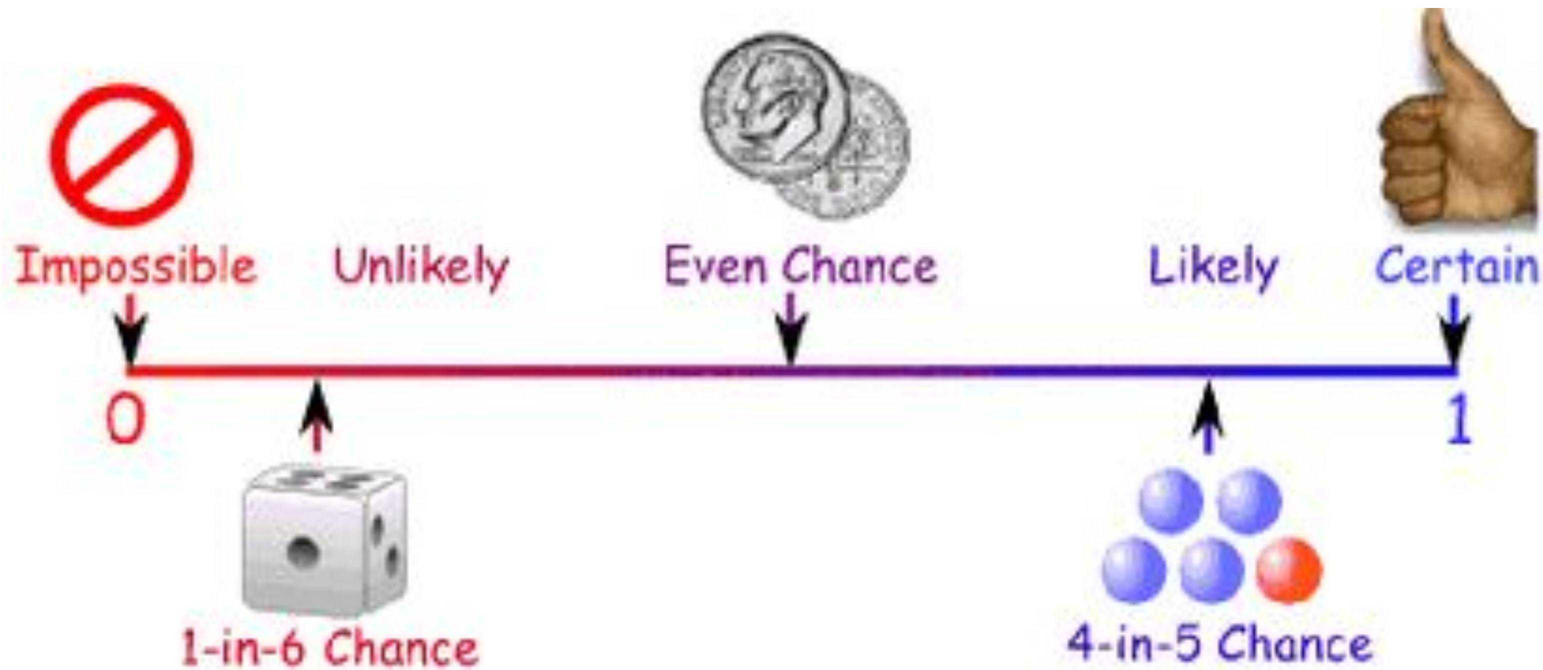
Artificial Intelligence
CC -721

Probability



Reasoning Under Uncertainty

- The knowledge-based systems we discussed used logic to perform certain inference.
- All conclusions were reached with certainty.
- In many domains, such as medical diagnosis, we are uncertain about our conclusions.
- For example, we may just conclude that it is very likely that the patient has a strep throat.

- Researchers initially tried to handle uncertainty by staying within the rule-based framework.
- The MYCIN system was designed in the 1970s to diagnose bacterial infection while reasoning under uncertainty.
- MYCIN used the rule-based approach while handling uncertainty with “Certainty Factors.”

- In the 1980s it was realized there were serious problems with this approach.
- Probability-based methods using Bayesian networks were developed.
- I will present the Bayesian networks approach.
- I start by reviewing probability.

