encyclopedia of Mathematics, EMS Press, 2001 [1994]
- "Multi-dimensional distribution" (https://www.encyclopediaofmath.org/index.php?title=Multi-dimensional_distribution), Encyclopedia of Mathematics, EMS Press, 2001 [1994]
- A modern introduction to probability and statistics : understanding why and how. Dekking, Michel, 1946-. London: Springer. 2005. ISBN 978-1-85233-896-1. OCLC 262680588.
- "Joint continuous density function" (https://planetmath.org/?op=getobj&from=objects&id=576). PlanetMath.
- Mathworld: Joint Distribution Function (http://mathworld.wolfram.com/JointDistributionFunction.html)