WIKIPEDIA

Marginal distribution

In probability theory and statistics, the **marginal distribution** of a subset of a collection of random variables is the probability distribution of the variables contained in the subset. It gives the probabilities of various values of the variables in the subset without reference to the values of the other variables. This contrasts with a conditional distribution, which gives the probabilities contingent upon the values of the other variables.

Marginal variables are those variables in the subset of variables being retained. These concepts are "marginal" because they can be found by summing values in a table along rows or columns, and writing the sum in the margins of the table.[1] The distribution of the marginal variables (the marginal distribution) is obtained by **marginalizing** – that is, focusing on the sums in the margin – over the distribution of the variables being discarded, and the discarded variables are said to have been **marginalized out**.

The context here is that the theoretical studies being undertaken, or the data analysis being done, involves a wider set of random variables but that attention is being limited to a reduced number of those variables. In many applications, an analysis may start with a given collection of random variables, then first extend the set by defining new ones (such as the sum of the original random variables) and finally reduce the number by placing interest in the marginal distribution of a subset (such as the sum). Several different analyses may be done, each treating a different subset of variables as the marginal variables.

Contents

Definition

Marginal probability mass function

Given a known underlined joint distribution of two **discrete** random variables, say, X and Y, the marginal distribution of either variable--X for example—is the probability distribution of X when the values of Y are not taken into consideration. This can be calculated by summing the joint probability distribution over all values of Y. Naturally, the converse is also true: the marginal distribution can be obtained for Y by summing over the separate values of X.

$$p_X(x_i) = \sum_j p(x_i, y_j) \text{ , and } p_Y(y_j) = \sum_i p(x_i, y_j)$$

Y \ X	x_1	x_2	x_3	x_4	$p_Y(y)$ ↓
y_1	4/32	2/32	1/32	1/32	8/32
y_2	3/32	6/32	3/32	3/32	15/32
y_3	9/32	0	0	0	9/32
$p_X(x)$ →	16/32	8/32	4/32	4/32	32/32

Table. 1 Joint and marginal distributions of a pair of discrete random variables, X and Y, dependent, thus having nonzero mutual information I(X; Y). The values of the joint distribution are in the 3×4 rectangle; the values of the marginal distributions are along the right and bottom margins.

A **marginal probability** can always be written as an expected value:

$$p_X(x) = \int_y p_{X|Y}(x \mid y)\, p_Y(y)\, \mathrm{d}y = \mathrm{E}_Y\left[p_{X|Y}(x \mid y)\right].$$

Intuitively, the marginal probability of X is computed by examining the conditional probability of X given a particular value of Y, and then averaging this conditional probability over the distribution of all values of Y.

This follows from the definition of expected value (after applying the law of the unconscious statistician)

$$\mathrm{E}_Y[f(Y)] = \int_y f(y) p_Y(y)\, \mathrm{d}y.$$

Therefore, marginalization provides the rule for the transformation of the probability distribution of a random variable Y and another random variable $X = g(Y)$:

$$p_X(x) = \int_y p_{X|Y}(x \mid y)\, p_Y(y)\, \mathrm{d}y = \int_y \delta\big(x - g(y)\big)\, p_Y(y)\, \mathrm{d}y.$$

Marginal probability density function

Given two **continuous** random variables X and Y whose joint distribution is known, then the marginal probability density function can be obtained by integrating the joint probability distribution, f, over Y, and vice versa. That is

$$f_X(x) = \int_c^d f(x,y)dy, \text{ and } f_Y(y) = \int_a^b f(x,y)dx$$

where $x \in [a,b]$, and $y \in [c,d]$.

Marginal cumulative distribution function

Finding the marginal cumulative distribution function from the joint cumulative distribution function is easy. Recall that

$$F(x,y) = P(X \leq x, Y \leq y) \text{ for } \textbf{discrete} \text{ random variables,}$$

$$F(x,y) = \int_a^x \int_c^y f(x',y')\,dy'dx' \text{ for } \textbf{continuous random variables,}$$

If X and Y jointly take values on [a, b] × [c, d] then

$$F_X(x) = F(x,d) \text{ and } F_Y(y) = F(b,y)$$

If d is ∞, then this becomes a limit $F_X(x) = \lim_{y \to \infty} F(x,y)$. Likewise for $F_Y(y)$.

