

We now shall study properties of a random walk process with only upward or downward steps that is stopped after the first upward movement and for which steps are of length $c = 1$. With regard to gambling, an upward movement represents a win. Again, we shall determine the mean and variance of the final height, the mean and variance of the trajectory arclengths, and the mean signed area between the trajectories and the x -axis. We also shall determine the probability distribution of all ending heights. We begin with some introductory background on the geometric distribution.

The Geometric Distribution

Suppose there is a fixed probability $p > 0$ of having a “success” on any single attempt. If we continue to make independent attempts, then the *geometric random variable* counts the total number of attempts needed to obtain the first success. We denote this discrete random variable by $Y \sim \text{geo}(p)$ and its range is the set of natural numbers $\{1, 2, 3, \dots\}$.

The probability of having a “success” on the first attempt is p . In general, to succeed for the first time on the k th attempt, there would have to be $k - 1$ “failures” beforehand. By independence, any sequence having $k - 1$ failures followed by a success occurs with probability $q^{k-1}p$, where $q = 1 - p$ is the probability of “failure” on any attempt. Thus, the *probability* of succeeding for the first time on the k th attempt is given by

$$P(Y = k) = q^{k-1}p, \text{ for } k \geq 1.$$

There is a useful closed-form formula for the cumulative distribution function (cdf) $P(Y \leq k)$. The event $\{Y \leq k\}$ means that the first success occurs *within* k attempts. Its complement is the event that there are no successes in any of the first k attempts, which has probability q^k . So the cdf of the geometric distribution is given by

$$P(Y \leq k) = 1 - q^k, \text{ for } k \geq 1.$$

We can use the geometric series formula to verify that the probability distribution sums to 1:

$$\sum_{k=1}^{\infty} P(Y = k) = \sum_{k=1}^{\infty} q^{k-1}p = p \sum_{k=0}^{\infty} q^k = p \times \frac{1}{1 - q} = 1.$$

Other Geometric Sums

The geometric series gives us $\sum_{k=m}^{\infty} x^k = x^m / (1 - x)$, for $-1 < x < 1$. Taking the derivative of both sides with respect to x , we obtain

$$\sum_{k=m} k x^{k-1} = \frac{m x^{m-1} - m x^m + x^m}{(1-x)^2}. \quad (1)$$

In particular, for $m = 1$, we have

$$\sum_{k=1} k x^{k-1} = \frac{1}{(1-x)^2} \quad (2)$$

Multiplying by x In Equation (2), we have for $-1 < x < 1$:

$$\sum_{k=0} k x^k = \sum_{k=1} k x^k = \frac{x}{(1-x)^2} \quad (3).$$

Taking the derivative of both sides of (3) with respect to x then multiplying by x , we obtain for $-1 < x < 1$:

$$\sum_{k=0} k^2 x^{k-1} = \sum_{k=1} k^2 x^{k-1} = \frac{1+x}{(1-x)^3} \quad (4)$$

Mean and Variance of Geometric Distribution

Using these variations of geometric series, we can derive the expected value and variance of the geometric random variable $Y \sim \text{geo}(p)$. From (2), the expected value is given by

$$E[Y] = \sum_{k=1} k P(Y = k) = p \sum_{k=1} k q^{k-1} = p \times \frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}.$$

Next from (4) we have

$$E[Y^2] = \sum_{k=1} k^2 P(Y = k) = p \sum_{k=1} k^2 q^{k-1} = p \times \frac{1+q}{(1-q)^3} = \frac{1+q}{p^2}; \text{ thus,}$$

$$\text{Var}(Y) = E[Y^2] - (E[Y])^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2}.$$

Proposition. Let $p > 0$ be the fixed probability of success on any single independent attempt. In order to have a probability of at least R of succeeding within k attempts, the number of attempts must satisfy

$$k \geq \frac{\ln(1-R)}{\ln(1-p)}.$$

Proof. Simply solve for k in the inequality $R \leq 1 - q^k$.

