

1 WYNTKA... Working with Random Variables

Any function of normal variables can be applied to random variables as well. They can be added, subtracted, multiplied and variously combined. However the results of doing these operations are often far from obvious, and depend on how the random variables are related to each other – in particular on their correlation, so I'll start with correlations.

1.1 Correlations and Joint Probabilities

Often, we need to be able to tell whether there is any correlation between statistical variables: in other words whether the two variables are “independent”¹. For example, the height of any person chosen from the population at random will be a random variable. However, since there is a genetic component to height, if your father is tall, there is a good chance that you will be tall as well. These two statistical variables (your height and your father’s height) are said to be *correlated*². It’s useful to be able to define a single parameter than describes whether two statistical variables are independent or not, and the most common one is the correlation coefficient.

Consider the expectation value (mean) of the product of the difference between two random variables x and y and their means:

$$\lim_{N \rightarrow \infty} \left(\frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{N} \right) = \overline{(x - \bar{x})(y - \bar{y})}$$

Now if y tends to be above its mean value when x is also above its mean value, then $(y_i - \bar{y})$ and $(x_i - \bar{x})$ will both tend to be positive, so the product will be positive. Equally, if y tends to be below the mean when x is also below the mean, then both $(y_i - \bar{y})$ and $(x_i - \bar{x})$ will be negative, and the product will again tend to be positive. Therefore, if these random variables are *positively correlated*, then this expectation value will tend to be positive. Equally, if there is a negative correlation (y tends to be above its mean when x is below its mean and vice versa), then this mean value will be negative. If there is no correlation between the two statistical variables at all, then this mean value will be zero: the chances of x being above the mean and below the mean for any particular value of y are the same.

If one of the statistical variables is just a multiple of the other (for example, if $y_i = k x_i$), then

$$\overline{(x - \bar{x})(kx - \bar{kx})} = k \overline{(x - \bar{x})^2} = k \sigma_x^2 = \sigma_x \sigma_y$$

¹ *Independent* and *correlated* are opposites here: if the chances of one variable having any particular value is not dependent on the value of the other variable, then the two variables are said to be *independent* or *uncorrelated*.

² In this case, *positively correlated*. A *negative correlation* would indicate that if your father were tall, then you would be more likely than average to be short.

since the standard deviation of kx is just k times the standard deviation of x . For this reason, we usually normalise the result above, and define a correlation coefficient, $\chi_{x,y}$, as:

$$\chi_{x,y} = \frac{(x - \bar{x})(y - \bar{y})}{\sigma_x \sigma_y}$$

after a bit of algebra, we can produce a version of this formula which is easier to calculate:

$$\begin{aligned} \sigma_x \sigma_y \chi_{x,y} &= \overline{(x - \bar{x})(y - \bar{y})} = \overline{(xy - \bar{x}y - \bar{y}x + \bar{xy})} \\ &= \overline{xy} - \overline{\bar{x}y} - \overline{\bar{y}x} + \overline{\bar{xy}} \\ &= \overline{xy} - \overline{xy} \end{aligned}$$

Normalising the correlation value in this way means that if each y_i is directly proportional to the corresponding x_i , the correlation coefficient χ is equal to one, and if each y_i is proportional to minus one times x_i , the correlation coefficient χ is minus one. If the two statistical variables x and y are independent, then the correlation coefficient is zero.

1.1.1 Conditional Probabilities and Joint Probabilities

If I had never met you, but I had met your father and he was very tall, I might expect you to be tall as well. In general, if you have two correlated statistical variables x and y , and you know the value of one, and the correlation coefficient, then it is possible to make a better guess about the value of the other. In other words, the distribution of the unknown variable changes if there is prior knowledge of the value of a correlated variable.

We can write the probability distributions of y given a certain value of $x = X$ as $p_y(y | X)$. This is known as the *conditional probability of y , given X* . Similarly, if we know y , then the probability distribution of x given that the value of y is Y is written $p_x(x | Y)$. This is the *conditional probability of x , given Y* .

