Bayesian Analysis for the Life Sciences

Bob O'Hara
Dept. of Mathematics and Statistics,
University of Helsinki
www.rni.helsinki.fi/~boh

Laws of Probability

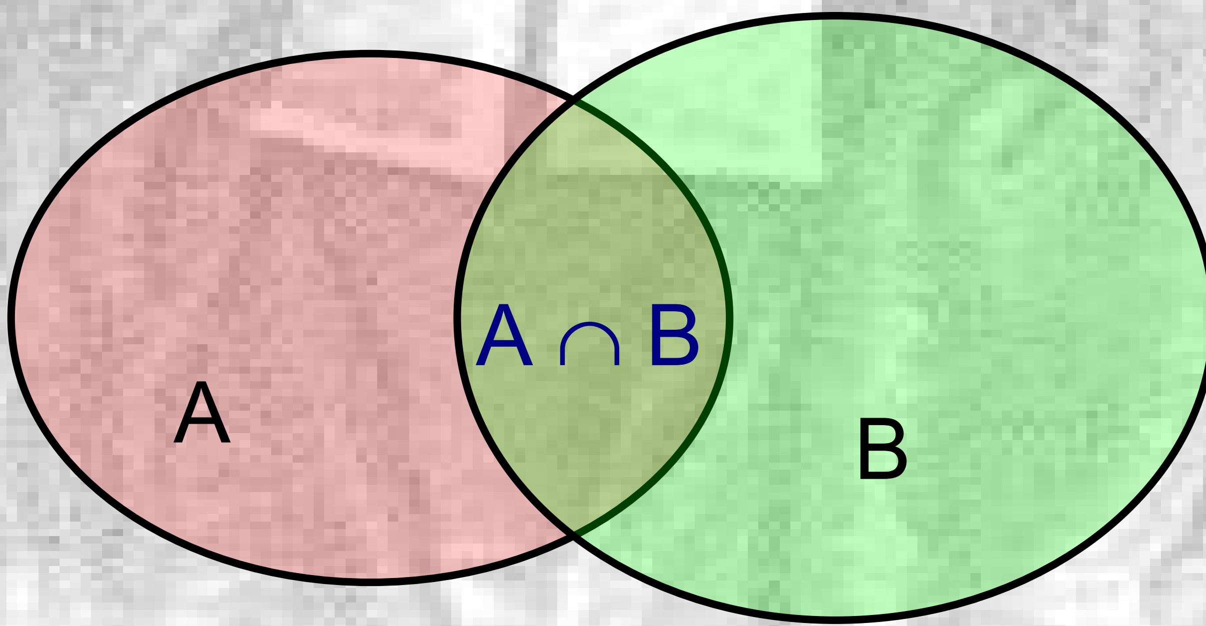
- These are mathematical theorems
- Problem (later): relating these to the real world
- Definitions:
  - A and B are events
  - A – Bus to work is late
  - B – Bus from work is late
  - Sometimes: C_i – i people waiting for the bus
- Notation: A^c – not A (i.e. bus not late)

Basics

- $0 \leq P(A) \leq 1$
- $P(A) + P(A^c) = 1$
  - either a bus is late, or is on time
- $\Sigma P(C) = 1$
  - all probabilities add up to 1
- If $P(A \text{ and } B) = 0$, then $P(A \text{ or } B) = P(A) + P(B)$

