# Theme 3: What is Probability & what is useful for

- Homicides
- Random var. & distributions
- Mathematics for data
- Simulation
- Bernoulli & Poisson distributions

AoS

CHAPTER 8

# Probability – the Language of Uncertainty and Variability

In 1650s France, the self-styled Chevalier de Méré had a gambling problem. It was not that he gambled too much (although he did), but he wanted to know which of two games he stood the greatest chance of winning -

?

- Game 1: Throw a fair die at most four times, and win if you get a six.
- Game 2: Throw two fair dice at most twenty-four times, and win if you get a double-six.

Which was his better bet?

Following good empirical statistical principles, the Chevalier de Méré decided to play both games numerous times and see how often he won. This took a great deal of time and effort, but in a bizarre parallel universe in which there were computers but no probability theory, the good Chevalier (real name Antoine Gombaud) would not have wasted his time collecting data on his successes – he would simply have simulated thousands of games.

Figure 8.1 displays the results of such a simulation, showing how the overall proportion of times that he wins each game changes as he 'plays' more and more. Although Game 2 looks the better bet for a while, after around 400 games of each it becomes clear that Game 1 is better, and in the (very) long run he can expect to win around 52% of Game 1, and only 49% of Game 2.

DATA WE COLLECTED

$$x_i = \begin{cases} 1, & \text{WIN } i\text{TH GAME} \\ 0, & \text{OTHERWISE} \end{cases} \qquad i = 1, \dots, n$$

$$\sum_{i=1}^{n} x_i \quad : \text{ FREQ OF WINS UP TO GAME } n$$

$$y_n = \frac{1}{n} \sum_{i=1}^{n} x_i \quad : \text{ PERC " " " } \dots$$

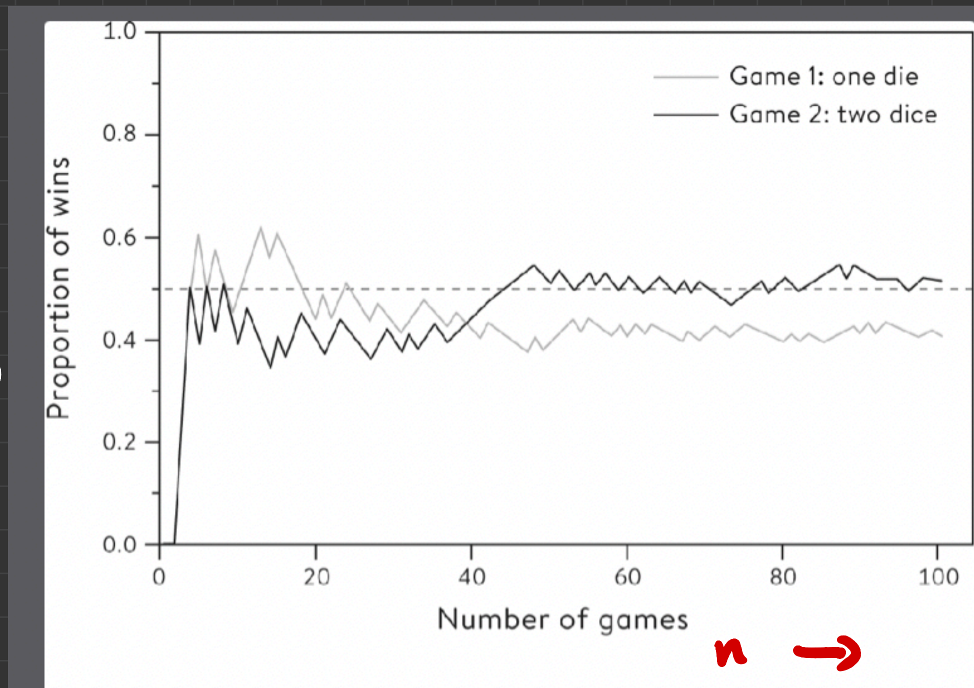
COMPUTER SIMULATION

↳ (PSEUDO) RANDOM NUMBERS

A.S

$y_n$

MANIFESTATION OF RANDOMNESS




$n \rightarrow$

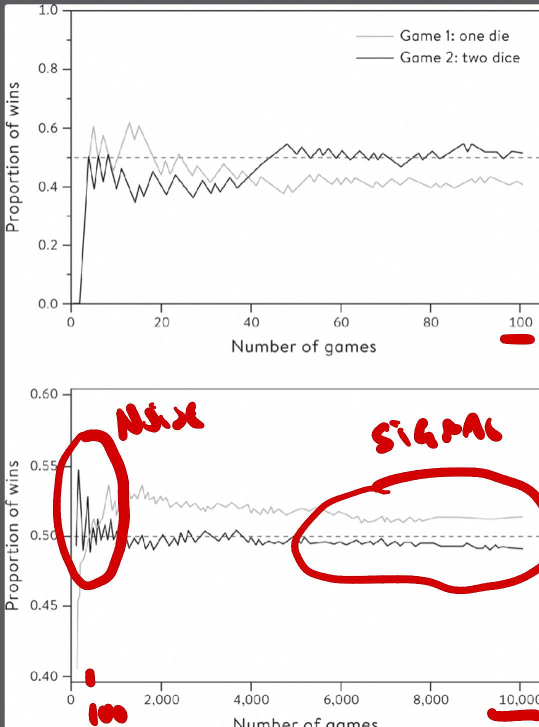



**Figure 8.1**
A computer simulation of 10,000 repeats of two games. In Game 1, you win if you throw a six in at most 4 throws of a fair die; in Game 2, you win if you throw a double-six in at most 24

- EFFECT OF RANDOMNESS
- CONCEPT OF "PROBABILITY" AS LONG-RUN PERCENTAGE

"PROBABILITY" AS LIMITING LONG-RUN PERCENTAGE

# MODELS & DATA

MODELS : PROBABILITIES / DISTRIBUTIONS

DATA : RANDOM VARIABLES

- BINARY VARIABLES

UPPER CASE $X$ : RANDOM VARIABLE

MODELS THE UNCERTAIN OUTCOME

BEFORE IT IS REALIZED

LOWER CASE $x$ : REALIZATION OF $X$ WHEN IT HAPPENS

## BERNOULLI MODEL / DISTRIBUTION

$$X \underset{\text{"FOLLOWS"}}{\sim} \text{Bern}(\mu) \qquad \mu \in (0,1)$$

$$P_X(x) = \begin{cases} \mu & x=1 \\ 1-\mu & x=0 \end{cases}$$

A R.V. $X$ FOLLOWS THE BERNOULLI MODEL/DISTR.

WITH PARAMETER $\mu$

## OBSERVATIONS :

i. $p_X(0) + p_X(1) = 1$

$\mu + (1-\mu) = 1$

ii. $p_X(x) \geq 0$

$\{0, 1\}$ : SET WITH ELEMENTS 0 & 1

$(0, 1)$ : INTERVAL OF ALL REAL NUMBERS BETWEEN 0 & 1

$x \in \{0, 1\}$ $x$ IS AN ELEMENT OF THE SET

↳ "BELONGS TO"

$a^0 = 1$ BY CONVENTION FOR ANY $a$

iii A MATHEMATICALLY MORE CONVENIENT WAY TO WRITE THE MODEL

$$P_X(x) = \mu^{x}(1-\mu)^{1-x} \quad , \quad x \in \{0,1\}, \ \mu \in (0,1)$$

BERNOULLI PROBABILITY MASS FUNCTION