Conditional probabilities are extremely useful in communications theory. More often than not, a communications receiver receives a distorted version of the transmitted signal, and the job of the receiver is to recover the original signal as accurately as possible. Clearly the received signal and the transmitted signal are correlated (if they weren't we couldn't communicate at all), but both are random variables unknown in advance (otherwise there wouldn't be any point in communicating, the receiver would already know what the transmitter was going to say). An ideal receiver should determine the conditional probabilities of each possible transmitted signal given knowledge of the received signal, and output the transmitted signal with the largest conditional probability³.

If there is no correlation between the statistical variables x and y , then knowing the value of one of them doesn't change our expectation of the value of the other one at all. So in this case we could write $p_x(x | Y) = p_x(x)$, and $p_y(y | X) = p_y(y)$. (However, if x was the received signal, and y the transmitted signal, this would indicate that there was no way of telling anything about

³ The basic problem in a lot of communications engineering is to find efficient ways to do exactly this.

the transmitted signal from the received signal, which would mean no information could be received.)

The *joint probability* of x and y is just the probability that x has some value X , and y has some value Y . It is written $p_{xy}(X, Y)$.

Now the probability that x has some value X and y has some value Y must be equal to the probability that y has some value Y , and x has some value X given that y has the value Y . In other words,

$$p_{xy}(X, Y) = p_x(X | Y) p_y(Y)$$

and it must also be equal to the probability that x has some value X , and y has some value Y given that x has the value X ,

$$p_{xy}(X, Y) = p_y(Y | X) p_x(X)$$

therefore:

$$p_{xy}(X | Y) p_y(Y) = p_{yx}(Y | X) p_x(X)$$

(This result is known as Bayes' theorem, after Thomas Bayes, an eighteenth century mathematician, and has been extensively studied in fields from jurisprudence to economics.)

Note also that if the x and y are uncorrelated: that is, it doesn't matter what value x has, we can't tell anything more about the likely value of y , then:

$$p_{xy}(X, Y) = p_x(X) p_y(Y)$$

in other words the probability of x having a value X at the same time as y has the value Y is just the product of the two individual probabilities.

1.2 Adding Random Variables

Quite often we have two random variables, and we're interested in knowing something about the distribution of the sums of these random variables. There are three important results:

- The mean of the sum is the sum of the means.
- The variance of the sum is the sum of the variances of the individual distributions, provided the individual distributions are independent.
- The probability distribution of the sum is the convolution of the probability distributions of the individual distributions, again provided the individual distributions are independent.

1.2.1 The Mean of the Sum

Consider two random variables x and y , with probability distributions $p_x(x)$ and $p_y(y)$. Let a set of samples of these variables be x_i and y_i as usual. Then, from the definition of the mean:

$$\overline{x_i + y_i} = \lim_{N \rightarrow \infty} \left(\frac{\sum_{i=1}^N x_i + y_i}{N} \right) = \lim_{N \rightarrow \infty} \left(\frac{\sum_{i=1}^N x_i}{N} \right) + \lim_{N \rightarrow \infty} \left(\frac{\sum_{i=1}^N y_i}{N} \right) = \overline{x_i} + \overline{y_i}$$

and note that it doesn't matter whether the statistical variables are independent or not, this result always holds.

