The Geometric Distribution

Prerequisites

You should have studied discrete probability distributions.

Definition of a discrete random variable

Let X be a variable such that

(*a*) It is discrete, meaning it can only take n exact values $x_1, x_2, \ldots, x_n$. When X takes the value x_i we write $X = x_i$.

(*b*) It is random, meaning that with each value x_i that the variable takes, there is associated a probability p_i. We write this $P(X = x_i) = p_i$. The assignment of probabilities to each value that X can take is called a discrete probability distribution.

(*c*) Because it is random it obeys the law of total probability. The sum of all the probabilities for all n values is equal to 1.

You should also have studied geometric progressions.

Example (1)

A geometric progression has first term $\frac{1}{6}$ and common ratio $\frac{5}{6}$. The nth term of this series is denoted by u_n.

(*a*) Find

(*i*) The third term u_3.

(*ii*) The nth term u_n

(*iii*) The sum to infinity.

(*b*) Find $\left(\frac{5}{6}\right)^2, \left(\frac{5}{6}\right)^3, \left(\frac{5}{6}\right)^{20}, \left(\frac{5}{6}\right)^{50}$ as decimals and conjecture to what number $\left(\frac{5}{6}\right)^n$ tends as $n \to \infty$; that is, as n gets larger and larger without limit.

(*c*) Explain why the variable X that assigns to the value n the probability u_n, $P(X = n) = u_n$, may be regarded as a discrete random variable.

Solution

(*a*) $a = \frac{1}{6} \quad r = \frac{5}{6}$

(*i*) $u_3 = ar^2 = \frac{1}{6} \times \left(\frac{5}{6}\right)^2 = \frac{25}{216}$

(*ii*) $u_n = ar^n = \frac{1}{6} \times \left(\frac{5}{6}\right)^{n-1}$

(*iii*) $S_\infty = \frac{a}{1-r} = \frac{\frac{1}{6}}{1-\frac{5}{6}} = 1$

(*b*) $\left(\frac{5}{6}\right)^2 = 0.6944...$ $\left(\frac{5}{6}\right)^3 = 0.5787...$ $\left(\frac{5}{6}\right)^{20} = 0.0260...$ $\left(\frac{5}{6}\right)^{50} = 0.0001...$

This number is getting smaller and smaller the larger *n* gets. We conjecture that as $n \to \infty, \left(\frac{5}{6}\right)^n \to 0$ which we write as

$\lim_{n\to\infty}\left(\frac{5}{6}\right)^n = 0$.

(*c*) By part (*a*)(*iii*) the sum of all the assignments given in part (*a*)(*ii*) as

$P(X = n) = u_n = ar^n = \frac{1}{6} \times \left(\frac{5}{6}\right)^{n-1}$

is equal to 1. This assignment obeys the law of total probability. Thus the values $u_n = ar^n = \frac{1}{6} \times \left(\frac{5}{6}\right)^{n-1}$ may be interpreted as probabilities. The entire set of these values is as follows.

n	1	2	3	4	...	n	...
$P(X = n)$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{25}{216}$	$\frac{125}{1296}$	...	$\frac{1}{6} \times \left(\frac{5}{6}\right)^{n-1}$	...

This table is a probability distribution.

Remark

In the definition of a probability density function given above we wrote that such a function must take *n* exact values $x_1, x_2, ... , x_n$. For the geometric progression above to be interpreted as a probability distribution we must allow that *n* may be potentially infinite. We state (without proof) that the conjecture $n \to \infty, \left(\frac{5}{6}\right)^n \to 0$ is true, which goes further to justify this *extension* of the concept of a discrete probability distribution to allow for an *infinite* set of values and their associated probabilities because the probability of the *n*th term gets smaller and smaller and converges on 0 as *n* gets larger and larger.

The geometric distribution

The geometric distribution is the discrete probability distribution that arises when a trial is repeated only until a "success" (or "failure") is obtained.

