

Coin Toss Experiment

Mariam Lomishvili

Summary:

I simulated 10,000 fair coin tosses in MATLAB (2,500 rows and 4 columns). For each row of four tosses I counted the number of heads (0-4) and compared the empirical distribution with the theoretical Binomial(4, 0.5). Over the entire 10,000-toss sequence I computed run lengths of heads (1 in a row, 2 in a row, ...) and compared the empirical(experimented) histogram and CDF with the correct theoretical model.

Figures

Figure 1 — Groups of 4: PDF (histogram)

Empirical bars for $k = 0-4$ with Binomial(4, 0.5) curve overlay

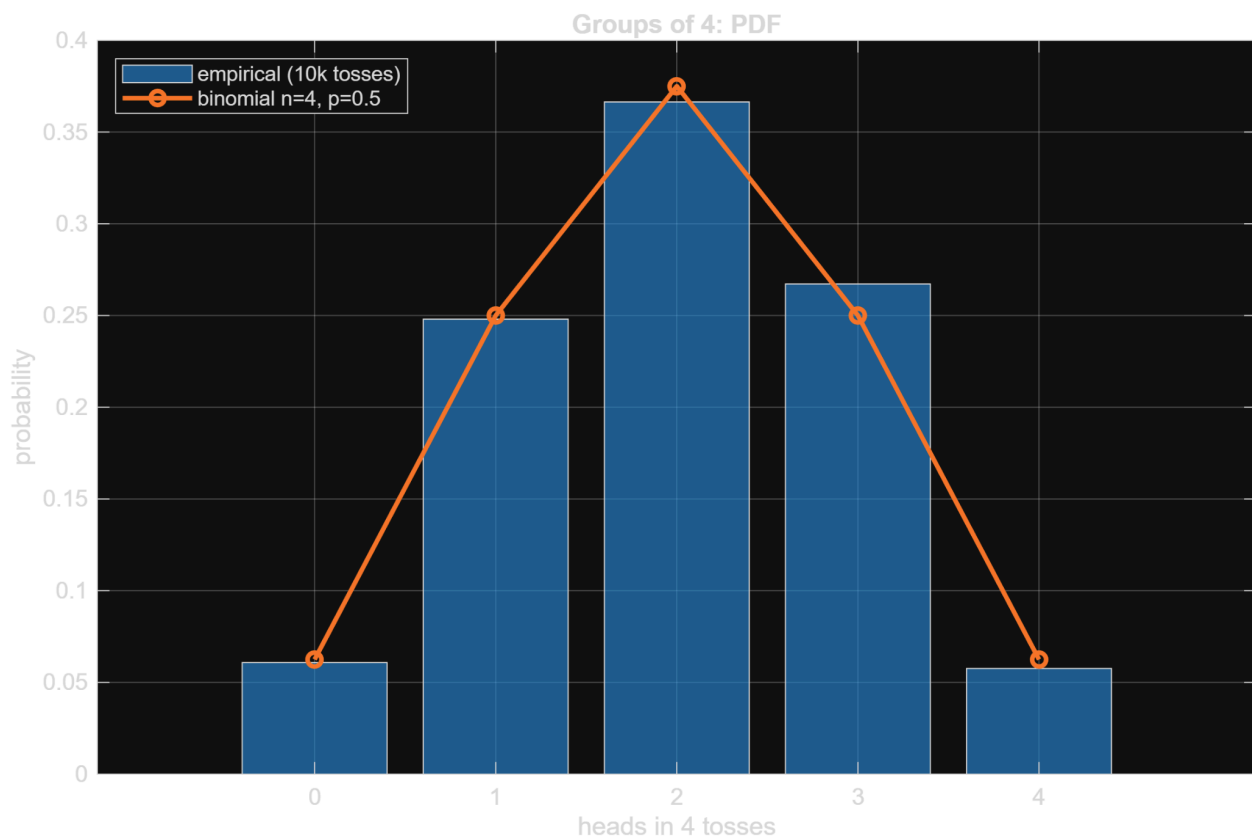


Figure 2 — Groups of 4: CDF

Empirical CDF steps vs binomial CDF points.