1.2.2 The Variance of the Sum

Using the previous result, that the expectation value (means) of the sum of two random variables is the sum of the expectation values of the individual random variables, we can determine how to work out the variance of the sum of two random quantities:

$$\begin{aligned} \sigma_{x+y}^2 &= \overline{(x+y)^2} - \overline{(x+y)}^2 = \overline{x^2 + 2xy + y^2} - \overline{(x+y)}^2 \\ &= \overline{x^2} + 2\overline{(xy)} + \overline{y^2} - \overline{x}^2 - 2\overline{(x)}\overline{(y)} - \overline{y}^2 \\ &= \overline{x^2} - \overline{x}^2 + \overline{y^2} - \overline{y}^2 + 2\left(\overline{(xy)} - \overline{(x)}\overline{(y)}\right) \\ &= \sigma_x^2 + \sigma_y^2 + 2\left(\overline{(xy)} - \overline{(x)}\overline{(y)}\right) \end{aligned}$$

Noting the formula we worked out for the correlation coefficient, we can write this as:

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y\chi_{x,y}$$

This is a useful result. This says that if the two statistical variables are independent (the correlation coefficient is zero), then the variance of the sum is just equal to the sum of the variances of each variable. However if the two variables are correlated, then this is not true. In the case where the two statistical variables are always equal or are directly proportional to each other, then the correlation coefficient is one, and we have:

$$\sigma_{2x}^2 = \sigma_x^2 + \sigma_x^2 + 2\sigma_x\sigma_x = 4\sigma_x^2$$

This is an important result when considering what happens when two copies of the same signal are added: remembering that the variance is proportional to the mean power in the signal (provided the mean value of the signal is zero, which it usually is): the sum has four times as much power as each individual signal. In the case of two totally uncorrelated noise sources, the mean power in the sum is just twice the mean power in each individual noise source.

1.2.3 The Probability Distribution of the Sum

Another question that often arises is: given two random variables (x and y as usual), what is the probability distribution of their sum? Suppose the sum is $z = x + y$. There are (for continuous random variables at least) an infinite number of different ways to get the sum z : x can have any value X , provided y has the value $(z - X)$. The probability of x and y having these values is given by the joint probability $p_{xy}(X, z-X)$; and the total probability that the sum has a value z is the sum of all these different ways to make z :

$$p_z(z) = \int_{-\infty}^{\infty} p_{xy}(X, z - X) dX = \int_{-\infty}^{\infty} p_x(x) p_y(z - X | X) dX$$

and in the case where the two variables x and y are independent, $p_y(y)$ is not a function of the value of x , so we can just write:

$$p_z(z) = \int_{-\infty}^{\infty} p_x(X) p_y(z - X) dX$$

which is a convolution integral⁴. If x and y were independent discrete statistical variables with distributions $q_x(x)$ and $q_y(y)$ respectively, then we'd get:

$$q_z(z) = \sum_{X=-\infty}^{\infty} q_x(X) q_y(z - X)$$

1.3 Distributions of Functions of Statistical Variables

Let's start this one with an example. Suppose you have a statistical variable x , with a probability distribution $p_x(x)$. What is the probability distribution of $y = 1/x$?

First, let the probability distribution of y be $p_y(y)$. For any value X , the probability that x is lower than X must be the same as the probability that y is greater⁵ than $Y = 1/X$. In other words⁶:

⁴ Tip: when working these out, pay close attention to the limits of the integration. For example, try working out the probability distribution of the sum of two independent statistical variables with uniform distributions from zero to one:

$$p_z(z) = \int_{-\infty}^{\infty} p_x(X) p_y(z - X) dX = \begin{cases} 0 & z < 0 \\ \int_0^z dX = z & 0 < z < 1 \\ \int_{z-1}^1 dX = 2 - z & 1 < z < 2 \\ 0 & z > 2 \end{cases}$$

since when $z < 0$, $(z - X)$ is negative and when $z > 2$, $(z - X)$ is always greater than one: in both cases $p_y(z - X)$ is zero. When $0 < z < 1$, X must be greater than zero (for $p_x(X)$ to be one) and less than z (so $p_y(z - X)$ is also one), etc.

⁵ Greater, since as x increases, y decreases: larger values of x correspond to smaller values of y . All values of x less than X correspond to values of y greater than $1/X$.