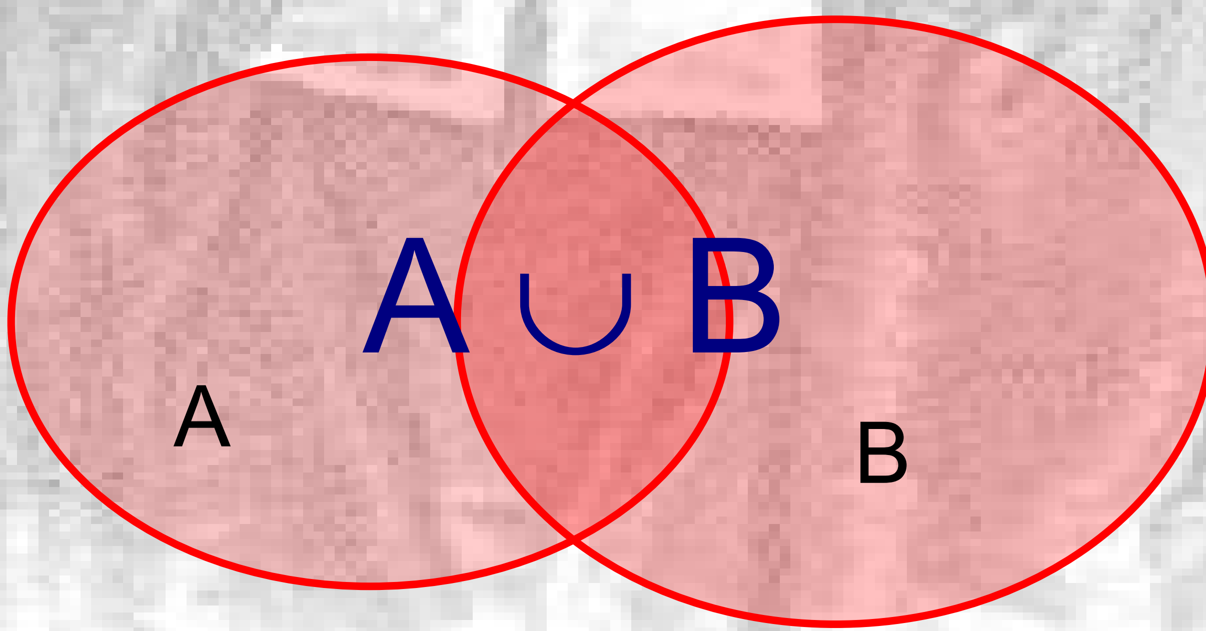
Intersection: $P(A \cap B)$

- e.g. $P(A) = 0.4$, $P(B) = 0.2$, $P(A \text{ and } B) = 0.1$
- Question: What is the probability of both buses being late?



Union: P(A U B)

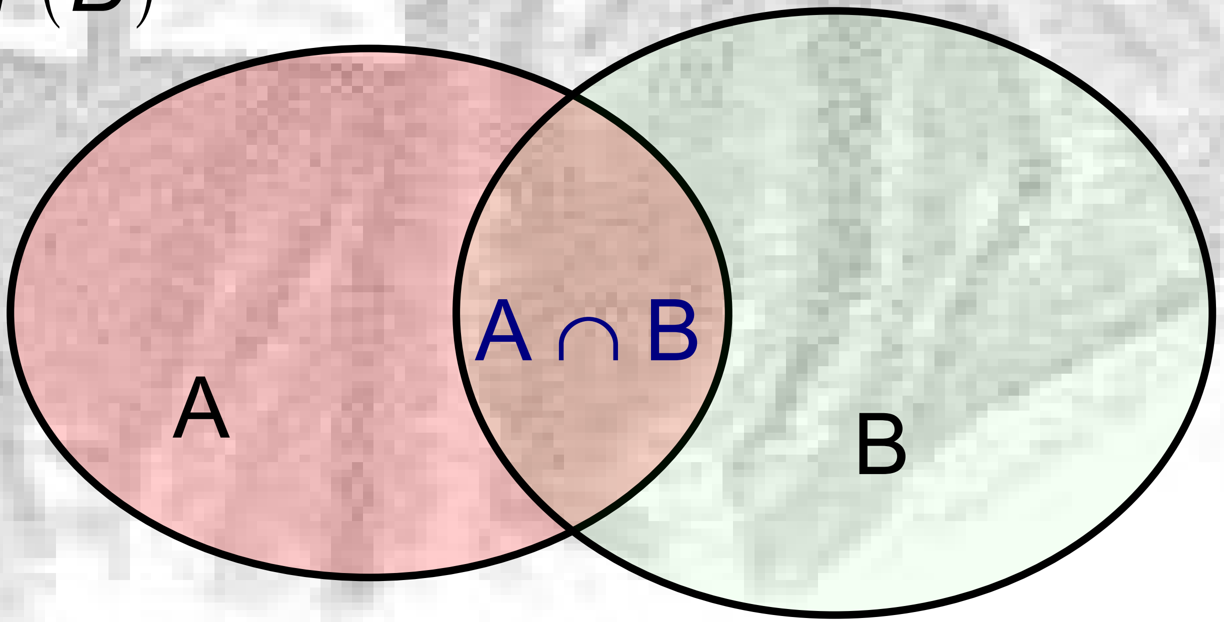
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- What is the probability of at least 1 bus being late?
  - P(A and B) = 0.4 + 0.2 – 0.1 = 0.5



