Evaluating Contrarian Signals with Conditional Random Walk

Minh Le

May 2025

1 Problem

A coin is flipped n times with probability p of being heads. This experiment is repeated k times. We keep only those replications that have at least m heads after the first j flips. This selected subset of replications is used for all subsequent analyses.

1. Let N_H denote the total number of heads in n flips for a replication. Considering only the selected subset of replications (those with at least m heads in the first j flips):
   (a) What is the expected value of N_H?
   (b) Determine the probability distribution (i.e., the probability mass function) of N_H.
2. The overall process of n flips can be viewed as a 1D random walk. If we associate a Head (H) with a step of $+1$ and a Tail (T) with a step of -1, the position after k_{flips} flips is $S_{k_{flips}} = H_{k_{flips}} - T_{k_{flips}}$. Given that $H_{k_{flips}} + T_{k_{flips}} = k_{flips}$ (where $H_{k_{flips}}$ is the number of heads and $T_{k_{flips}}$ is the number of tails in k_{flips} flips), the position can also be expressed as $S_{k_{flips}} = 2H_{k_{flips}} - k_{flips}$. A "zero-crossing" event occurs at the first flip $k^* > j$ such that $S_{k^*} = 0$.
   (a) Prove that for a zero-crossing event to occur (i.e., $S_{k^*} = 0$), the flip number k^* must be an even integer.
   (b) Let $t^* = k^* - j$ denote the number of additional flips after the j-th flip until this event occurs. What is the probability of such a zero-crossing occurring after the j-th flip and up to (and including) the n-th flip (i.e., $j < k^* \leq n$)?
3. Now, further restrict the analysis. From the subset of replications selected based on the properties of the first j flips, consider only those that *also* exhibit such a zero-crossing (where $S_{k^*} = 0$ for the first time with $j < k^* \leq n$). For this specific group of iterations:
   (a) What is the expected value of t^* (the number of additional flips after j until this first zero-crossing)?
   (b) What is the distribution of t^*?

The random walk formulation presented, based on coin flips, offers a framework for modeling asset price movements and evaluating potential trading signals. In this analogy, a 'Head' signifies an upward price movement and a 'Tail' a downward one. The core condition of this study, selecting replications with at least m heads in the first j flips for all subsequent analyses, is akin to identifying an asset that has exhibited a strong initial directional signal, or an 'overextension'. If such an initial signal results in a significant positive displacement (where the position $S_j = 2H_j - j$ is high), a subsequent 'zero-crossing' event (where $S_{k^*} = 0$, with $k^* > j$) would signify a complete reversion of this initial trend. This scenario is of particular interest for contrarian investment strategies, which capitalize on the premise that strong market moves in one direction are often followed by a corrective move (i.e., mean reversion). The quantitative exploration undertaken in this project, specifically determining the probability of such a zero-crossing occurring after the j-th flip and the expected number of additional flips (t^*) until this event [see Problem 2b and 3a], can therefore provide valuable measures for assessing the potential viability and timing characteristics of contrarian positions taken against such initial market overextensions.

2 Solution

Let N_H be the total number of heads in n flips, $N_{H,j}$ be the number of heads in the first j flips, and $N_{H,n-j}$ be the number of heads in the flips from $j+1$ to n. Let p be the probability of heads for a single flip. The experiment considers only replications where the condition $C : N_{H,j} \geq m$ is met.

1. Expected Value and Distribution of Total Heads N_H

a. Expected number of heads after n flips

This calculation is independent of the random walk interpretation used for later parts. The expected number of heads after n flips, conditional on C, is:

$$E[N_H \mid C] = E[N_{H,j} \mid C] + E[N_{H,n-j} \mid C]$$

The number of heads in the $n-j$ flips after the first j flips ($N_{H,n-j}$) is independent of the first j flips. Thus, $E[N_{H,n-j} \mid C] = E[N_{H,n-j}] = (n-j)p$. The term $E[N_{H,j} \mid C]$ is the conditional expectation of a binomial random variable $X \sim B(j,p)$, given $X \geq m$:

$$E[N_{H,j} \mid N_{H,j} \geq m] = \frac{\sum_{x=m}^{j} x\binom{j}{x}p^x(1-p)^{j-x}}{\sum_{i=m}^{j} \binom{j}{i}p^i(1-p)^{j-i}}$$

Therefore, the total expected number of heads is:

$$E[N_H \mid N_{H,j} \geq m] = \frac{\sum_{x=m}^{j} x\binom{j}{x}p^x(1-p)^{j-x}}{\sum_{i=m}^{j} \binom{j}{i}p^i(1-p)^{j-i}} + (n-j)p$$

b. Distribution of the total number of heads after n flips

We want to find the probability mass function $P(N_H = h \mid C)$ for $h \in [m,n]$.
We have $N_H = N_{H,j} + N_{H,n-j}$.

