# Mathematics of Uncertainty

Lecture 2

# Topics

- Probability on finite sample spaces
- Conditional probability
- Law of total probability
- Independence

# Formal definitions

A ***probability space*** is defined by a triplet $(\mathbf{S}, \mathbf{F}, \mathbf{P})$ and satisfy the following laws of probability:

**1)** **S** is the *sample space*, a non-empty set of all possible outcomes

**1)** **F** is a family of events, a class of subsets of **S**, satisfying the following conditions:
- $\mathbf{S} \in \mathbf{F}$ (there is at least one event to talk about)
- If an event $A \in \mathbf{F}$, then $A^c \in \mathbf{F}$ (if *A* is an event, then the complement of *A* is also an event)
- If all of $A_1, A_2, \ldots \in \mathbf{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathbf{F}$ (if *A* and *B* are both events, then $A \cup B$ is also be an event)

Such a class of sets **F** is called **σ-field** on **S**. The σ-field of subsets of the real line is called a **Borel** σ-field.

1) The *probability set function* **P**, defined on **F**, assigns numbers to events, $\mathbf{P}: \mathbf{F} \rightarrow [0,1]$ and we set
- $\mathbf{P}(\mathbf{S}) = 1$ (the sample space includes all events that might happen)
- For all sequences of pairwise disjoint (i.e. $A_i \cap A_j = \emptyset$ for $i \neq j$) events $A_1, A_2, A_3 \ldots$, we assume that $0 \leq \mathbf{P}(A_i) \leq 1$ (other values are meaningless) and $\mathbf{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$ (***countable additivity***)

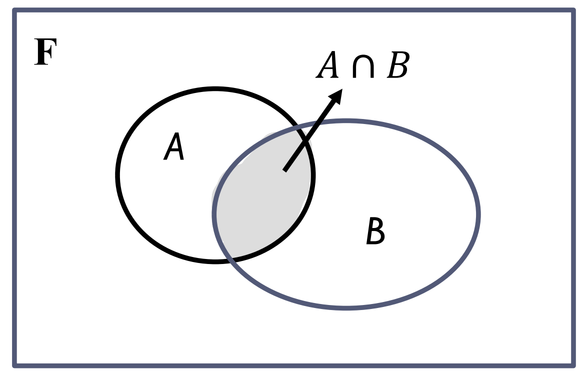
A function satisfying these properties is called a ***probability measure***.

# Formal definitions

Keeping the set of events **F** constant, we can define a new probability function $\mathbf{Q}: \mathbf{F} \to [0,1]$, again satisfying $\mathbf{Q}(\mathbf{S}) = 1$ and countable additivity. Such a change is called ***change of probability measure***. The change of measure is a useful trick often used to value derivative assets and complex financial contracts.

In financial computations, we will not use set functions **P** and deal with only real-valued functions $Pr(A)$. The three laws of probability imply:

- $Pr(A^c) = 1 - Pr(A)$
- If $A \subseteq B$, then $Pr(A) \leq Pr(B)$ and $Pr(B \backslash A) = Pr(B) - Pr(A)$
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$
- Then, $Pr(A \cup B) \leq Pr(A) + Pr(B)$ (***Boole's inequality***)



# Examples to clarify

**Example**: Tossing a coin once (assuming it does not land on its edge!), **S** = {H, T} and **F** = {{H}, {T}, {H,T}, ∅}. A proper probability measure on **S** can be $Pr(H) = p$ and $Pr(T) = 1 - p$.

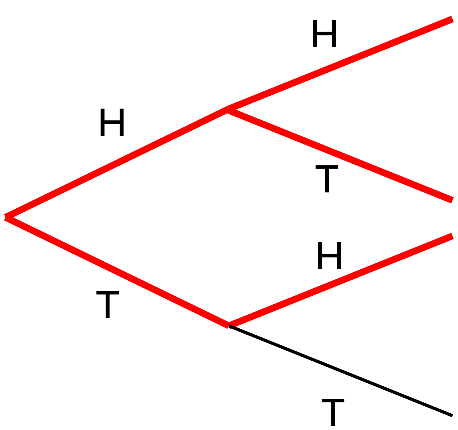
**Example**: Tossing a coin twice, **S** = {HT, HH, TH, TT} and **F** = $\{$∅, **S**, {HH}, {TT}, {HT}, {TH}, {HT, TH}, {HH, TT}, {HH, HT}, {HH, TH}, {TT, TH}, {TT, HT}, {TT, HT, TH}, {HH, HT, TH}$\}$

**Example**: Suppose BIST sets the minimum price tick as $\pm 0.05\, TL$. If the current price of a stock is 20, the sample space **S** = {19.50, 19.55, ... ,20.45, 20.50} will describe possible prices after 10 ticks. **F** would contain all possible combinations of up-ticks, down-ticks, and zero-ticks, where the total number of ticks is 10.