Conditional Probability

- If my bus to work is late, what is the probability my bus home is late too?

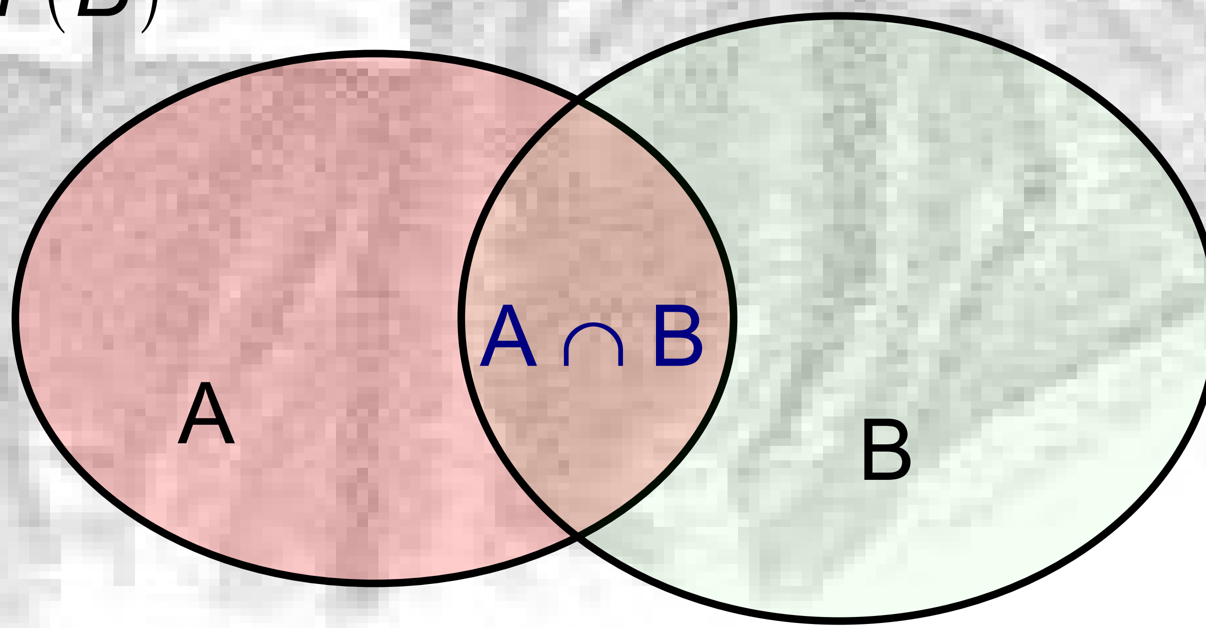
$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$