·A probability function P is a function satisfying the following conditions:

We have a set of elementary events

$$\Omega = \{e_1, e_2, \cdots, e_n\}.$$

1. $$0 \le P(e_i) \le 1$$

2. $$P(e_1) + P(e_2) + \cdots + P(e_n) = 1$$

3. The probability of a subset is the sum of the probabilities of its members.

$$\mathrm{E} = \{e_3, e_6, e_8\}$$

$$P(\mathrm{E}) = P(e_3) + P(e_6) + P(e_8).$$

Example

$$\Omega = \{\text{heads}, \text{tails}\}$$

$$P(\text{heads}) = P(\text{tails}) = .5$$

Works for a symmetrical coin.

What if we are flipping a thumbtack?

Relative Frequency Approach to Probability

S_n is number of heads in n tosses.

$$P(\text{heads}) \equiv \lim_{n\to\infty} \frac{S_n}{n}.$$

3761 heads in 10,000 tosses:

$$P(\text{heads}) \approx \frac{3761}{10{,}000} = .3761.$$

Another Example

$$\Omega = 52\, cards$$

$$P(\text{jack of hearts}) = \frac{1}{52}$$

$$P(\text{Jack}) = P(\text{jack of hearts}) + P(\text{jack of clubs}) +$$
$$P(\text{jack of spades}) + P(\text{jack of diamonds})$$
$$= \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52} = \frac{1}{13}$$

Conditional Probability

The conditional probability of *E* given *F*, written

$$P(E|F),$$

is the probability of *E* given that *F* occurs.

$$P(\texttt{Jack}) = \frac{4}{52} = \frac{1}{13}$$

$$P(\texttt{Jack} \mid \texttt{RoyalCard}) = \frac{4}{12} = \frac{1}{3}$$

Independence

If

$$P(E|F) = P(E)$$

then E is independent of F.

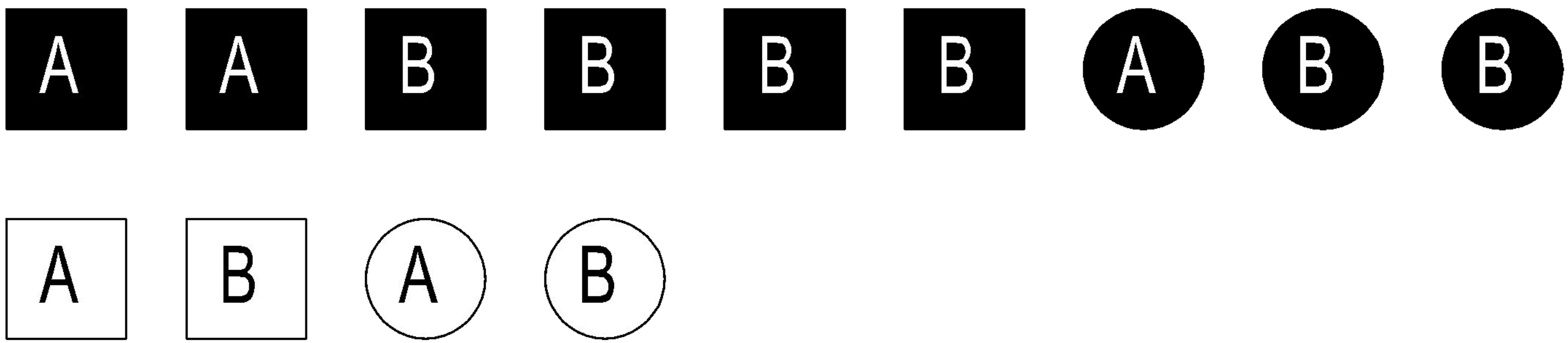
$$P(\texttt{Jack}) = \frac{4}{52} = \frac{1}{13}$$