# Examples to clarify

**Example**: The probability of the event $A = \{\text{at least one head}\}$ in 2 tosses of a fair coin

$$Pr(A) = 1 - Pr(A^c) = 1 - Pr(\text{no H}) = 1 - Pr(\text{TT}) = 1 - (\frac{1}{2} \times \frac{1}{2}) = \frac{3}{4}$$




$Pr(\text{HH}) = 1/4$

$Pr(\text{HT}) = 1/4$

$Pr(\text{TH}) = 1/4$

$Pr(\text{TT}) = 1/4$

$sum = 1$

$\mathbf{S} = \{HH, HT, TH, TT\}$

$\mathbf{A} = \{HH, HT, TH\}$

# Gambling

Players select 6 different numbers from the list {1, 2, 3, . . . , 49}

- The sample space includes $\binom{49}{6} = \frac{49!}{6! \times 43!} = 13{,}983{,}816$ different orders of 6 numbers.
- One ticket then has a probability of $\frac{1}{13{,}983{,}816} = 7.15 \times 10^{-8}$ of matching the winning number.
- There are $\binom{6}{k}$ ways of matching $k$ of the particular 6 winning numbers. The remaining $(6 - k)$ numbers will be the losing numbers and there are $\binom{43}{6-k}$ ways of getting these losing numbers. As a result, the probability of matching $k$ numbers in this lottery is $\frac{\binom{6}{k} \times \binom{43}{6-k}}{\binom{49}{6}}$

| k | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Pr | 0.43596 | 0.41302 | 0.13238 | 0.01765 | 0.00097 | $1.84 \times 10^{-5}$ | $7.15 \times 10^{-8}$ |

Note that $\Pr(k \leq 2) > 0.98$. That is, more than 98% of the tickets will not match more than 2 of the 6 winning numbers! Then… why play?

# Binomial probabilities

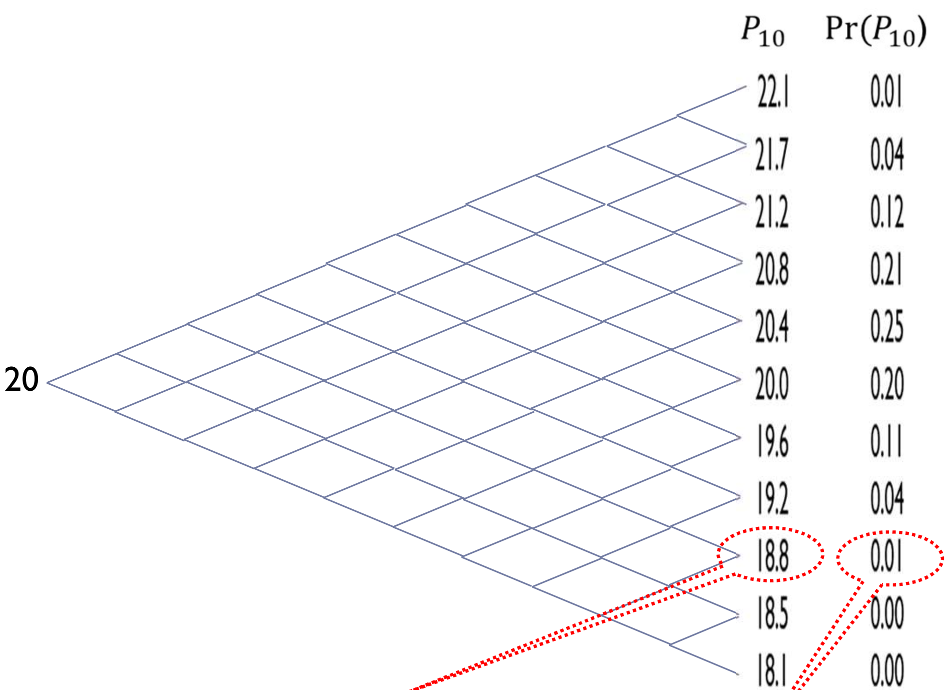
**Example:** To prevent negative prices, define one price tick as $\pm 1\%$, and ignore zero-ticks. Starting with an initial price of $P_0 = 20$, the sample space after 10 ticks is $S = \{20(1.01)^k(0.99)^{10-k}; k = 0,1,\ldots,10\}$, where $k$ is the number of up-ticks and $10 - k$ is the number of down-ticks. The number of possible price paths is given by the binomial coefficients:

$$\binom{10}{k} = \frac{10!}{k!\,(10-k)!}$$

Letting $p$ be the probability of an up-tick and $1 - p$ be that of a down-tick, the probability of the price level after 10 ticks is given by $\Pr\left(P_{10} = 20(1.01)^k(0.99)^{10-k}\right) = \binom{10}{k} p^k(1-p)^{10-k}$. This is called the ***binomial probability*** law.

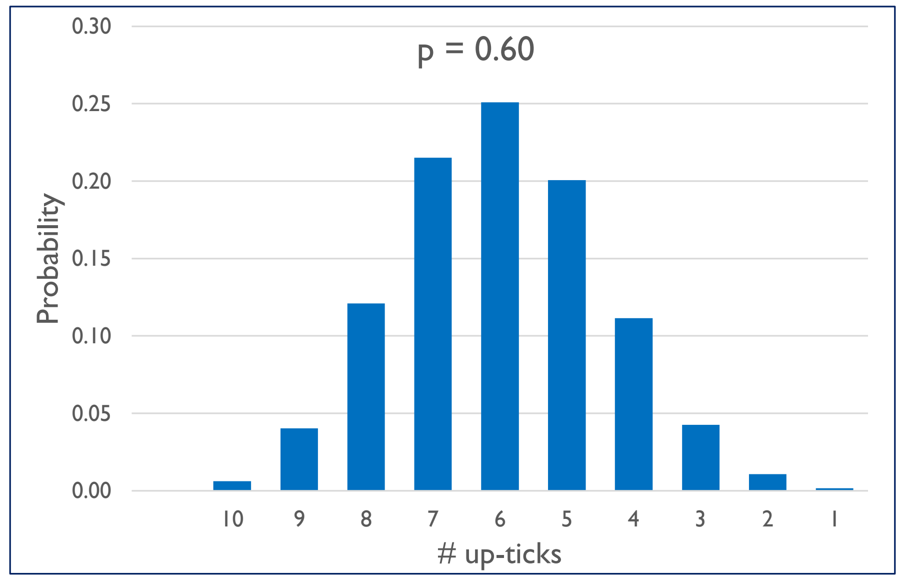
Note that $\sum_{k=0}^{10} \binom{10}{k} p^k(1-p)^{10-k} = 1$, satisfying the laws of probability.