Marginal distribution vs. conditional distribution

Definition

The **marginal probability** is the probability of a single event occurring, independent of other events. A **conditional probability**, on the other hand, is the probability that an event occurs given that another specific event *has already* occurred. This means that the calculation for one variable is dependent on another variable.[2]

The conditional distribution of a variable given another variable is the joint distribution of both variables divided by the marginal distribution of the other variable.[3] That is,

$$p_{Y|X}(y|x) = P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P_X(x)} \text{ for } \textbf{discrete random variables,}$$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \text{ for } \textbf{continuous random variables.}$$

Example

Suppose there is data from classroom of 200 students on the amount of time studied (X) and the percent correct (Y).[4] Assuming that X and Y are discrete random variables, the joint distribution of X and Y can be described by listing all the possible values of $p(x_i,y_j)$, as shown in Table.3.

Y \ X		Time studied (minutes)				
		x_1 (0–20)	x_2 (21–40)	x_3 (41–60)	x_4 (>60)	$p_Y(y)$ ↓
% correct	y_1 (0–20)	2/200	0	0	8/200	10/200
	y_2 (21–40)	10/200	2/200	8/200	0	20/200
	y_3 (41–59)	2/200	4/200	32/200	32/200	70/200
	y_4 (60–79)	0	20/200	30/200	10/200	60/200
	y_5 (80–100)	0	4/200	16/200	20/200	40/200
	$p_X(x)$ →	14/200	30/200	86/200	70/200	1

Table.3 Two-way table of dataset of the relationship in a classroom of 200 students between the amount of time studied and the percent correct

The **marginal distribution** can be used to determine how many students that scored 20 or below: $p_Y(y_1) = P_Y(Y = y_1) = \sum_{i=1}^{4} P(x_i, y_1) = \frac{2}{200} + \frac{8}{200} = \frac{10}{200}$, meaning 10 students or 5%.

The **conditional distribution** can be used to determine the probability that a student scored 20 or below while also studying for 60 minutes or more: $p_{Y|X}(y_1|x_4) = P(Y = y_1|X = x_4) = \frac{P(X = x_4, Y = y_1)}{P(X = x_4)} = \frac{8/200}{70/200} = \frac{8}{70} = \frac{4}{35}$, meaning there is about a 11% probability of scoring 20 after having studied for at least 60 minutes.

Real-world example

Suppose that the probability that a pedestrian will be hit by a car, while crossing the road at a pedestrian crossing, without paying attention to the traffic light, is to be computed. Let H be a discrete random variable taking one value from {Hit, Not Hit}. Let L (for traffic light) be a discrete random variable taking one value from {Red, Yellow, Green}.

Realistically, H will be dependent on L. That is, P(H = Hit) will take different values depending on whether L is red, yellow or green (and likewise for P(H = Not Hit)). A person is, for example, far more likely to be hit by a car when trying to cross while the lights for perpendicular traffic are green than if they are red. In other words, for any given possible pair of values for H and L, one must consider the joint probability distribution of H and L to find the probability of that pair of events occurring together if the pedestrian ignores the state of the light.

However, in trying to calculate the **marginal probability** P(H = Hit), what is being sought is the probability that H = Hit in the situation in which the particular value of L is unknown and in which the pedestrian ignores the state of the light. In general, a pedestrian can be hit if the lights are red OR if the lights are yellow OR if the lights are green. So, the answer for the marginal probability can be found by summing P(H | L) for all possible values of L, with each value of L weighted by its probability of occurring.

Here is a table showing the conditional probabilities of being hit, depending on the state of the lights. (Note that the columns in this table must add up to 1 because the probability of being hit or not hit is 1 regardless of the state of the light.)

Conditional distribution: $P(H \mid L)$

H \ L	Red	Yellow	Green
Not Hit	0.99	0.9	0.2
Hit	0.01	0.1	0.8

To find the joint probability distribution, more data is required. For example, suppose P(L = red) = 0.2, P(L = yellow) = 0.1, and P(L = green) = 0.7. Multiplying each column in the conditional distribution by the probability of that column occurring results in the joint probability distribution of H and L, given in the central 2×3 block of entries. (Note that the cells in this 2×3 block add up to 1).