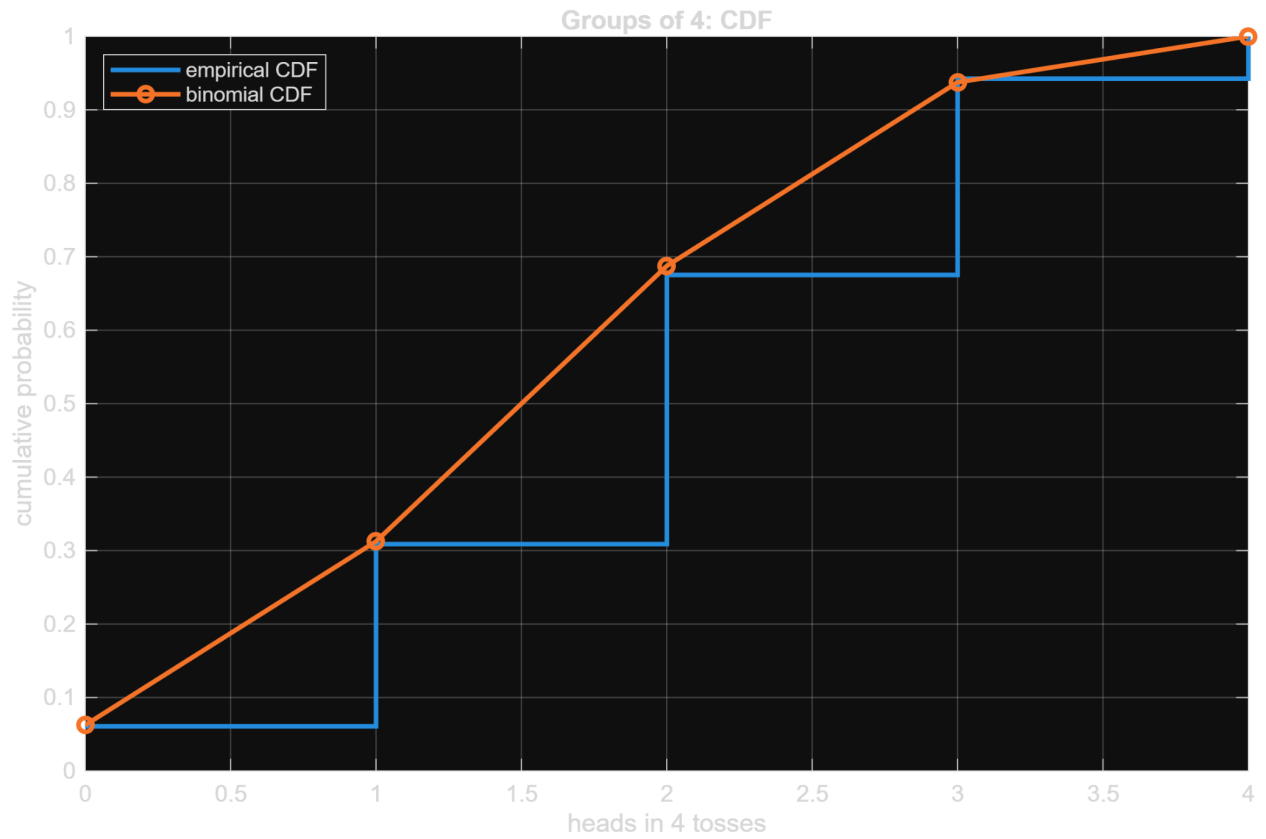


Figure 3 — Head Run-Lengths: PDF (histogram)

Empirical bars for run length $k \geq 1$ with geometric overlay $P(L=k) = (1-p)p^{k-1}$, $p=0.5$.