# Binomial probabilities




$20 \times 1.01^2 \times 0.99^8$

$\binom{10}{2} 0.6^2 0.4^8$

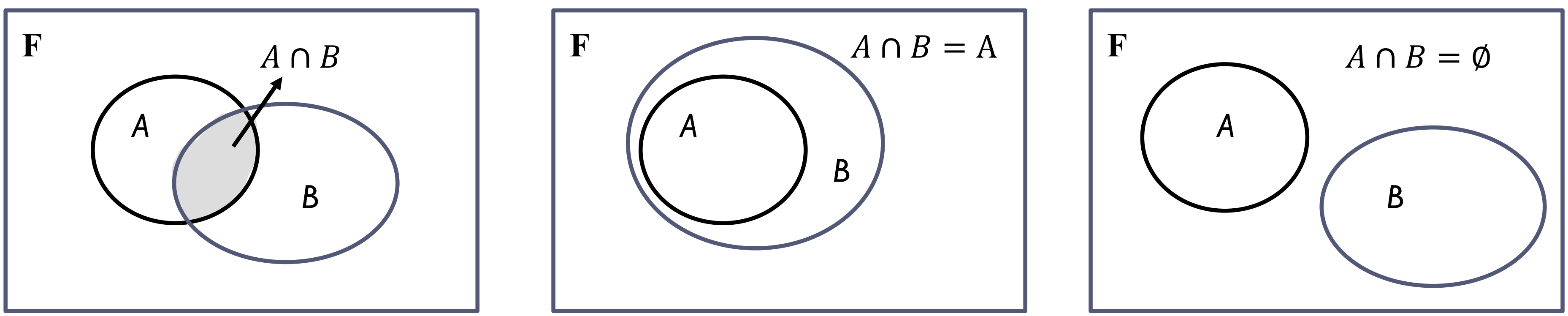


# Conditional probability

Given a probability space $(S, F, P)$, let two events $\{A, B\} \in F$ with $Pr(B) > 0$. Provided $Pr(A \cap B) > 0$, the *conditional probability of A, given B*, is defined as

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

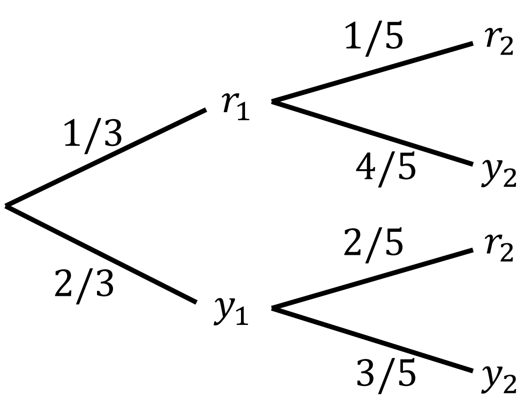
Any conditional probability is also an ordinary probability, satisfying all three laws of probability.



# Example

Draw a ball randomly: $\Pr(r_1) = 1/3$ and $\Pr(y_1) = 2/3$

*Without replacement*, draw a second ball: $\Pr(r_2|r_1) = 1/5$, $\Pr(y_2|r_1) = 4/5$ etc.




$$\Pr(r_1 \cap r_2) = \left(\frac{1}{3}\right)\left(\frac{1}{5}\right) = \frac{1}{15}$$

$$\Pr(r_1 \cap y_2) = \left(\frac{1}{3}\right)\left(\frac{4}{5}\right) = \frac{4}{15}$$

$$\Pr(y_1 \cap r_2) = \left(\frac{2}{3}\right)\left(\frac{2}{5}\right) = \frac{4}{15}$$

$$\Pr(y_1 \cap y_2) = \left(\frac{2}{3}\right)\left(\frac{3}{5}\right) = \frac{6}{15}$$

$$\Pr(y_2) = \Pr(r_1 \cap y_2) + \Pr(y_1 \cap y_2) = \frac{10}{15}$$

$$\Pr(r_1 \mid y_2) = \frac{\Pr(r_1 \cap y_2)}{\Pr(y_2)} = \frac{4/15}{10/15} = \frac{2}{5}$$

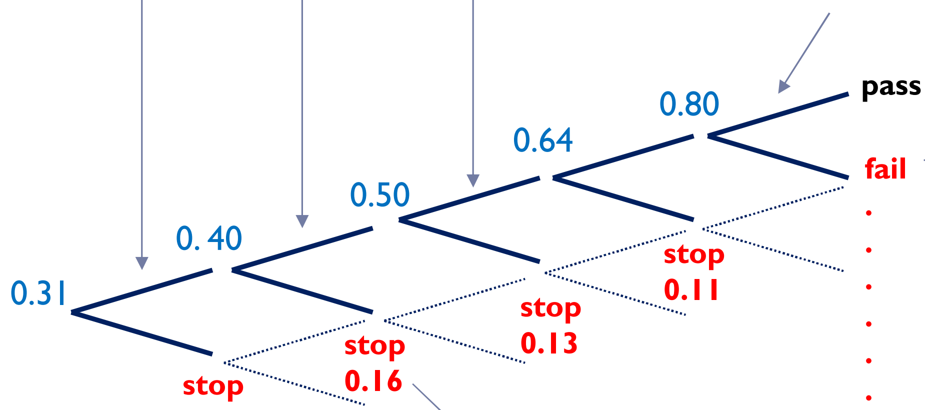
Note that $\Pr(r_1 \mid y_2) > \Pr(r_1)\ldots$ why?