Conditional Probability

- P(Bus late in evening | late in morning)
- = 0.1/0.4
- = 0.25

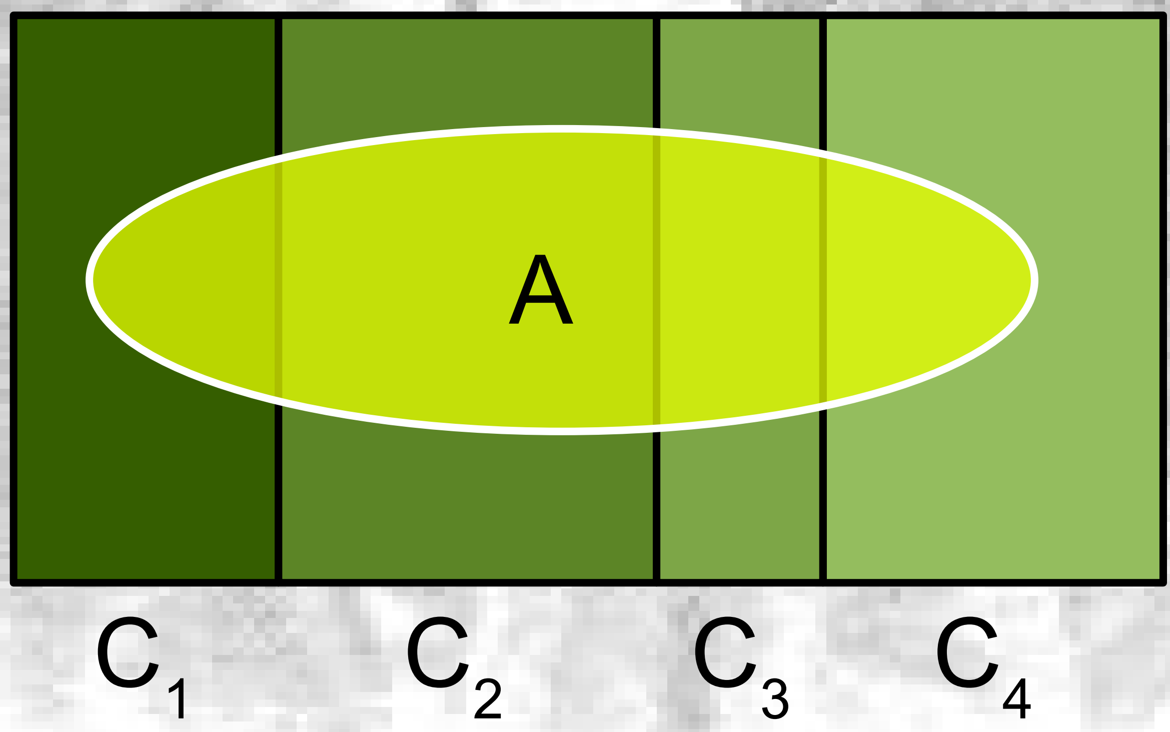
$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$



Law of Total Probability

- Everything adds up!

$$Pr(A) = \sum_{i=1}^{\infty} Pr(A|C_i)\,P(C_i)$$



Independence

- Definition: A and B are independent if

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

- The point:

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

- i.e. if the probability of the my bus home being late is independent of probability of the my bus to work being late, then knowing one says nothing about the other

Late buses

- Bus late in morning: P(A) = 0.4
- Bus late in evening: P(B) = 0.2
- Both buses late: P(AnB) = 0.1
- P(A).P(B) = 0.4 x 0.2 = 0.08
  - So A and B not independent
- P(B|A) = 0.25
  - if my bus is late in the morning, it is more likely to be late in the evening

Bayes' Rule: “The Law of Inverse Probablity”

- If I know P(A|B), can I find P(B|A)?