$$P(\texttt{Jack}\,|\,\texttt{Club}) = \frac{1}{13}.$$

Conditional Independence

Two events E and F are conditionally independent given G if

$$P(E|F \cap G) = P(E|G).$$

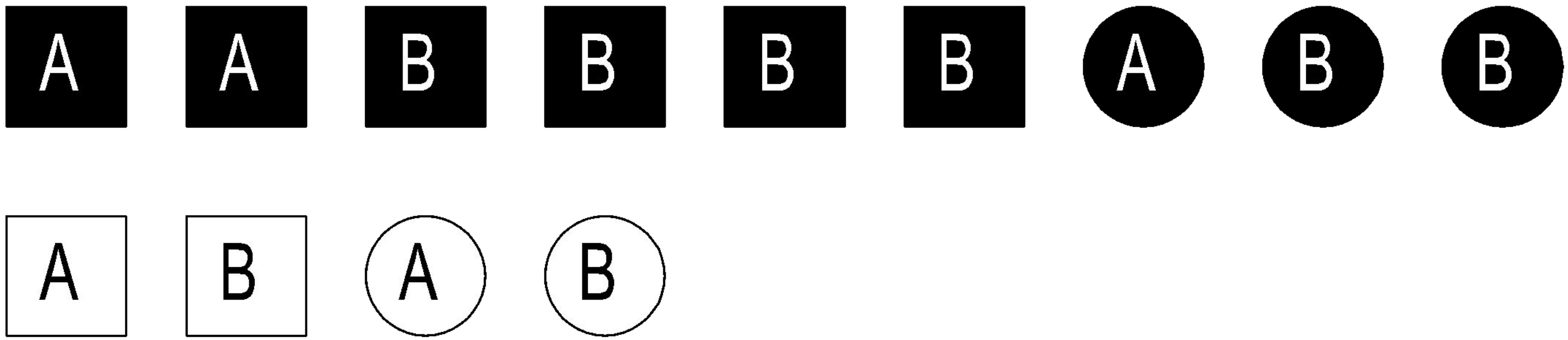


$$P(\mathtt{A}) = \frac{5}{13}$$

$$P(\mathtt{A}\,|\,\mathtt{Square}) = \frac{3}{8}$$

$$P(\mathtt{A}\,|\,\mathtt{Black}) = \frac{3}{9} = \frac{1}{3}$$

$$P(\mathtt{A}\,|\,\mathtt{Square} \cap \mathtt{Black}) = \frac{2}{6} = \frac{1}{3}$$



$$P(\mathtt{A}\,|\,\mathtt{White}) = \frac{2}{4} = \frac{1}{2}$$

$$P(\mathtt{A}\,|\,\mathtt{Square} \cap \mathtt{White}) = \frac{1}{2}.$$

Random Variables

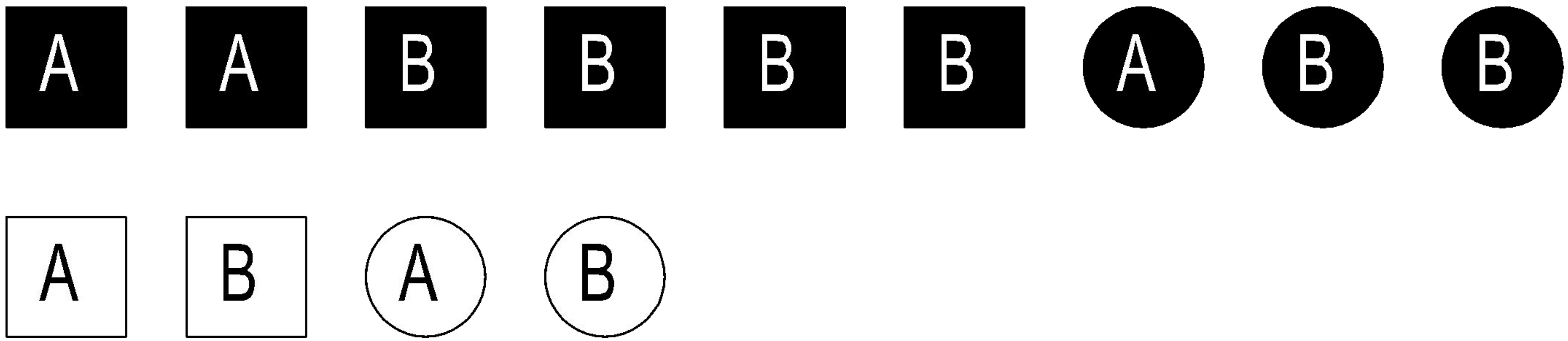
$P(\text{Suit} = \text{Club}) = \frac{1}{4}$

$P(\text{Suit} = \text{Heart}) = \frac{1}{4}$

$P(\text{Suit} = \text{Spade}) = \frac{1}{4}$

$P(\text{Suit} = \text{Club}) = \frac{1}{4}$

Suit is called a random variable.



