Probability Basics

16-385 Computer Vision (Kris Kitani)

Carnegie Mellon University

Random Variable

What is it?

Is it ‘random’?

Is it a ‘variable’?

Random Variable

What is it?

Is it ‘random’?

not in the traditional sense

Is it a ‘variable’?

not in the traditional sense

Random Variable:

a variable whose possible values are numerical outcomes of a random phenomenon

http://www.stat.yale.edu/Courses/1997-98/101/ranvar.htm

Random variable:

a measurable function from a probability space into a measurable space known as the state space (Doob 1996)

http://mathworld.wolfram.com/RandomVariable.html

Random variable:

a function that associates a unique numerical value with every outcome of an experiment

http://www.stats.gla.ac.uk/steps/glossary/probability_distributions.html

$$\text{outcome} \xrightarrow{\text{random variable}} \text{value}$$

(index)

outcome
(face of a penny)

random variable

value
(heads or tails)

0: heads

1: tails

What kind of random variable is this?

outcome
(face of a penny)

random variable

value
(heads or tails)

0: heads

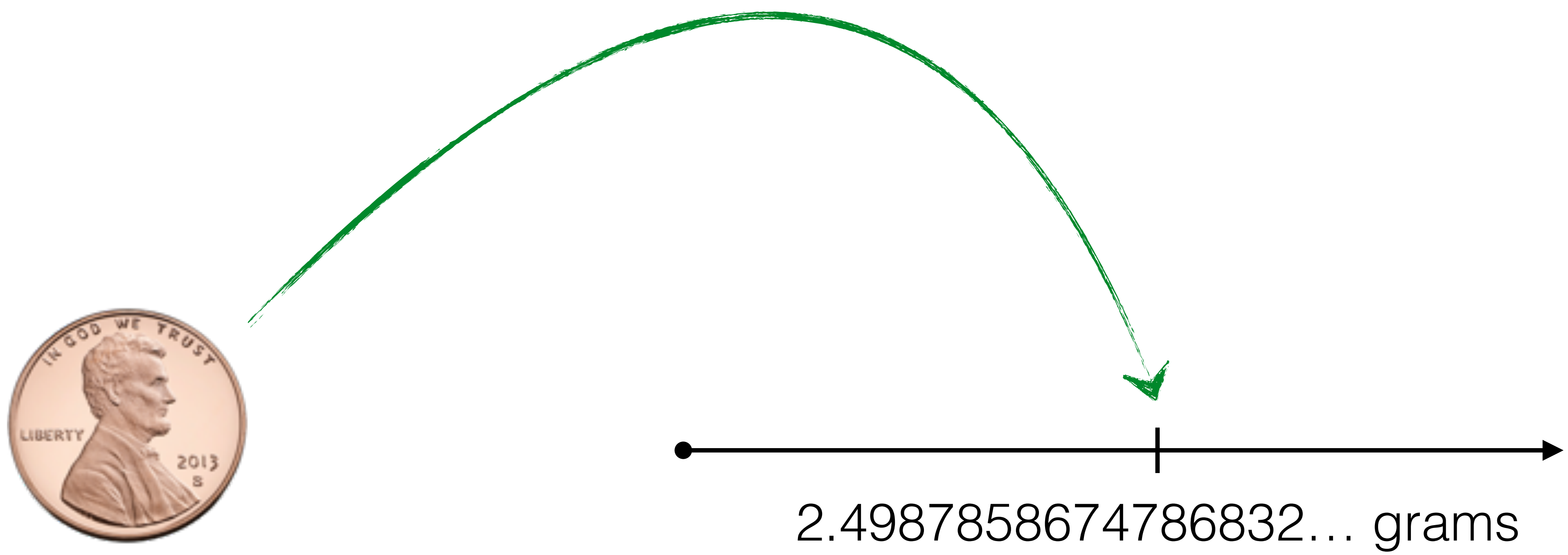
1: tails

Discrete.
Can enumerate all possible outcomes

outcome
(mass of a penny)