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

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

so

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

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

Bayes' Rule

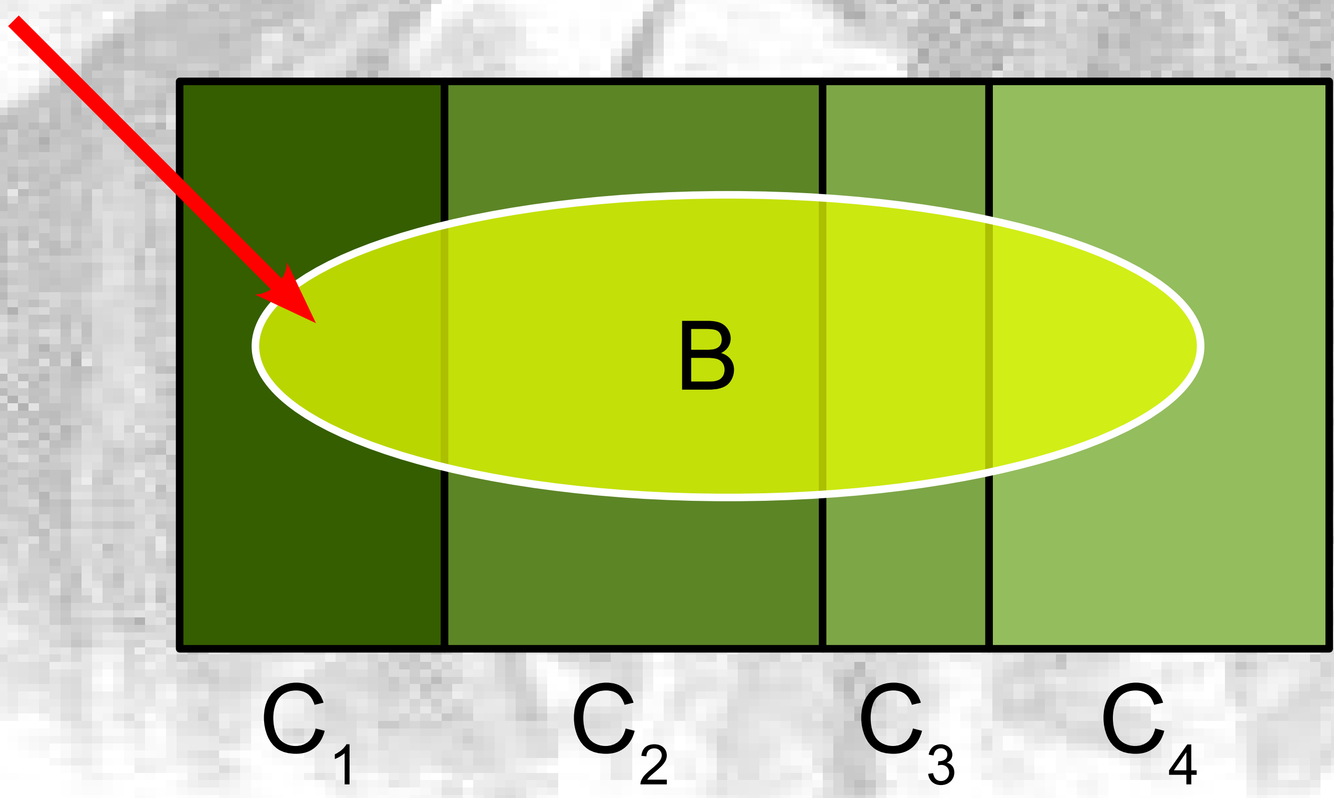
$$Pr(A|B) = \frac{Pr(B|A)\,Pr(A)}{Pr(B)}$$