Joint distribution: $P(H, L)$

H \ L	Red	Yellow	Green	Marginal probability P(H)
Not Hit	0.198	0.09	0.14	0.428
Hit	0.002	0.01	0.56	0.572
Total	0.2	0.1	0.7	1

The marginal probability P(H = Hit) is the sum 0.572 along the H = Hit row of this joint distribution table, as this is the probability of being hit when the lights are red OR yellow OR green. Similarly, the marginal probability that P(H = Not Hit) is the sum along the H = Not Hit row.

Multivariate distributions

For multivariate distributions, formulae similar to those above apply with the symbols X and/or Y being interpreted as vectors. In particular, each summation or integration would be over all variables except those contained in X.[5]

That means, If $X_1, X_2, ..., X_n$ are **discrete random variables**, then the marginal probability mass function should be

$$p_{X_i}(k) = \sum p(x_1, x_2, \ldots, x_{i-1}, k, x_{i+1}, \ldots x_n);$$

if $X_1, X_2, ... X_n$ are **continuous random variables**, then the marginal probability density function should be

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_{i-1} dx_{i+1} \ldots dx_n.$$

See also

- Compound probability distribution
- Joint probability distribution
- Marginal likelihood
- Wasserstein metric
- Conditional distribution

References

1. Trumpler, Robert J. & Harold F. Weaver (1962). Statistical Astronomy. Dover Publications. pp. 32–33.
2. "Marginal & Conditional Probability Distributions: Definition & Examples" (https://study.com/academy/lesson/marginal-conditional-probability-distributions-definition-examples.html). Study.com. Retrieved 2019-11-16.
3. "Exam P [FSU Math]" (https://www.math.fsu.edu/~paris/Pexam/). www.math.fsu.edu. Retrieved 2019-11-16.
4. Marginal and conditional distributions (https://www.khanacademy.org/math/ap-statistics/analyzing-categorical-ap/distributions-two-way-tables/v/marginal-distribution-and-conditional-distribution), retrieved 2019-11-16
5. A modern introduction to probability and statistics : understanding why and how. Dekking, Michel, 1946-. London: Springer. 2005. ISBN 9781852338961. OCLC 262680588 (https://www.worldcat.org/oclc/262680588).
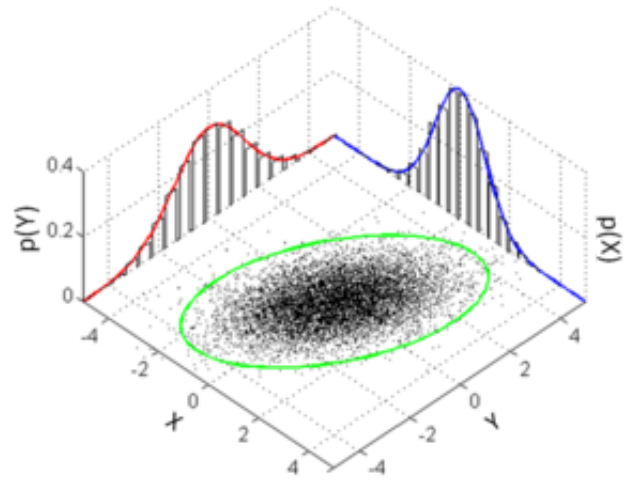
Many samples from a bivariate normal distribution. The marginal distributions are shown in red and blue. The marginal distribution of X is also approximated by creating a histogram of the X coordinates without consideration of the Y coordinates.

Bibliography

- Everitt, B. S.; Skrondal, A. (2010). Cambridge Dictionary of Statistics. Cambridge University Press.
- Dekking, F. M.; Kraaikamp, C.; Lopuhaä, H. P.; Meester, L. E. (2005). A modern introduction to probability and statistics. London : Springer. ISBN 9781852338961.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Marginal_distribution&oldid=991746821"

This page was last edited on 1 December 2020, at 16:43 (UTC).