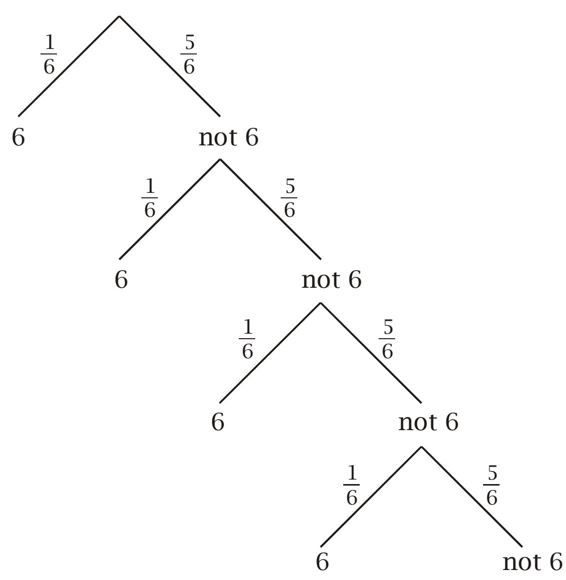
Example (2)

A fair cubical die is tossed until a 6 is thrown. Let X be the random variable representing the number of throws of the die until a 6 is obtained. Draw a probability tree to represent this situation and find the probability distribution of X.

Solution

$P(6)=\frac{1}{6}$ $\qquad$ $P(\text{not } 6)=P(\bar{6})=\frac{5}{6}$

The probability tree is



The tree carries on indefinitely, but the probability of not-6 becomes less and less, tending to zero. The probability distribution is precisely the geometric progression that we investigated in example (1).

n	1	2	3	4	...	n	...

$P(X=n)$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{25}{216}$	$\frac{125}{1296}$	...	$\frac{1}{6}\times\left(\frac{5}{6}\right)^{n-1}$	...

In this example we immediately recognise that this is common situation that does not depend on the precise probability $p=\frac{1}{6}$. The probability p is the parameter of the distribution. In general, if the probability of a success is p, and the probability of a failure is q, then the probability that n trials are required before a success is obtained is

$P(X=n)=p\times q^{n-1}$ where $q=1-p$.

Definition of a geometric distribution

A discrete random variable X having probability density function of the form

$P(X=n)=p\times q^{n-1}$ where $0\le p\le 1$ and $q=1-p$

is said to follow a geometric distribution for $n=1, 2, 3, ...$

This is written $X\sim Geo(p)$

The expression $X\sim Geo(p)$ is read, "X follows a geometric distribution with parameter p". As usual we are interested in finding the mean (expectation) and variance of a geometric distribution. This we state as follows.

Expectation and variance of a geometric distribution

Given $X\sim Geo(p)$ then

$$E(X)=\frac{1}{p}$$

$$Var(X)=\frac{q}{p^2}$$

Example (3)

During March the probability of rain is 0.25 on any day. Let X be the discrete random variable denoting the number of days until it rains following a dry day.

(*a*) Find the probability that the first rainy day will be the 4th day following a dry day.

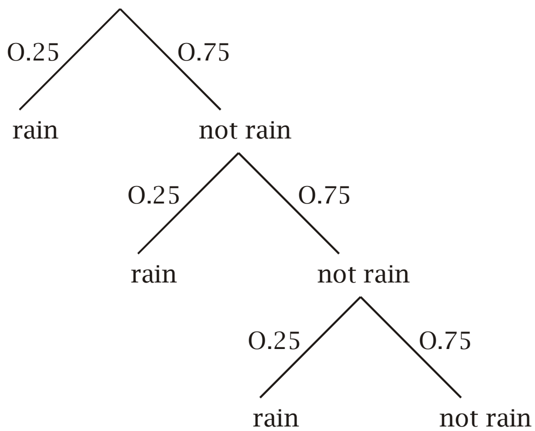
(*b*) Find the probability that it will not rain until after the 4th day following a dry day.

(*c*) Given that a certain day is dry, find the number of days that one would expect to wait on average until there is another rainy day.

(*d*) Find the variance of the probability distribution of X.

Solution

The probability tree illustrates that this is a geometric probability distribution



This illustrates that the problem may be modelled as a geometric distribution with parameter 0.25. Hence $X \sim Geo(0.25)$.

(*a*) $P(X = 4) = 0.25 \times (0.75)^3 = 0.1055$ (4 s.f.)

(*b*) $P(X > 4) = 1 - P(X = 1) - P(X = 2) - P(X = 3) - P(X = 4)$

$$= 1 - 0.25 - 0.25 \times 0.75 - 0.25 \times 0.75^2 - 0.25 \times 0.75^3$$

$$= 1 - 0.25 - 0.1875 - 0.1406 - 0.1055$$

$$= 0.3164$$

(*c*) $E(X) = \dfrac{1}{p} = \dfrac{1}{0.25} = 4$

We expect on average that it will rain on the 4th day following a day on which it does not rain.

(*d*) $Var(X) = \dfrac{q}{p^2} = \dfrac{0.75}{(0.25)^2} = 12$