Distribution of the Final Heights

We now describe the range of all possible final heights X_T upon stopping after the first upward movement and the probabilities of ending at such heights. We again assume that movements are always either up with probability p or down with probability $q = 1 - p$ over intervals of length $c = 1$. We assume an initial height of X_0 , which may be random. The only paths now under consideration are

$$a \quad -b, a \quad -b, -b, a \quad -b, -b, -b, a \quad -b, -b, -b, -b, a \quad \text{etc.}$$

The distribution of X_T is then shown in the following table:

Steps Needed T	Ending Height X_T	Probability
1	$X_0 + a$	p
2	$X_0 - b + a$	qp
3	$X_0 - 2b + a$	$q^2 p$
.	.	.
.	.	.
k	$X_0 - (k - 1)b + a$	$q^{k-1} p$
.	.	.

Thus, $P(X_T = X_0 - (k - 1)b + a) = q^{k-1} p$, for $k \geq 1$.

Moreover, we can write the final height as

$$X_T = X_0 - b(T - 1) + a. \tag{5}$$

Mean of the Final Height

We next shall derive the mean of the final height X_T . Again, we assume that the process begins with an initial height X_0 that may be random and that is independent of the steps to follow. From (5), the average final height is then found by

$$\begin{aligned} E[X_T] &= E[X_0] + a - b(E[T] - 1) = E[X_0] + a - b \left(\frac{1}{p} - 1 \right) \\ &= E[X_0] + a - b \frac{1 - p}{p} \\ &= E[X_0] + a - \frac{q}{p} b. \end{aligned}$$

Notes: (i) With regard to gambling, we generally assume that $a < \frac{q}{p} b$, so that the casino is assured to make a profit in the long run. Under this condition, we see that $E[X_T] < E[X_0]$.

(ii) In order for us to have $X_T = X_0$, we can only sustain n successive losses such that $nb = a$. Thus, we can only have $n = a/b$ losses.

(iii) The probability of coming out at least even upon the first win is obtained by having the first win within $a/b + 1$ attempts. Thus,

$$P(X_T = X_0) = 1 - q^{a/b + 1}.$$

(iv) The average of the general random walk at time t has previously been derived as

$$E[X_t] = E[X_0] + \frac{t(ap - bq)}{c}.$$

Now suppose that $c = 1$, and we use the time $t = E[T] = 1/p$. Then

$$E[X_{1/p}] = E[X_0] + \frac{1}{p}(ap - bq) = E[X_0] + a - \frac{q}{p}b = E[X_T].$$

$$\text{That is, } E[X_{E[T]}] = E[X_T].$$

So the average of the height upon stopping X_T is equal to the average of the height of all possible walks at the average time of stopping $X_{E[T]}$. Another way to interpret the result is to say that we multiply the average gain per step, $ap - bq$, by the average number of steps $1/p$ taken to find the total average gain in height.

Variance of the Final Height

We again consider the final height in the form $X_T = X_0 - b(T - 1) + a$. Because adding the constant a does not change the variance and because X_0 is independent of the steps that follow, we have

$$\begin{aligned} \text{Var}(X_T) &= \text{Var}(X_0 - b(T - 1) + a) \\ &= \text{Var}(X_0 - b(T - 1)) \\ &= \text{Var}(X_0) + \text{Var}(-b(T - 1)) \\ &= \text{Var}(X_0) + b^2 \text{Var}(T - 1) \\ &= \text{Var}(X_0) + b^2 \text{Var}(T) \\ &= \text{Var}(X_0) + b^2 \times \frac{1-p}{p^2}. \end{aligned}$$

Exercises

1. Recall that the mean arclength through time t of the general random walk process is

$$E[L_t] = \frac{t}{c} \left(\sqrt{c^2 + a^2} p + c r + \sqrt{c^2 + b^2} q \right).$$

Assume that $r = 0$ and $c = 1$. Let T be the number of steps needed for the first upward movement.

- (a) Explain how to write the arclength at time T , L_T , in terms of T .
- (b) Derive $E[L_T]$.
- (c) Determine if $E[L_T] = E[L_{E[T]}]$.

2. Assume that $r = 0$ and $c = 1$. Let T be the number of steps needed for the first upward movement.

- (a) Explain how to write the area under the curve up to time T , A_T , in terms of T .
- (b) Derive $E[A_T]$.