random variable

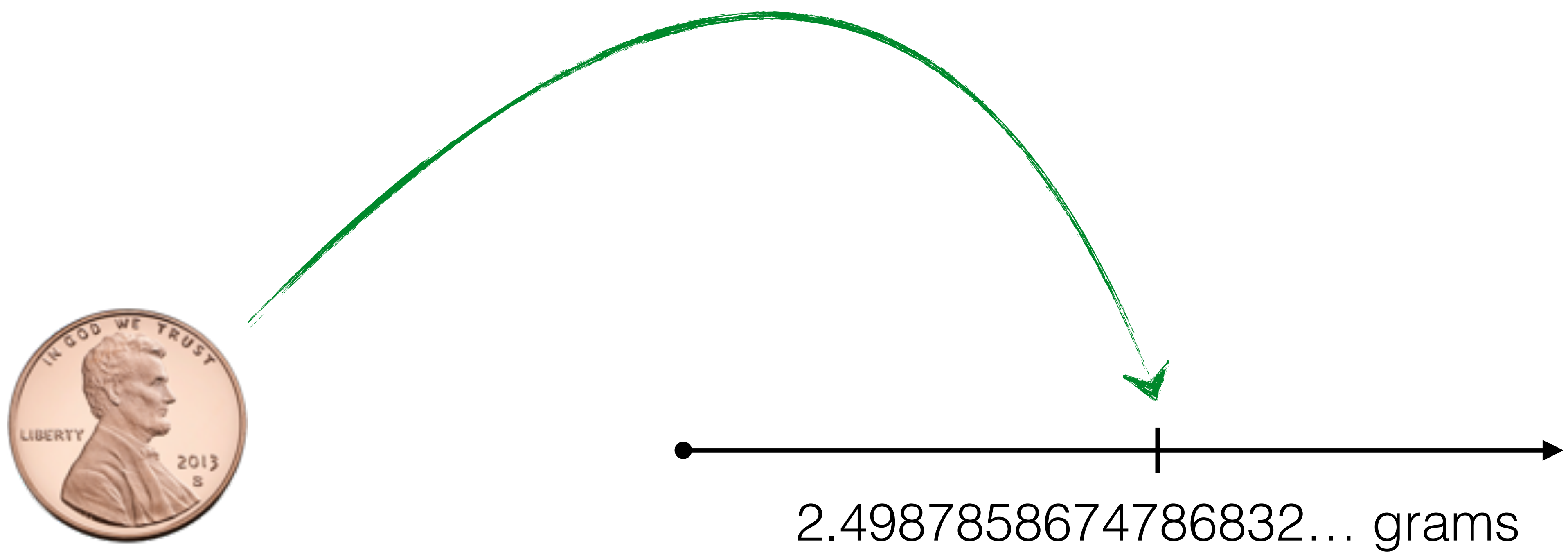
value
(a number)



outcome
(mass of a penny)

$\xrightarrow{\text{random variable}}$

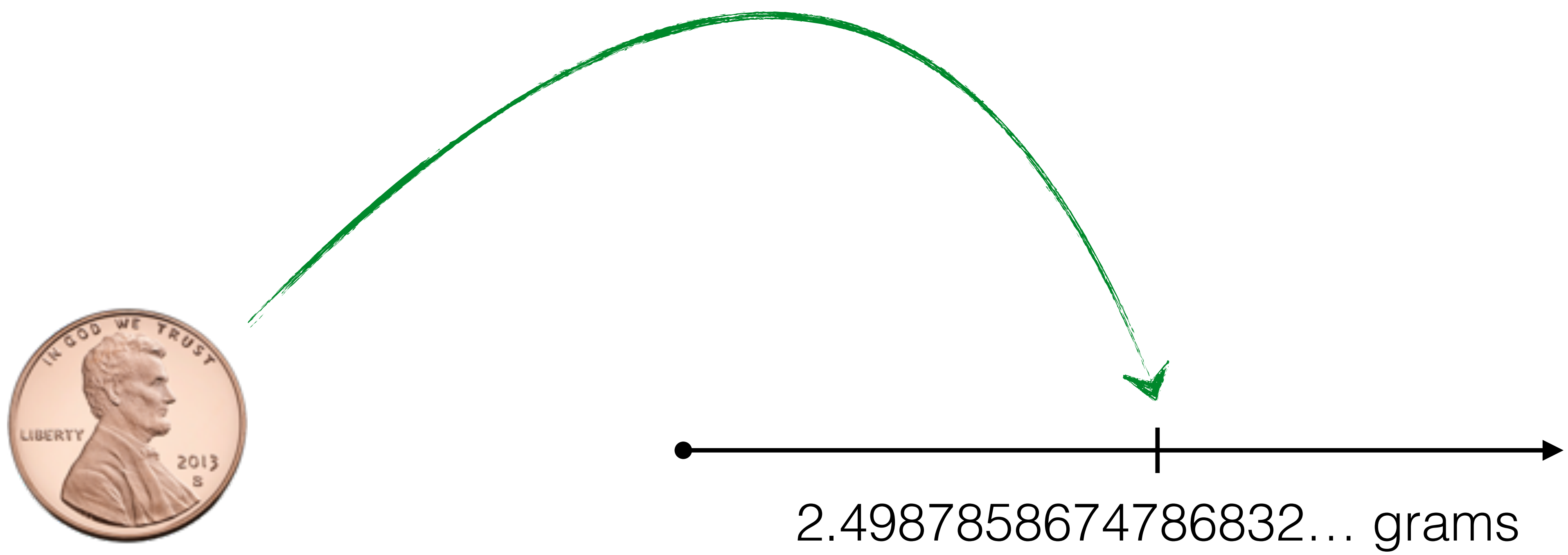
value
(a number)