⁶ A couple of notes about this integral: the probability that y is greater than some value Y (here $Y = 1/X$) is one minus the probability that y is less than that value. Then, I differentiate the integral: the small change in the value of an integral when its upper limit is changed is just the value of the integral at that upper limit times dX . (Think of the [continued on next page...])

$$\int_{-\infty}^X p_x(x) dx = \int_Y^{\infty} p_y(y) dy = 1 - \int_{-\infty}^Y p_y(y) dy$$

$$p_x(X) dX = -p_y(Y) dY$$

Now since $Y = 1/X$, $dY/dX = -1/X^2$, so:

$$p_x(X) = -p_y(Y) \left(-\frac{1}{X^2} \right)$$

$$p_y(Y) = \frac{p_x(X)}{X^2} = \frac{p_x(1/Y)}{Y^2}$$

So the probability distribution $p_y(y) = p_x(1/y)/y^2$.

That's a particularly simple example because the relationship between x and y was monotonic: as x increased, y decreased uniformly, and for every value of x there is one, and only one, value of y . It's possible that this might not be true for all cases: consider what would have happened if $y = x^2$. Then the probability that y is greater than X^2 is the sum of the probabilities that x is greater than X , and the probability that x is less than $-X$. In this case, we would have to write⁷:

$$\int_{-X}^X p_x(x) dx = \int_0^Y p_y(y) dy$$

$$(p(X) + p(-X)) dX = p(Y) dY$$

$$p(Y) = (p(X) + p(-X)) \frac{dX}{dY}$$

$$= \frac{1}{2X} (p(\sqrt{Y}) + p(-\sqrt{Y}))$$

$$= \frac{1}{2\sqrt{Y}} (p(\sqrt{Y}) + p(-\sqrt{Y}))$$

Consider a real example: the distribution of the amplitude of a fading radio wave can often be accurately modelled using the Rayleigh distribution:

$$p(x) = \frac{2x}{r^2} \exp\left(-\frac{x^2}{r^2}\right)$$

integral as being the area under the curve – I'm adding a small amount of area of width dX , and height $p(x)$.) This has got to be equal to the small change in area of the other integral, which is determined in the same way – a change of dX in X corresponds to a small change dY in Y .

⁷ Again, think of differentiating this integral in terms of the area under the curve – I'm adding a small area at the bottom of area $p(-X)dX$, and a small area at the top, of $p(X)dX$.

This is the distribution of the amplitude of the signal: but what is the distribution of the power in the signal? Power is related to amplitude by $P = kx^2$, where k represents the impedance of the media the signal is travelling in, which is often normalised to one. Since in this case the amplitude can never be negative, we don't have to worry about negative values of x . So, the probability distribution of the power in this case is:

$$\begin{aligned} p(P) &= \frac{1}{2\sqrt{P}} p(\sqrt{P}) \\ &= \frac{1}{2\sqrt{P}} \frac{2\sqrt{P}}{r^2} \exp\left(-\frac{P}{r^2}\right) \\ &= \frac{1}{r^2} \exp\left(-\frac{P}{r^2}\right) \end{aligned}$$

This is known as a negative exponential distribution, and it is a very common distribution. We'll be seeing a lot more of it later on.

1.4 Problems

- 1) Consider the weighted die described previously, which is twice as likely to come up as a one as any other number. Suppose you have two of these dice, and you add up the result each gives. What is the mean and variance of this sum? What is the probability distribution of the result?
- 2) Suppose you have N of these dice. What is the mean and variance of the sum of a roll of all N of the dice?
- 3) In a certain communications system, the probability of transmitting a '1' bit is 0.3. The probability of receiving a '1' given that a '1' was transmitted is 0.9, and the probability of receiving a '1' is 0.5. What is the probability that a '1' is transmitted and received correctly?
- 4) If x is a random variable with a uniform distribution between 1 and 2, what is x^3 ? And what is $x^3 + x$?
- 5) What is the probability density function of the distances from a point placed entirely randomly within a circle to the centre of that circle? Assume that the point is equally likely to be anywhere within the circle.