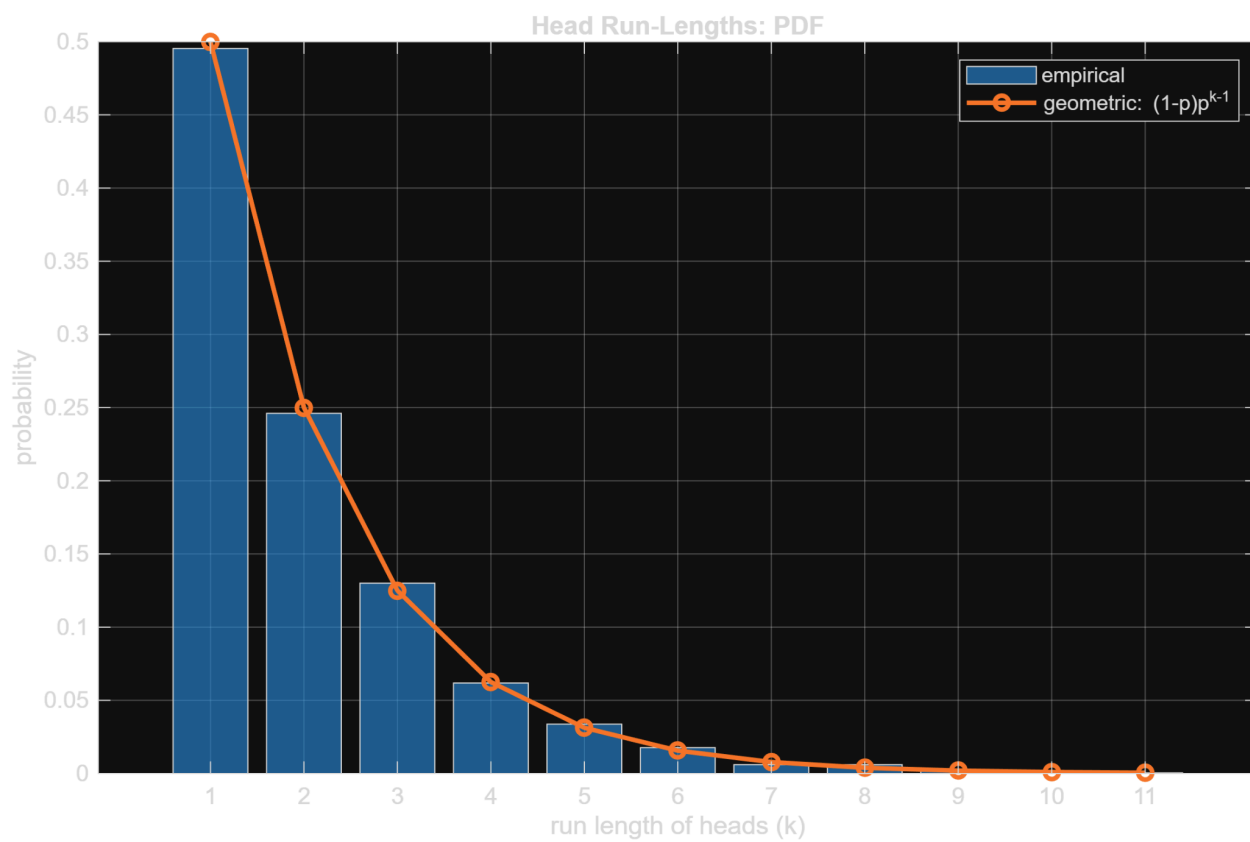
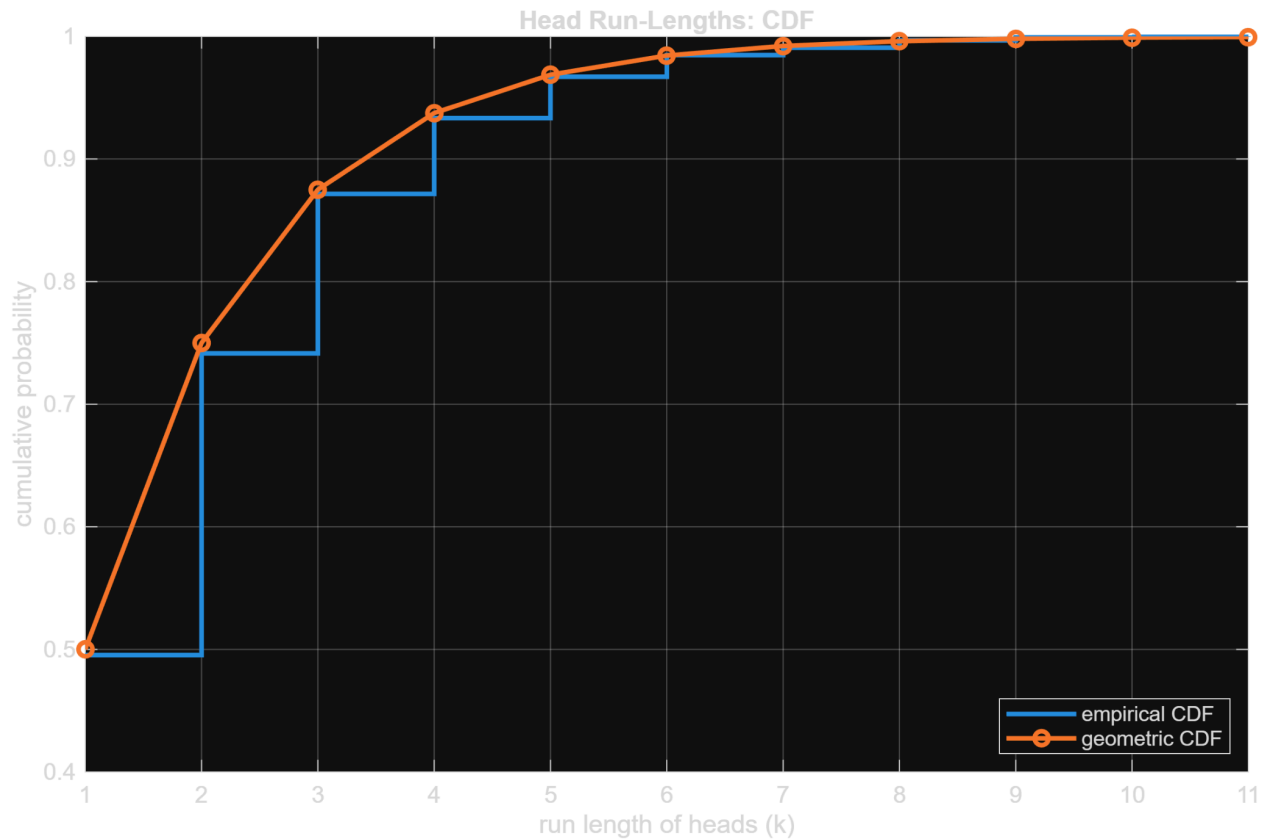