Let

C be a random variable for color:

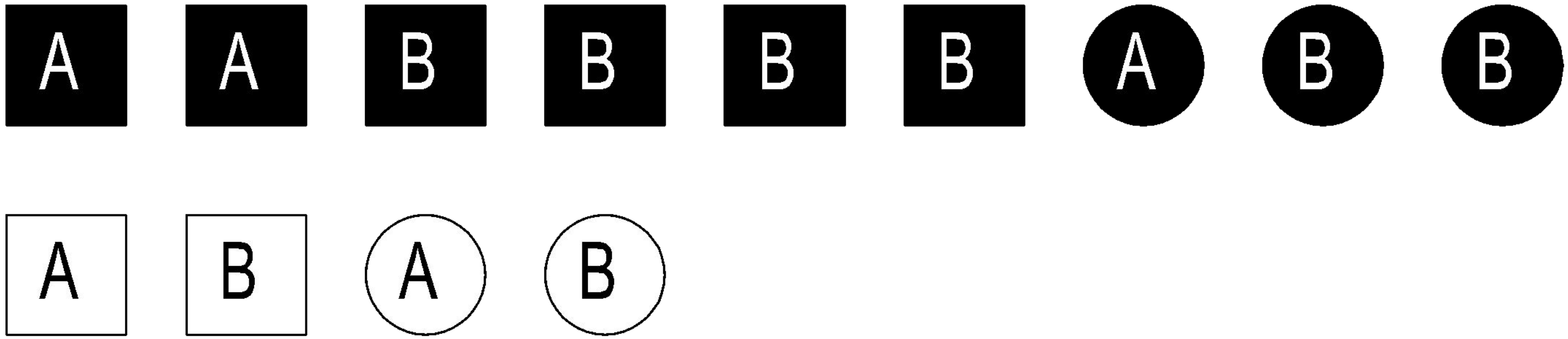
C = *black* or *white*

S be a random variable for shape:

S = *square* or *circle*

V be a random variable for value:

V = *A* or *B*



The random variables *V* and *S* are conditionally independent given the random variable *C*.

That is, for all values of *V*, *S*, and *C*

$$P(V|S,C) = \mathrm{P}(V|C).$$

$$P(V = \mathrm{A}|S = \mathrm{square}, C = \mathrm{black}) = P(V = \mathrm{A}|C = \mathrm{black}) = 1/3.$$

We write

$$I(V,S|C).$$

Case	Gender	Height	Wage
1	female	64	30,000
2	female	64	30,000
3	female	64	40,000
4	female	64	40,000
5	female	68	30,000
6	female	68	40,000
7	male	64	40,000
8	male	64	50,000
9	male	68	40,000
10	male	68	50,000
11	male	70	40,000
12	male	70	50,000

Wage and Height are not independent.

$P(W = 30) = 1/4$

$P(W = 30{,}000 \mid H = 64) = 1/3$

$P(W = 30{,}000 \mid H = 68) = 1/4$

$P(W = 30{,}000 \mid H = 70) = 0$

Case	Gender	Height	Wage
1	female	64	30,000
2	female	64	30,000
3	female	64	40,000
4	female	64	40,000
5	female	68	30,000
6	female	68	40,000
7	male	64	40,000
8	male	64	50,000
9	male	68	40,000
10	male	68	50,000
11	male	70	40,000
12	male	70	50,000

Wage and Height are conditionally independent given Gender.