# Example

An exam consists of 5 questions randomly selected from a list of 50 questions, and you know the answers to 40 of these. What is the probability of passing the exam if you have to answer all five questions?

Define $A_i$ $(i = 1, \dots, 5)$ as question $i$ being one of the 40 you know, then, by successive conditioning,

$$Pr\left(\cap_{i=1}^{5} A_i\right) = Pr(A_1)\, Pr(A_2 | A_1)\, Pr(A_3 | A_1 \cap A_2) \ldots Pr\left(A_n \middle| \cap_{i=1}^{4} A_i\right) = \frac{40}{50} \times \frac{39}{49} \times \frac{38}{48} \times \frac{37}{47} \times \frac{36}{46} = 0.31$$




$$Prob = \frac{40}{50} \times \frac{39}{49} \times \frac{38}{48} \times \frac{37}{47} \times \frac{10}{46} = 0.09$$

$$\frac{40}{50} \times \frac{10}{49}$$

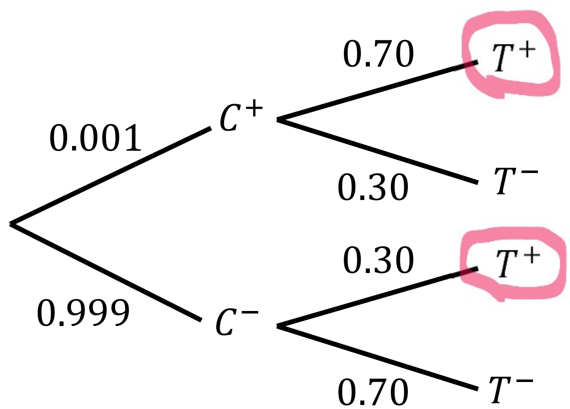
***Filtration***: sequence of revealed information

(As new information comes, we revise our believes)

# Example

- Chances of catching Covid19 is one out of 1,000 people: $\Pr(C^{+}) = 0.001$
- Accuracy of Covid19 PCR test: $\Pr(T^{+} \mid C^{+}) = \Pr(T^{-} \mid C^{-}) = 0.70$

If you test positive, what is the probability that you actually have Covid19: $\Pr(C^{+} \mid T^{+})$=?




$\Pr(C^{+} \cap T^{+}) = 0.001 \times 0.70 = 0.0007$

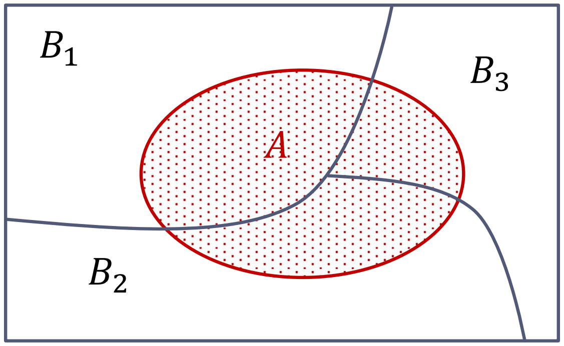
$\Pr(C^{-} \cap T^{+}) = 0.999 \times 0.30 = 0.2997$

$$\Pr(C^{+} \mid T^{+}) = \frac{\Pr(C^{+} \cap T^{+})}{\Pr(T^{+})} = \frac{0.0007}{0.0007 + 0.299} = 0.0023 \neq 0.70$$

# Law of total probability

Let $\{B_1, B_2, \ldots\}$ be pairwise disjoint events in **S** such that $\bigcup_i B_i = \mathbf{S}$, then for any event $A$

$$Pr(A) = \sum_i Pr(A \cap B_i) = \sum_i Pr(A|B_i)\,Pr(B_i)$$




**Example:** A football player has a good game two times in three. Empirically, his probability of scoring is 3/4 in a good game and 1/4 in a bad game.