$$P(N_H = h \mid C) = \frac{P(N_H = h \text{ and } C)}{P(C)}$$

The denominator is $P(C) = P(N_{H,j} \geq m) = \sum_{i=m}^{j} \binom{j}{i}p^i(1-p)^{j-i}$.
The numerator is $P(N_{H,j} + N_{H,n-j} = h \text{ and } N_{H,j} \geq m)$. Let $X = N_{H,j}$ and $Y = N_{H,n-j}$.

$$P(X+Y = h \text{ and } X \geq m) = \sum_x P(X = x \text{ and } Y = h-x \text{ and } x \geq m)$$

Since X and Y are independent, $P(X = x, Y = h-x) = P(X = x)P(Y = h-x)$. The sum is over x such that $m \leq x \leq j$ (for $X = x$ and $X \geq m$) and $0 \leq h-x \leq n-j$ (for $Y = h-x$). The conditions for x are: $\max(m, h-(n-j)) \leq x \leq \min(j,h)$.

$$\begin{aligned} P(X=x)P(Y=h-x) &= \left[\binom{j}{x}p^x(1-p)^{j-x}\right]\left[\binom{n-j}{h-x}p^{h-x}(1-p)^{(n-j)-(h-x)}\right] \\ &= \binom{j}{x}\binom{n-j}{h-x}p^h(1-p)^{n-h} \end{aligned}$$

So, the numerator becomes $p^h(1-p)^{n-h}\sum_{x=\max(m,h-(n-j))}^{\min(j,h)} \binom{j}{x}\binom{n-j}{h-x}$. Therefore, the conditional distribution of N_H is:

$$P(N_H = h \mid N_{H,j} \geq m) = \frac{p^h(1-p)^{n-h}\sum_{x=\max(m,h-(n-j))}^{\min(j,h)} \binom{j}{x}\binom{n-j}{h-x}}{\sum_{i=m}^{j} \binom{j}{i}p^i(1-p)^{j-i}}$$

This distribution is defined for $h \in [m,n]$.

2. Zero-Crossing Properties

As defined in Problem 2, the random walk position is $S_{k_{flips}} = 2H_{k_{flips}} - k_{flips}$. A zero-crossing event occurs at the first flip $k^* > j$ such that $S_{k^*} = 0$.

a. Proof that k^* for a zero-crossing must be even

Proof. *Let k^* be the flip number at which a zero-crossing event occurs. The state of the random walk at this flip is $S_{k^*} = H_{k^*} - T_{k^*}$, where H_{k^*} is the number of heads and T_{k^*} is the number of tails in the first k^* flips. The total number of flips is k^*, so: $H_{k^*} + T_{k^*} = k^*$.*

A zero-crossing event means $S_{k^*} = 0$. Thus, $H_{k^*} - T_{k^*} = 0$, which implies $H_{k^*} = T_{k^*}$. Substituting $T_{k^*} = H_{k^*}$ into the equation for the total number of flips: $H_{k^*} + H_{k^*} = k^*$ $2H_{k^*} = k^*$ Since H_{k^*} (the number of heads) must be an integer, k^* must be an integer multiple of 2. Therefore, k^* must be an even integer. □

b. Probability of a zero-crossing after the j-th flip

Let $t^* = k^* - j$ be the number of additional flips after the j-th flip until the zero-crossing. Since k^* must be even (from 2a), and $k^* = j + t^*$, then t^* must have the same parity as j. The maximum value for t^* is $n - j$.

The state of the walk after j flips is $S_j(x) = 2x - j$, where $x = N_{H,j}$. The analysis is conditional on $N_{H,j} \geq m$. Let $q(x)$ be the PMF of $N_{H,j}$ given this condition:

$$q(x) = P(N_{H,j} = x \mid N_{H,j} \geq m) = \frac{\binom{j}{x} p^x (1-p)^{j-x}}{\sum_{i=m}^{j} \binom{j}{i} p^i (1-p)^{j-i}} \quad \text{for } m \leq x \leq j$$

We seek the first $t^* \in [1, n-j]$ such that a random walk of t^* additional steps, starting effectively from $S_j(x)$, reaches overall state 0. This is equivalent to a standard 1D random walk starting at $Y_0 = 0$ that first reaches target state $A_x = -S_j(x) = j - 2x$ at step t^*.

Let $g(t', A)$ be the probability that such a 1D random walk, starting from 0, first reaches state A at step t'.