$$P(W = 30 | G = \text{female}) = 1/2$$

$$P(W = 30 \mid H = 64, G = \text{female}) = 1/2$$

$$P(W = 30 \mid H = 68, G = \text{female}) = 1/2$$

$$\mathrm{I}(W,H \mid G)$$

Bayes' Theorem

Theorem: Given two events E and F,

$$P(\mathrm{E} \mid \mathrm{F}) = \frac{P(\mathrm{F} \mid \mathrm{E})P(\mathrm{E})}{P(\mathrm{F} \mid \mathrm{E})P(\mathrm{E}) + P(\mathrm{F} \mid \bar{\mathrm{E}})P(\bar{\mathrm{E}})}.$$

To prove Bayes' Theorem, we first prove the following:

$$P(\mathtt{E} \mid \mathtt{F}) = \frac{P(\mathtt{F} \mid \mathtt{E})P(\mathtt{E})}{P(\mathtt{F})} \tag{1}$$

Owing to the definition of conditional probability

$$P(\mathtt{E} \mid \mathtt{F}) = \frac{P(\mathtt{E} \cap \mathtt{F})}{P(\mathtt{F})} \quad \text{and} \quad P(\mathtt{F} \mid \mathtt{E}) = \frac{P(\mathtt{F} \cap \mathtt{E})}{P(\mathtt{E})}.$$

We then have that

$$P(\mathtt{E} \mid \mathtt{F})P(\mathtt{F}) = P(\mathtt{F} \mid \mathtt{E})P(\mathtt{E}) \tag{2}$$

because they both equal $P(\mathtt{E} \cap \mathtt{F})$.

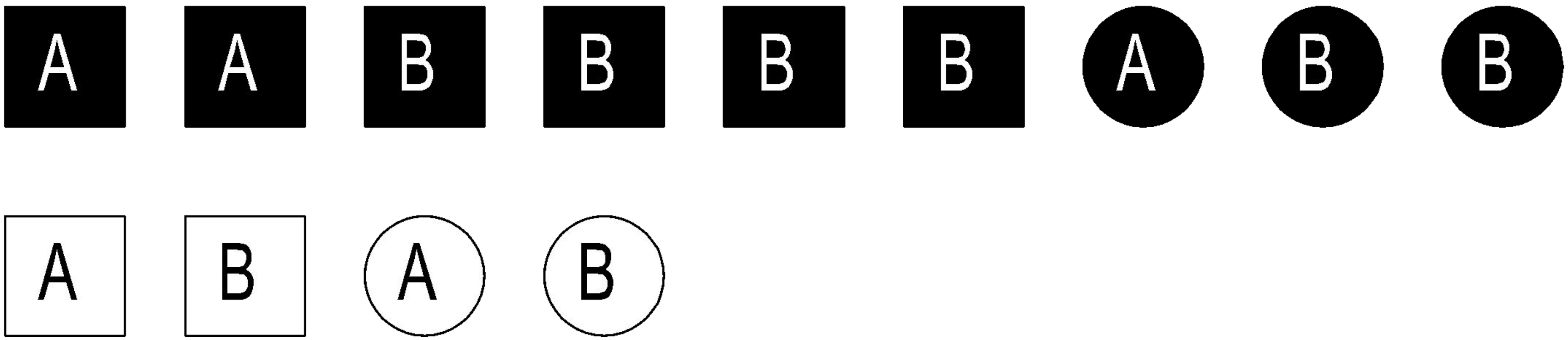
The proof of Equality 1 is now completed by dividing Equality 2 by

$$P(\mathtt{F}).$$

It is left as an exercise to complete the proof of Bayes' Theorem by proving the law of total probability, which is

$$P(\mathbf{F}) = P(\mathrm{F} \mid \mathrm{E})P(\mathrm{E}) + P(\mathrm{F} \mid \overline{\mathrm{E}})P(\overline{\mathrm{E}}).$$

Example



$$P(\texttt{Black} \mid \texttt{A}) = \frac{3}{5}$$

$$P(\texttt{Black} \mid \texttt{A}) = \frac{P(\texttt{A} \mid \texttt{Black})P(\texttt{Black})}{P(\texttt{A} \mid \texttt{Black})P(\texttt{Black}) + P(\texttt{A} \mid \texttt{White})P(\texttt{White})}$$

$$= \frac{\left(\frac{1}{3}\right)\left(\frac{9}{13}\right)}{\left(\frac{1}{3}\right)\left(\frac{9}{13}\right) + \left(\frac{1}{2}\right)\left(\frac{4}{13}\right)} = \frac{3}{5}$$