What kind of random variable is this?

outcome $\xrightarrow{\text{random variable}}$ value

(mass of a penny) (a number)



Continuous.

Cannot enumerate all possible outcomes

Random Variables are typically denoted with a capital letter

$$X, Y, A, \ldots$$

Probability:

the chance that a particular event (or set of events) will occur expressed on a linear scale from 0 (impossibility) to 1 (certainty)

http://mathworld.wolfram.com/Probability.html

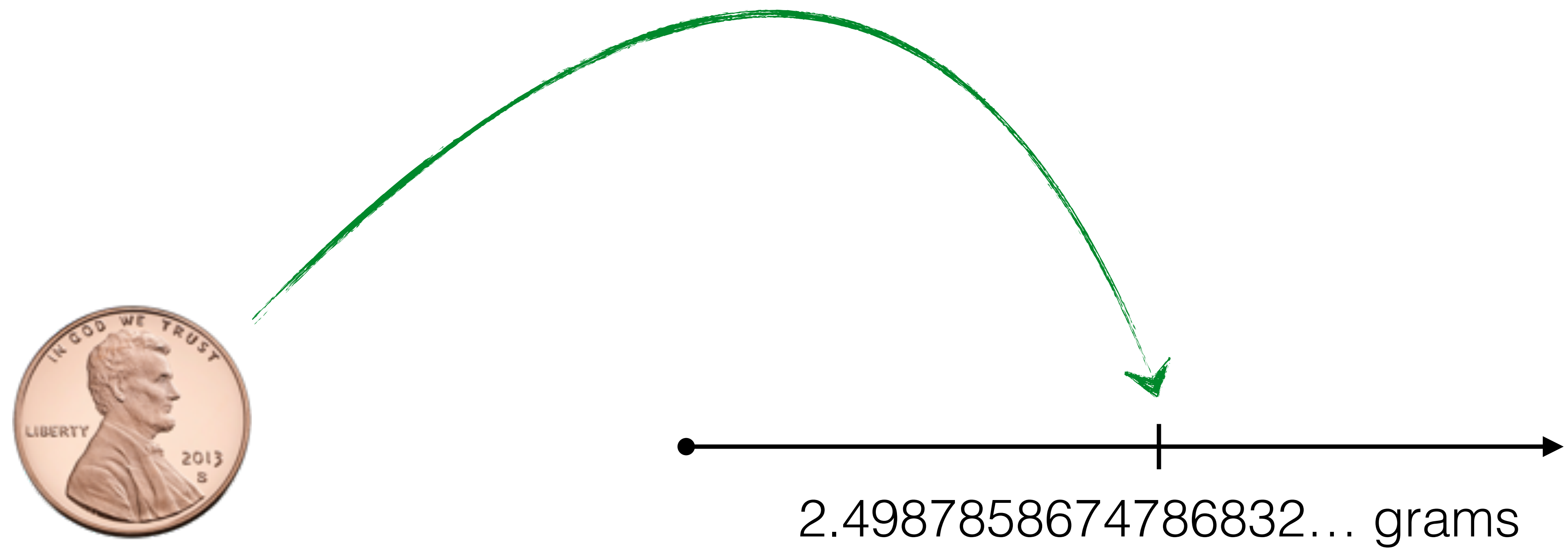
0: heads

1: tails

$$p(X = 0) = 0.5$$

$$p(X = 1) = 0.5$$

$$p(X = 0) + p(X = 1) = 1.0$$



$$\int p(x)dx = 1$$

Probability Axioms:

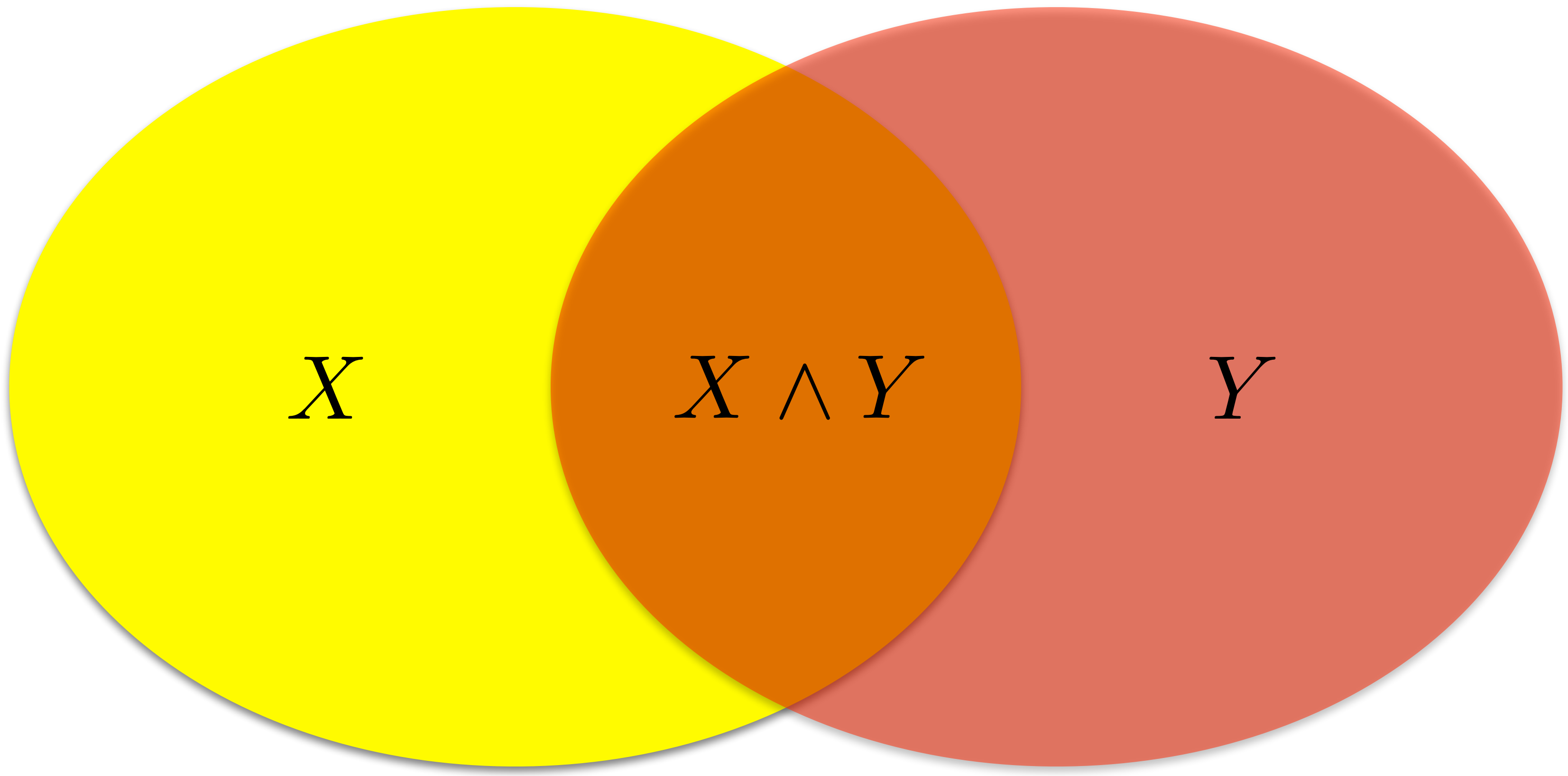
$$0 \leq p(x) \leq 1$$

$$p(\text{true}) = 1$$

$$p(\text{false}) = 0$$

$$p(X \vee Y) = p(X) + p(Y) - P(X \wedge Y)$$

$$p(X \vee Y) = p(X) + p(Y) - P(X \wedge Y)$$



Joint Probability

$$p(x, y)$$

When random variables are **independent**
(a sequence of coin tosses)