Figure 4 — Head Run-Lengths: CDF
Empirical CDF for run lengths vs geometric CDF $1-p^k$, $p=0.5$



A sample size much larger than 100 tosses is needed for a smooth approximation. with 10,000 tosses the histogram aligns with Binomial(4, 0.5). The run-length distribution is not binomial. when counted per run it follows a geometric law on $k \geq 1$ because a head-run continues with probability p and ends with a tail with probability $1-p$. The figures show what was expected.

MATLAB code:

```
clear; close all; clc %this was crucial

% basics

num_toss_total = 10000; % should be >= 10000 (so >= 2500 sets of 4)

p_head = 0.5; % fair coin

rng(1729); % not totally random so it is reproducible

% output folder with timestamps so i can make out which run it is
```

```

tstamp = datestr(now, 'yyyymmdd_HHMMSS');

out_dir_tag = ['coin_outputs_' tstamp];

if ~exist(out_dir_tag, 'dir'); mkdir(out_dir_tag); end

% generate tosses

num_sets4 = floor(num_toss_total/4);

coinroll_stream = rand(1, num_sets4*4) < p_head; % 1=head, 0=tail, <p_head
makes bools

% groups of 4 analysis

grp4_mat = reshape(coinroll_stream, 4, num_sets4); % reshape to be 4xN blocks

grp4_headsums = sum(grp4_mat, 1); % 0-4 heads per block

% empirical part

edges4 = -0.5:1:4.5; %centers bins on integers

cnt_emp_blk4 = histcounts(grp4_headsums, edges4); %count

pmf_emp_blk4 = cnt_emp_blk4 / num_sets4; %count to PMF

cdf_emp_blk4 = cumsum(pmf_emp_blk4); %cdf

% theoretical (binomial) part for n=4

%pmf(k) = C(4,k) * p^k * (1-p)^(4-k)

k4 = 0:4;

pmf_binom4 = arrayfun(@(k) nchoosek(4,k) * p_head.^k * (1-p_head).^(4-k), k4);
%PMF

cdf_binom4 = cumsum(pmf_binom4); %CDF

% PDF plot: empirical vs theoretical

f1 = figure('Color', 'w');

bar(k4, pmf_emp_blk4, 'FaceAlpha', 0.6); hold on

plot(k4, pmf_binom4, '-o', 'LineWidth', 2)

xlabel('heads in 4 tosses'); ylabel('probability')

legend('empirical (10k tosses)', 'binomial n=4, p=0.5', 'Location', 'northwest')