- If $A \neq 0$: $g(t', A) = \frac{|A|}{t'} \binom{t'}{(t'+A)/2} p^{(t'+A)/2} (1-p)^{(t'-A)/2}$. This is non-zero if $t' \geq |A|$, and t' and A have the same parity.
- If $A = 0$ (requiring j to be even, $x = j/2$, and $x \geq m$): $g(2N, 0) = \frac{2}{N} \binom{2N-2}{N-1} (p(1-p))^N$ for $N \geq 1$. $g(t', 0) = 0$ if t' is odd or $t' < 2$.

Let $P_Z(x)$ be the probability of a zero-crossing occurring with $1 \leq t^* \leq n - j$, given $N_{H,j} = x$:

$$P_Z(x) = \sum_{t^*=1}^{n-j} g(t^*, A_x) = \sum_{t^*=1}^{n-j} g(t^*, j - 2x)$$

The sum respects the conditions for $g(t^*, j-2x)$ to be non-zero. The overall probability of a zero-crossing, P_Z, is:

$$P_Z = \sum_{x=m}^{j} q(x) P_Z(x) = \sum_{x=m}^{j} q(x) \left(\sum_{t^*=1}^{n-j} g(t^*, j - 2x) \right)$$

If $P_Z = 0$, the conditional quantities below are not well-defined.

3. Analysis conditional on zero-crossing

We assume $P_Z > 0$ and consider only replications that exhibit a zero-crossing $S_{k^*} = 0$ for the first time with $j < k^* \leq n$.

a. Expected value of t^*

$$E[t^* \mid \text{crossing } j < k^* \leq n] = \frac{1}{P_Z} \sum_{x=m}^{j} q(x) \left(\sum_{t^*=1}^{n-j} t^* \cdot g(t^*, j - 2x) \right)$$

b. Distribution of t^*

For $\tau \in [1, n-j]$ (and τ having the same parity as j):

$$P(t^* = \tau \mid \text{crossing } j < k^* \leq n) = \frac{1}{P_Z} \sum_{x=m}^{j} q(x) \cdot g(\tau, j - 2x)$$

3 Theoretical Result vs. Empirical Simulation

Parameters: $k_{\text{replications}} = 700,000$; $n = 250$; $p = 0.5$; $m = 8$; $j = 10$.

1. Expected Value and Distribution of Total Heads N_H

a. Expected number of heads after n flips

$$E[N_{H,j} \mid N_{H,j} \geq 8] = \frac{460}{56} \approx 8.2143$$

$$(n - j)p = (250 - 10) \times 0.5 = 120$$

$$E[N_H \mid N_{H,j} \geq 8] \approx 8.2143 + 120 = 128.2143$$

Empirically Simulated Result:

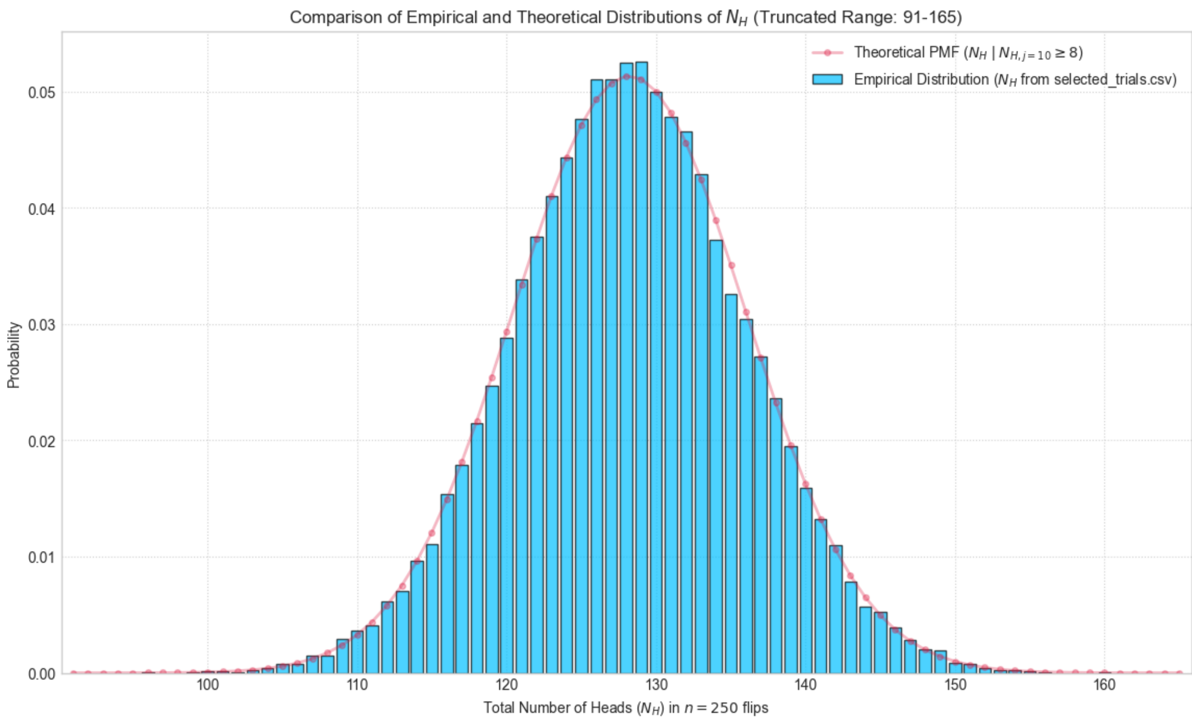
- Expected number of heads: 124.9968
- Variance of number of heads: 62.4472
- Standard deviation of number of heads: 7.9024
- 95% confidence interval for number of heads: (109.5082, 140.4854)

