1 Defining the Process

The Bernoulli Process is historically the first stochastic process ever studied. It can be thought of as a sequence of tosses of a (not necessarily fair) coin.

Let $Y_1, Y_2, ...$ be i.i.d. Bernoulli random variables with parameter p with

$$Y_i = \begin{cases} 0 & \text{with probability } \ p \\ 1 & \text{with probability } \ 1-p=q \end{cases}$$

To keep with the coin flipping analogy, getting a head would be represented by a 1 and tails by 0. Now let

$$N_k = \sum_{i=1}^{k} Y_i$$

be the number of heads up to the k-th toss, which is distributed Binomial (k, p).

1.1 Waiting Times

Let S_n be the time at which the n-th head occurs. Then,

$$S_n = \inf\{k : N_k = n\}$$

Let $X_n = S_n - S_{n-1}$ be the number of tosses to get the n-th head starting from the $(n-1)$-th head.

2 Properties and Useful Results

1. The waiting times $X_1, X_2, ...$ are i.i.d. Geometric(p) random variables.
2. The time at which the n-th head occurs (S_n) is Negative Binomal(n, p).
3. Given $N_k = n$, the distribution of $(S_1, ..., S_n)$ is the same as the distribution of a random sample of n numbers chosen without replacement from $\{1, 2, ..., k\}$.
4. Given $S_n = k$, the distribution of $(S_1, ..., S_{n-1})$ is the same as the distribution of a random sample of $n-1$ numbers chosen without replacement from $\{1, 2, ..., k-1\}$.
5. We have as sets:
$$\{S_n > k\} = \{N_k < n\}$$
6. Central limit theorem
$$\frac{N_k - E[N_k]}{\sqrt{V(N_k)}} = \frac{N_k - kp}{\sqrt{kp(1-p)}} \longrightarrow N(0,1)$$

7. Central limit theorem

$$\frac{S_n - E[S_n]}{\sqrt{V(S_n)}} = \frac{S_n - \frac{nq}{p}}{\sqrt{nq}/p} \longrightarrow N(0,1)$$

8. As $p \to 0$

$$\frac{X_1}{E[X_1]} = \frac{X_1}{1/p} \longrightarrow \text{Exponential}(\lambda = 1)$$

9. As $p \to 0$

$$\mathbf{P}\{N_{[\frac{t}{p}]} = j\} \to \frac{t^j}{j!}e^{-t}$$

3 Proofs of Properties

1. Waiting times are geometric(p) by the definition of the geometric distribution.

   To show that the X_i's are independent, let $X_n = S_n - S_{n-1}$ be the waiting time for the n-th success starting from the $(n-1)$-th success. We know $S_0 = 0$ and $S_n = \inf\{k : N_k = n\}$. So $X_1 = S_1 - S_0 = S_1 - 0 = S_1$ and therefore S_1 is independent of S_0.

   Now assume that $S_1 = k$, then by the same logic $X_2 = S_2 - S_1 = S_2 - k$. Note that

   - The Bernoulli trial $Y_k + 1$ is independent of $Y_k, Y_{k-1}, \ldots, Y_1$.
   - The event $Y_{k+1} = 1$ is therefore independent of $Y_k, Y_{k-1}, \ldots, Y_1$.

   Therefore, in the case $S_1 = k$, we see $S_2 - S_1 = k + 1 - k = 1$. In other words,$S_2 - S_1$is independent of $S_1 = k$. The same result occurs if $Y_{k+1} = 0, Y_{k+2} = 0, \ldots, Y_{k+n} = 1$. In that case $S_2 - S_1 = k + n - k = n$. This shows that X_1 is independent of X_2.

   In general, we can repeat this logic as long as we want. e.g., X_3 is independent of X_2 since if $X_2 = n$, then $X_3 = S_3 - S_2 = (k+n+m) - (k+n) = m$, which is independent of X_2 no matter what value n takes (or k), and so on...

2. The distributional assertions are easy to prove; we may just use the definition of geometric (p) and negative binomial random variables. We need only to show that the Xi's are independent.

3.

4.

5.

6. The Central Limit Theorem (CLT) states: If $X_1, X_2, ..., X_n, ...$ are random samples drawn from a population with overall mean μ and finite variance σ^2, and if $\bar{X}_n$ is the sample mean of the first n samples, then the limiting form of the distribution, $Z = \lim_{n\to\infty}\left(\frac{\bar{X}_n-\mu}{\sigma_{\bar{X}}}\right)$ with $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, is a standard normal distribution.

   First note that $N_k = X_1 + X_2 + \cdots + X_n$ where $X \sim \text{Bernoulli}(p)$, so $\mu = p$ and $\sigma = \sqrt{pq} \implies \sigma_{\bar{X}} = \frac{\sqrt{pq}}{\sqrt{k}}$. Then we have

$$\frac{N_k - E[N_k]}{\sqrt{\text{Var}(N_k)}} = \frac{N_k - kp}{\sqrt{kpq}} = \frac{k(\frac{1}{k}N_k - p)}{k\sqrt{pq/k}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{k}}$$

   And by the CLT $\frac{\bar{X}-\mu}{\sigma/\sqrt{k}} \longrightarrow N(0,1)$ as $k \to \infty$

7. Let $X_n = S_{n+1} - S_n$ be the waiting time for the n^{th} head starting from the $(n-1)^{th}$ head. S_n is the time to flip n heads, or in other words S_n is just the sum of the waiting times: $S_n = X_1 + X_2 + \cdots + X_n$. Therefore S_n is Negative-Binomial(n, p) and the X_i's are geometric(p) (see properties 1 and 2).

   Applying the central limit theorem to S_n gives

$$\frac{S_n - E[S_n]}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - \frac{n}{p}}{\frac{\sqrt{nq}}{p}} = \frac{n(\frac{1}{n}S_n - \frac{1}{p})}{n(\frac{\sqrt{q/n}}{p})}$$

   Under the "number of failures before a success" definition of the geometric distribution, $\mu = \frac{1}{p}$ and $\sigma^2_{\bar{X}} = \frac{q}{p^2} \implies \sigma_{\bar{X}} = \frac{\sqrt{q}}{p}$. Letting $\frac{1}{n}S_n = \bar{X}$, we have

$$\frac{n(\frac{1}{n}S_n - \frac{1}{p})}{n(\frac{\sqrt{q/n}}{p})} = \frac{\frac{1}{n}S_n - \frac{1}{p}}{\frac{\sqrt{q}}{p}\frac{1}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\sigma_x/\sqrt{n}} \longrightarrow N(0,1)$$

   as $n \to \infty$.

8.
$$P\left(\frac{X_1}{1/p} >\right) = P\left(X_1 > \frac{t}{p}\right) = P\left(X_1 > \left[\frac{t}{p}\right]\right)$$
$$= (1-p)^{\left[\frac{t}{p}\right]} = \left[(1-p)^{-\frac{1}{p}}\right]^{-p\left[\frac{t}{p}\right]} \to e^{-t}$$

   because

$$\lim_{p\to 0} -p\left[\frac{t}{p}\right] = \lim_{p\to 0} -p\left(\frac{t}{p} + \left[\frac{t}{p}\right] - \frac{t}{p}\right)$$
$$= -t + \lim_{p\to 0} p\left(\frac{t}{p} - \left[\frac{t}{p}\right]\right) = -t$$

   since $\left(\frac{t}{p} - \left[\frac{t}{p}\right]\right) \in [0,1]$

9.