- The state of Indiana requires everyone who wants to get married to take a routine blood test for HIV called ELISA.
- Joe takes the test.
- It comes back positive.
- Does Joe have HIV?

$$P(ELISA = positive \mid HIV = present) = .999$$

$$P(ELISA = positive \mid HIV = absent) = .002$$

$$P(HIV = present) = .00001$$

$$P(present \mid positive)$$

$$= \frac{P(positive \mid present)P(present)}{P(positive \mid present)P(present) + P(positive \mid absent)P(absent)}$$

$$= \frac{(.999)(.00001)}{(.999)(.00001) + (.002)(.99999)}$$

$$= .00497.$$

- A certain pregnancy test has the exact same accuracy as ELISA.
- Mary takes that test.
- It comes back positive.
- Is Mary pregnant?

$$P(test = positive \mid pregnant = yes) = .999$$

$$P(test = positive \mid pregnant = no) = .002$$

$$P(pregnant = yes) = .2$$

$$
\begin{aligned}
&P(yes \mid positive) \\
&= \frac{P(positive \mid yes)P(yes)}{P(positive \mid yes)P(yes) + P(positive \mid no)P(no)} \\
&= \frac{(.99)(.2)}{(.99)(.2) + (.02)(.8)} \\
&= .92523.
\end{aligned}
$$

Meaning of Probability

There are two predominant approaches to probability:

1. The frequentist approach

    - Egon Pearson (1895 - 1980)
    - Richard von Mises (1883 - 1953)
    - Ronald A. Fisher (1890 -1962)
    - Jerzy Neyman (1894 - 1981)

2. The Bayesian approach

    - Karl Pearson (1857 - 1936)
    - Frank Ramsey (1903 - 1930)
    - W.E. Johnson (1858 - 1931)
    - Bruno de Finetti (1906 - 1985)
    - Jimmy Savage (1917 - 1971)
    - Dennis Lindley (1923 -)
    - Harold Jeffreys (1891 - 1989)
    - I.J. Good (1916 - 2009)

Frequentist View

$$p = \lim_{n \longrightarrow \infty} \# \text{ heads } / \ n$$

Bayesian view

I will give you \$60 if the Bears lose if you will give me \$40 if they win. I will take either side of this bet.

$p_{Neo} = 0.6.$

- Introducing Bayesian probability with a sporting event clearly demonstrates that it applies to events that are not repeatable.
- However, the use of this domain seems to trivialize the subject.
- Actually, many if not most uncertain events are not repeatable.
- E.g. what will happen in the stock market in the coming month?

Hypothesis Testing Under the Two Approaches

- In statistical hypothesis testing we update our belief about some hypothesis based on data.