```

```

title('Groups of 4: PDF'); grid on

exportgraphics(f1, fullfile(out_dir_tag, 'groups_of_4_pdf.png'),
'Resolution',300)

% CDF plot: empirical vs theoretical

f2 = figure('Color','w');

stairs(k4, cdf_emp_blk4, 'LineWidth', 2); hold on

plot(k4, cdf_binom4, '-o', 'LineWidth', 2)

xlabel('heads in 4 tosses'); ylabel('cumulative probability')

legend('empirical CDF','binomial CDF','Location','northwest')

title('Groups of 4: CDF'); grid on

exportgraphics(f2, fullfile(out_dir_tag, 'groups_of_4_cdf.png'),
'Resolution',300)

% counts vs expected counts ( extra)

f5 = figure('Color','w');

bar(k4, cnt_emp_blk4, 'FaceAlpha', 0.6); hold on

exp_counts = pmf_binom4 * num_sets4;

plot(k4, exp_counts, '-o', 'LineWidth', 2)

xlabel('heads in 4 tosses'); ylabel('count per N sets')

legend('observed counts','expected counts (binomial)','Location','northwest')

title('Groups of 4: Counts vs Expected'); grid on

exportgraphics(f5, fullfile(out_dir_tag, 'groups_of_4_counts_vs_expected.png'),
'Resolution',300)

% run-lengths of heads over the entire stream (NOT ROWS OF 4)

runlens_heads = get_run_lengths(coinroll_stream); % k>=1 for heads runs

%histogram run lengths over k=1-Kmax

Kmax = max(runlens_heads);

edgesR = 0.5:1:(Kmax+0.5); %integer centers

```

```

%convert run lentgths fo PMF and CDF

%i kept naming some variables PMF, others PDF...

cnt_emp_runs = histcounts(runlens_heads, edgesR);

kR = 1:Kmax;

pmf_emp_runs = cnt_emp_runs / sum(cnt_emp_runs);

cdf_emp_runs = cumsum(pmf_emp_runs);

% theoretical model for run lengths of heads

%  $P(L=k) = (1-p) * p^{(k-1)}$ ,  $k=1,2,3,\dots$ ,  $p = P(\text{head})$ .

pmf_geom_runs = (1 - p_head) * (p_head).^(kR - 1);

cdf_geom_runs = 1 - (p_head).^kR;

% PDF plot for run lengths

f3 = figure('Color','w');

bar(kR, pmf_emp_runs, 'FaceAlpha',0.6); hold on

plot(kR, pmf_geom_runs, '-o', 'LineWidth', 2)

xlabel('run length of heads (k)'); ylabel('probability')

legend('empirical','geometric:  $(1-p)p^{k-1}$ ','Location','northeast')

title('Head Run-Lengths: PDF'); grid on

exportgraphics(f3, fullfile(out_dir_tag,'runlength_heads_pdf.png'),
'Resolution',300)

% CDF plot for run lengths

f4 = figure('Color','w');

stairs(kR, cdf_emp_runs, 'LineWidth', 2); hold on

plot(kR, cdf_geom_runs, '-o', 'LineWidth', 2)

xlabel('run length of heads (k)'); ylabel('cumulative probability')

legend('empirical CDF','geometric CDF','Location','southeast')

title('Head Run-Lengths: CDF'); grid on

```

```

exportgraphics(f4, fullfile(out_dir_tag,'runlength_heads_cdf.png'),
'Resolution',300)

% save raw data too, just in case!!

save(fullfile(out_dir_tag,'sim_data.mat'), ...

'coinroll_stream','grp4_headsums','cnt_emp_blk4','pmf_emp_blk4','cdf_emp_blk4',
...

'runlens_heads','cnt_emp_runs','pmf_emp_runs','cdf_emp_runs', ...

'p_head','num_sets4','num_toss_total');

disp('saved figures+data to folder:'); disp(out_dir_tag)

% HELPER FUNCTIONS

function rlen = get_run_lengths(bitstream)

    x = bitstream(:)'; % 0/1 row

    d = diff([0 x 0]); % mark edges: prepend and append a 0 so that a run at the
very beginning or end is handled correctly

    run_starts = find(d == 1);

    run_ends    = find(d == -1) - 1;

    rlen = run_ends - run_starts + 1; % positive lengths only

    rlen = rlen(rlen > 0);

end

```