1. What is the probability he scores in a game? By the law of total probability,

$$Pr(score) = \Pr(score|good)\Pr(good) + \Pr(score|bad)\Pr(bad) = (3/4)(2/3) + (1/4)(1/3) = 7/12$$

2. Given that he has scored, what is the probability he had a good game? Using conditional probabilities,

$$Pr(good|score) = \frac{\Pr(good \cap score)}{\Pr(score)} = \frac{(2/3)(3/4)}{7/12} = 6/7$$

# Bayes – Laplace rule of learning

Given the ***prior probabilities*** $\{Pr(B_j)\}$ , the law of total probability calculates the probability of an event $A$ by conditioning on $\{B_j\}$: $B_j \xrightarrow{caus} A$. After $A$ occurs, we want to revise our prior believes by changing the order of conditioning: $A \xrightarrow{infer} B_j$. This is done as

$$Pr(B_j|A) = \frac{Pr(A|B_j)\,Pr(B_j)}{Pr(A)} = \frac{Pr(A|B_j)\,Pr(B_j)}{\sum_i Pr(A|B_i)\,Pr(B_i)}$$

where the numerator is the conditional probability formula and the denominator is the total law formula. The revised probabilities $\{Pr(B_j|A)\}$ are called the ***posterior probabilities***.

**Example**: In the Covid case, the prior scenario was that $\Pr(C^+) = 0.001$ and $\Pr(T^+ \mid C^+) = 0.70$. Then, conditioning on the new information of a positive test, the Bayes formula revealed that

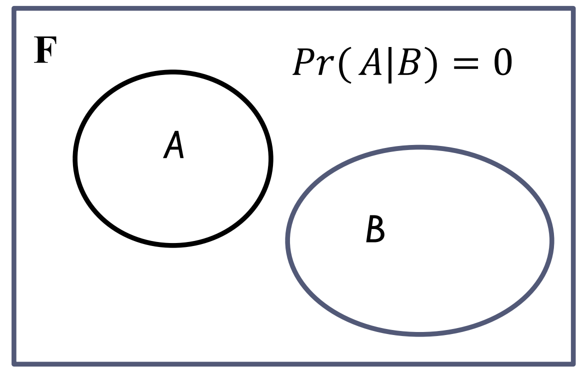
$$\Pr(C^+ \mid T^+) = \frac{\Pr(T^+ \mid C^+)\Pr(C^+)}{\Pr(T^+ \mid C^+)\Pr(C^+) + \Pr(T^+ \mid C^-)\Pr(C^-)} = 0.0023$$

# Independence

Definition: $Pr(A|B) = Pr(A)$, meaning that $A$ and $B$ are unrelated (physically or numerically)

Better definition: $A$ and $B$ are ***independent*** when $Pr(A \cap B) = Pr(A)\,Pr(B)$.

- Independence and disjointness (mutual exclusivity) are not the same things. For example, A and B are disjoint but dependent.




- If A, B, and C are pairwise independent, it is possible that $Pr(A \cap B \cap C) \neq Pr(A)\,Pr(B)\,Pr(C)$
- Sometimes, dependent events $\{A, B\}$ may be reformulated as ***conditionally independent,*** by conditioning on event ***C:***

$$Pr(A \cap B|C) = Pr(A|C)\,Pr(B|C)$$

(This trick is useful in getting ***martingales*** in asset prices.)

- The opposite is also possible: $Pr(A \cap B) = Pr(A)\,Pr(B)$ but $Pr(A \cap B|C) \neq Pr(A|C)\,Pr(B|C)$

# Stock prices

In finance, rates of return ($R_t = (P_t - P_{t-1})/P_{t-1}$ or $R_t = ln( P_t/P_{t-1})$) are often modeled as random variables. A classical question is whether the "probability law" generating $R_t$ is independent from past realizations $\{R_s : s < t\}$. There are two approaches:

- ***Technical analysis*** is based on the belief that past realizations of prices contain "usable information" to predict future prices: that is, rates of return over time are not independent. The probability law conditional on past returns is better for prediction purposes.

- ***Efficient markets*** approach claims that, if tomorrow's price were predictable today, it would become unpredictable through trade today. Hence, rates of return over non-overlapping units of time must be statistically independent.

***Let the data speak for itself!***

# Words to conclude

- $\Pr(event) = 0$ does <u>not</u> imply that it is impossible.
- $\Pr(event) = 1$ does <u>not</u> imply absolute certainty.
  - "***Knightian uncertainty***" is when popular probability models fail to capture existing uncertainties in financial markets … today?
- Bayesian way of "learning" may be useful in big data analytics and machine learning algorithms