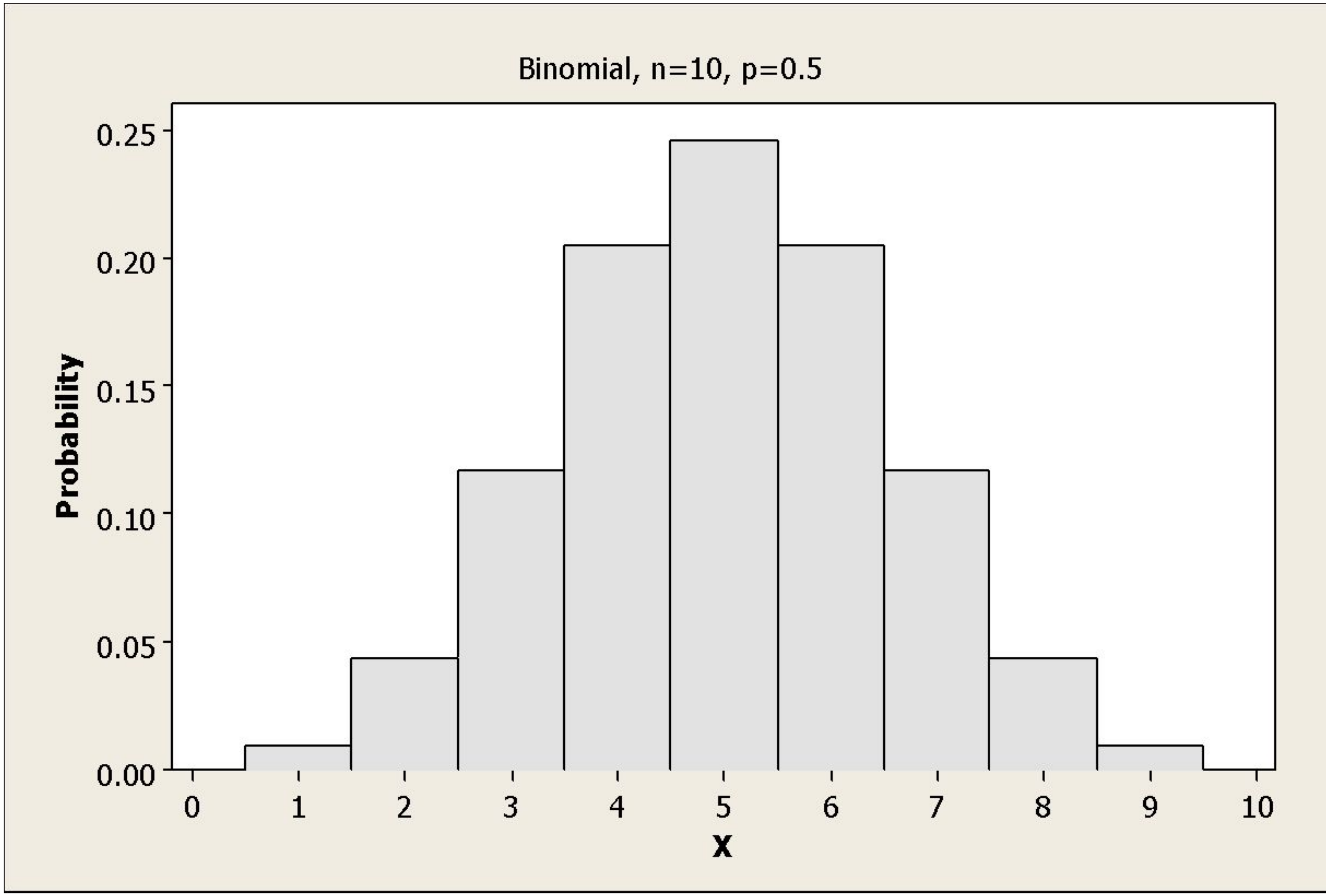
- Next we show how the two approaches handle hypothesis testing.

Frequentist Approach

Take a coin from your pocket and toss it 10 times.

Suppose it lands heads every time.

Should we reject the hypothesis that it is fair?



Probability of getting 10 heads if the coin is fair is 0.0009.

Called the *significance* or *p*-value of the result.

Perform a Randomized Control Experiment to test a new drug.

	Reduced Blood Pressure	No Reduction in Blood Pressure
Blood Pressure X	70	30
Placebo	51	49

Using a chi-square test the probability of getting this result or something more extreme if there is no difference in the two groups is 0.006.

Bayesian Approach

- Bayesian statistics involves updating prior belief based on Bayes' Theorem.
- Suppose Joe is about to get married, and he takes a routine blood test for HIV.
- The test comes back positive.

$$P(ELISA = positive \mid HIV = absent) = .002$$ significance

$$P(ELISA = positive \mid HIV = present) = .999$$ power

$$P(HIV = present) = .00001$$ prior probability

$$P(present \mid positive)$$

$$= \frac{P(positive \mid present)P(present)}{P(positive \mid present)P(present) + P(positive \mid absent)P(absent)}$$

$$= \frac{(.999)(.00001)}{(.999)(.00001) + (.002)(.99999)}$$

$$= .00497.$$

- In the early 20th century forefathers of current statistical methodology were largely Bayesians (e.g. Karl Pearson and R.A. Fisher).

- The Bayesian approach "depends upon an arbitrary assumption, so the whole method has been widely discredited."

  – R.A. Fisher, 1921

- "Inverse probability, which like an impenetrable jungle arrests progress towards precision of statistical concepts."

  – R.A. Fisher, 1922