Random Variables

A random variable is a function that assigns a number to every sample point in a sample space. Random variables are usually denoted by capital letters such as X, but they can also be denoted by $X(\bullet)$ in which the dot signifies a sample point. As an example let's define a random variable $X(\bullet)$ on the sample space for the four coin tosses experiment. Let $X(\bullet)$ = 1 if the first toss is heads and 0 if the first toss is tails, then X(HTTH) = 1, X(THTH) = 0, X(HHHT) = 1 and so on, where the HTTH, THTH, and the HHHT are sample points for the experiment.

Let $Y(\bullet)$ = 1 if the third toss is a head and 0 if the third toss is tails, then Y(HTTH) = 0, Y(HHHT) = 1 and so on.

Random variables can be defined in terms of other random variables as the following 3 examples show.

Let $Z(\bullet) = X(\bullet) + Y(\bullet)$ (Z = X + Y in the more common notation)

then Z(HHHH) = X(HHHH) + Y(HHHH) = 1 + 1 = 2.

Z(HTTT) = X(HTTT) + Y(HTTT) = 1 + 0 = 1.

Z(THTH) = X(THTH) + Y(THTH) = 0 + 0 = 0.

Z(THHH) = X(THHH) +Y(THHH) = 0 + 1 = 1.

and so on.

Let $W(\bullet) = 5Y(\bullet)$ (W = 5Y in the more common notation)

then W(HTHH) = 5Y(HTHH) = 5 x 1 = 5.

W(HHTH) = 5Y(HHTH) = 5 x 0 = 0.

And so on.

Let $V(\bullet) = X(\bullet)Z(\bullet)$ (V = XZ in the more common notation)

then V(HTHT) = X(HTHT)Z(HTHT) = 1 x 2 = 2.

V(HTTT) = X(HTTT)Z(HTTT) = 1 x 1 = 1.

V(THHT) = X(THHT)Z(THHT) = 0 x 1 = 0 and so on.

The probability function p(x) of a random variable is a function whose value for a given x is the probability that the random variable takes on the value x.
The probability that a random variable takes on the value x is the sum of the probabilities of the sample points which the random variable assigns the value x.
Two random variables are identically distributed if they have identical probability functions.

Independent random variables.
Two random variables X and Y are independent if
$P(X = x \text{ and } Y = y) = P(X = x)P(Y = y)$ for every possible value for x and every possible value for y.

Pairwise independence
The random variables $X_1, X_2, \ldots, X_N$ are pairwise independent if every pair of them are independent.

Expectation of a random variable

The "Lots of Coins" slot machine can return 0, 2, 5, 10, 20 or 50 coins with respective probabilities 414/500, 1/10, 1/25, 1/50, 1/100 and 1/500 .
Let f_i be how many times in N spins i coins were returned on a spin, then $0 \bullet f_0 + 2 \bullet f_2 + 5 \bullet f_5 + 10 \bullet f_{10} + 20 \bullet f_{20} + 50 \bullet f_{50}$ is the total number of coins returned by the slot machine in N spins and the average number of coins that were returned per spin is:

$$(0 \bullet f_0 + 2 \bullet f_2 + 5 \bullet f_5 + 10 \bullet f_{10} + 20 \bullet f_{20} + 50 \bullet f_{50)}) / N$$

$$= \ 0 \bullet \frac{f_0}{N} + 2 \bullet \frac{f_2}{N} + 5 \bullet \frac{f_5}{N} + 10 \bullet \frac{f_{10}}{N} + 20 \bullet \frac{f_{20}}{N} + 50 \bullet \frac{f_{50}}{N}$$

$\frac{f_i}{N}$ is the relative frequency that i coins were returned on a spin in N spins.
In the long run we expect $\frac{f_i}{N}$ to be close to p_i, the probability of getting a return of i coins on a spin. So the expected average return per spin for this slot machine is:

$0 \bullet 414/500 + 2 \bullet 1/10 + 5 \bullet 1/25 + 10 \bullet 1/50 + 20 \bullet 1/100 + 50 \bullet 1/500$ which equals .90.

The return on a spin is a random variable taking on the values 0, 2, 5, 10, 20, and 50 with probabilities 414/500, 1/10, 1/25, 1/50, 1/100 and 1/500.

In general, E(X) the expected value or expectation of a random variable equals $x_1\, p(x_1) + x_2\, p(x_2) + \ldots + x_k\, p(x_k)$ where k is the number of different values the random variable can take on, x_1, x_2, ..., x_k are the different values and the $p(x_i)$'s are the probabilities of getting those values.

In the slot machine example, we did not describe what the sample space was. In antique slot machines there were three reels with 20 stopping positions on each reel for a total of 8,000 stopping positions. The sample space would be these 8,000 stopping positions which each had a probability of 1/8000. Each stopping position is a sample point. The return on a spin random variable assigns a value to a sample point based on what the return would be for the combination of symbols appearing on the payline for that position.

In the formula $E(X) = x_1 p(x_1) + x_2 p(x_2) + . . . + x_k p(x_k)$, no reference is made to the sample space. We now give a definition of E(X) which is equivalent to the previous one and which makes reference to the sample space.

$E(X) = \Sigma X(\bullet)p(\bullet)$, where $\bullet$ is a sample point , $X(\bullet)$ is the value of X for that point, $p(\bullet)$ is the probability of the sample point, and where the sum ranges over all the sample points.
The event that a random variable takes on the value x_i is the set of sample points for which $X(\bullet) = x_i$ and the probability for this event is the sum of the probabilities for those sample points.
For any x_i, the sum $\Sigma X(\bullet)p(\bullet)$ will include all the products $X(\bullet)p(\bullet)$ for which $X(\bullet) = x_i$. Factoring x_i out of the sum of these products we have x_i times the sum of the probabilities for the sample points for which $X(\bullet) = x_i$ and this equals $p(x_i)$.
So for any x_i, the sum $\Sigma X(\bullet)p(\bullet)$ includes the product $x_i\, p(x_i)$.
This shows that $E(X) = \Sigma X(\bullet)p(\bullet)$ is equivalent to
$E(X) = x_1 p(x_1) + x_2 p(x_2) + . . . + x_k p(x_k)$.

Variance of a random variable

Consider the random variables X and Y whose probability functions are:

probability function of X: p(-1) = .4, p(0) = .3, and p(1) = .3
probability function of Y: p(-10) = .04, p(0) = .93, and P(10) = .03

Both these random variables have 3 possible values and both have an expectation of -.1, but the non zero values of random variable Y's possible values -10, 0, and 10 are farther away from the expected value of -.1 than the non zero values of X whose possible values are -1, 0 and 1.

To measure the average deviation of a random variable X from it's expectation, we might try calculating the expectation of $X - E(X)$. But we would always get zero since the negative deviations would cancel out the positive ones. Two alternatives are either to calculate the expectation of $|X - E(x)|$ or to calculate the expectation of $(X - E(X))^2$. The variance is the expectation of $(X - E(X))^2$. Why choose $(X - E(X))^2$? Well for one thing, it makes it easy to prove the law of large numbers and that's good enough for me.

The variance of a random variable X is:
$V(X) = (x_1 - E(X))^2p(x_1) + (x_2 - E(X))^2p(x_2) + . . . + (x_k - E(X))^2p(x_k)$
where x_1, x_2, . . . , x_k are the possible values that X can take on , and $p(x_1)$, $p(x_2)$, ..., $p(x_k)$ are the probabilities that X will take on those values.

Daniel Daniels updated 12/10/2022