$$p(x, y) = p(x)p(y)$$

When random variables are **dependent**

$$p(x, y) = p(x|y)p(y)$$

this is a conditional probability defined next …

Conditional Probability

$$p(x|y)$$

Conditional probability of x given y

$p(x|y)$ is the short hand for $\qquad ?$

in terms of the random variables **X** and **Y**

Conditional Probability

$$p(x|y)$$

Conditional probability of x given y

$p(x|y)$ is the short hand for $p(X = x|Y = y)$

How is it related to the joint probability?

$$p(x|y) = \frac{p(x, y)}{?}$$

Conditional Probability

$$p(x|y)$$

Conditional probability of x given y

$p(x|y)$ is the short hand for $p(X = x|Y = y)$

$$p(x|y) = \frac{p(x, y)}{p(y)}$$

Conditional probability is the probability of the union of the events x and y divided by the probability of event y

Bayes' Rule

likelihood

$$p(x|y) = \frac{p(y|x)\ ?}{?}$$

posterior

What's the relationship between the posterior and the likelihood?

Bayes' Rule

$$\underbrace{p(x|y)}_{\text{posterior}} = \frac{\overbrace{p(y|x)}^{\text{likelihood}}\overbrace{p(x)}^{\text{prior}}}{\underbrace{p(y)}_{\text{evidence (observation prior)}}}$$

How do you compute the evidence (observation prior)?

The use of evidence under Bayes' theorem relates to the likelihood of finding evidence in relation to the accused, where Bayes' theorem concerns the probability of an event and its inverse. Specifically, it compares the probability of finding particular evidence if the accused were guilty, versus if they were not guilty. An example would be the probability of finding a person's hair at the scene, if guilty, versus if just passing through the scene. Another issue would be finding a person's DNA where they lived, regardless of committing a crime there.

In Bayesian statistical inference, a prior probability distribution, often simply called the prior, of an uncertain quantity is the probability distribution that would express one's beliefs about this quantity before some evidence is taken into account. For example, the prior could be the probability distribution representing the relative proportions of voters who will vote for a particular politician in a future election. The unknown quantity may be a parameter of the model or a latent variable rather than an observable variable.

Bayes' theorem calculates the renormalized pointwise product of the prior and the likelihood function, to produce the posterior probability distribution, which is the conditional distribution of the uncertain quantity given the data.

Similarly, the prior probability of a random event or an uncertain proposition is the unconditional probability that is assigned before any relevant evidence is taken into account.

Bayes' Rule

likelihood prior

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

posterior

evidence (observation prior)

How do you compute the evidence (observation prior)?

$$p(x|y) = \frac{p(y|x)p(x)}{\sum_{x'} p(y|x')p(x')}$$

evidence (expanded)

Bayes' Rule

likelihood prior

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

posterior

evidence

Evidence (observation prior) is also called the **normalization factor**

$$p(x|y) = \eta p(y|x)p(x)$$

$$p(x|y) = \frac{1}{Z}p(y|x)p(x)$$

Bayes' Rule with 'evidence'

$$p(x|y,e) = \frac{p(y|x,e)p(x|e)}{p(y|e)}$$

Marginalization

$$p(x) = \sum_{y} p(x, y)$$

Marginalize out y

Conditioning

$$p(x) = \sum_{y} p(x|y)p(y)$$

Conditioned on y

Joint probability over three (dependent) variables

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

$$p(cavity) = ?$$

Recall: $p(x) = \sum_y p(x, y)$

Joint probability over three (dependent) variables

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

$$p(cavity) = ?$$

$$p(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

Joint probability over three (dependent) variables

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

$$p(cavity|toothache) = ?$$

Joint probability over three (dependent) variables

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

$$
\begin{aligned}
p(cavity|toothache) &= \frac{p(cavity, toothache)}{p(toothache)} \\
&= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} \\
&= 0.6
\end{aligned}
$$