b. Distribution of the total number of heads after n flips For $h \in [8, 250]$:

$$P(N_H = h \mid N_{H,j} \geq 8) = \frac{(0.5)^{240}}{56} \sum_{x=\max(8,h-240)}^{\min(10,h)} \binom{10}{x}\binom{240}{h-x}$$

Example for $h = 8$: The sum is $\binom{10}{8}\binom{240}{0} = 45$.

$$P(N_H = 8 \mid N_{H,j} \geq 8) = \frac{(0.5)^{240} \times 45}{56}$$



2. Zero-Crossing Properties

a. Proof that k^* for a zero-crossing must be even: The proof provided in the general solution (Section 2.a) applies directly. Since $S_{k^*} = 2H_{k^*} - k^*$, if $S_{k^*} = 0$, then $k^* = 2H_{k^*}$, which must be even.

b. Probability of a zero-crossing P_Z: $q(8) = \frac{45}{56}$, $q(9) = \frac{10}{56}$, $q(10) = \frac{1}{56}$. $A_8 = -6$, $A_9 = -8$, $A_{10} = -10$. (All non-zero, so $g(t', 0)$ case is not used for these A_x).

$$g(t', A) = \frac{|A|}{t'} \binom{t'}{(t' + A)/2} (0.5)^{t'}$$

t^* must be even. Max $t^* = 240$.

$$P_Z(8) = \sum_{t^*=6, t^* \text{even}}^{240} \frac{6}{t^*} \binom{t^*}{(t^* - 6)/2} (0.5)^{t^*}$$

$$P_Z(9) = \sum_{t^*=8, t^* \text{even}}^{240} \frac{8}{t^*} \binom{t^*}{(t^* - 8)/2} (0.5)^{t^*}$$

$$P_Z(10) = \sum_{t^*=10, t^* \text{even}}^{240} \frac{10}{t^*} \binom{t^*}{(t^* - 10)/2} (0.5)^{t^*}$$

$$P_Z = q(8)P_Z(8) + q(9)P_Z(9) + q(10)P_Z(10)$$

These sums require numerical computation. P_Z is expected to be high.

Numerically Computed Theoretical Solution: Probability of zero-crossing $P_Z = 0.679436$

Empirically Simulated Result:

- Number of zero crossings: 25916
- Percentage of zero crossings: 67.94431481530032%
- Number of non-zero crossings: 12227
- Percentage of non-zero crossings: 32.05568518469968%

3. Analysis conditional on zero-crossing

a. Expected value of t^*

$$E[t^* \mid \text{crossing by 240 steps}] = \frac{1}{P_Z} \sum_{x=8}^{10} q(x) \left(\sum_{t^*=|A_x|, t^* \text{even}}^{240} t^* \cdot g(t^*, A_x) \right)$$

Requires numerical evaluation of P_Z and the inner sums.

Numerically Computed Theoretical Solution:
$E[t^* \mid \text{crossing}] = 65.7955 \Rightarrow E[k^* \mid \text{crossing}] = 75.7955$

Empirically Simulated Result:

- Expected value of the first-crossing-position: 75.47646241703967
- Variance of the first-crossing-position: 3287.6610835805836
- Standard deviation of the first-crossing-position: 57.33812940426801
- 95% confidence interval for the expected value: (-36.91, 187.86)

b. Distribution of t^* For $\tau \in [1, 240]$ (τ even):

$$P(t^* = \tau \mid \text{crossing by 240 steps}) = \frac{1}{P_Z} \sum_{x=8}^{10} q(x) \cdot g(\tau, A_x)$$

Example for $\tau = 10$: Numerator $N(\tau = 10) \approx 0.022599$.

$$P(t^* = 10 \mid \text{crossing by 240 steps}) \approx \frac{0.022599}{P_Z}$$