- as $Pr(B) = \sum_{i=1}^{\infty} Pr(B|C_i)\,P(C_i)$

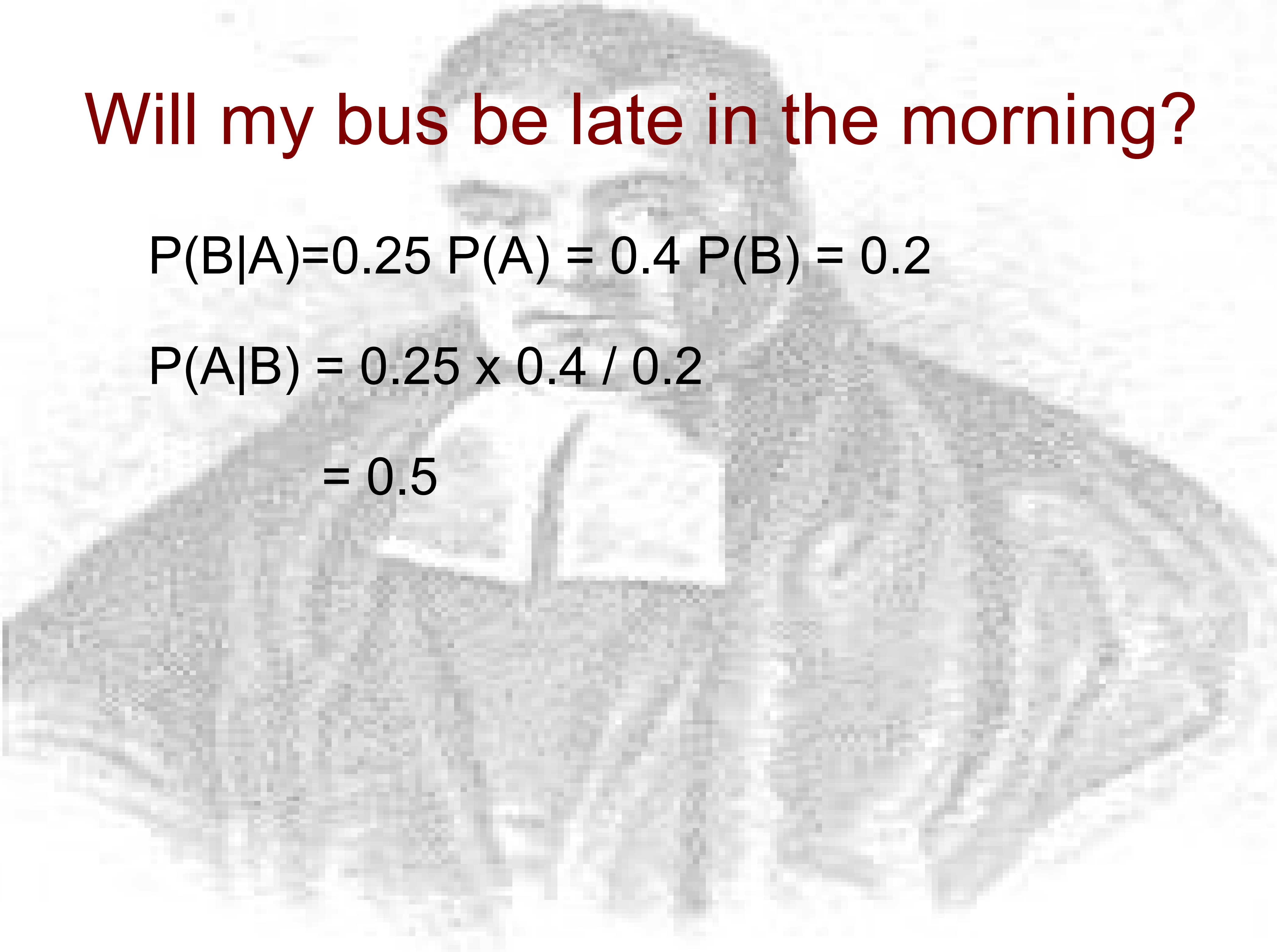
$$Pr(C_1|B) = \frac{Pr(B|C_1)\,Pr(C_1)}{\sum_i Pr(B|C_i)\,Pr(C_i)}$$

$$Pr\left(C_1 \middle| B\right) = \frac{Pr\left(B \middle| C_1\right) Pr\left(C_1\right)}{\sum_i Pr\left(B \middle| C_i\right) Pr\left(C_i\right)}$$

$$Pr\left(C_1 \middle| B\right)$$



Will my bus be late in the morning?



$P(B|A)=0.25$ $P(A) = 0.4$ $P(B) = 0.2$

$P(A|B) = 0.25 \times 0.4 / 0.2$

$= 0.5$

Probability Distributions

- Not everything random is just measured as TRUE/FALSE
  - e.g. number of people waiting for a bus
- describe these as a *probability distribution*
  - e.g. probability that *n* people are waiting for a bus
- Describes mathematically these probabilities

How many buses to work will be late?

- 5 working days per week, on how many of them will my bus be late?
- Probability of one bus being late is p
- Assume constant and independent
- Define N is the *random variable* “number of late buses”
- Then $\Pr(N=5) = p^5$

The Binomial Distribution

- $Pr(N=4)$
  - one bus on time. Could be on Monday, Tuesday, Wednesday etc.
  - 5 ways of this happening, with probabilities
    - On time on Monday: $p(1-p)^4$
    - On time on Tuesday: $(1-p)p(1-p)^3$
    - On time on Wednesday: $(1-p)^2p(1-p)^2$
    - On time on Thursday: $(1-p)^3p(1-p)$
    - On time on Friday: $(1-p)^4p$
  - Total probability: $5p(1-p)^4$

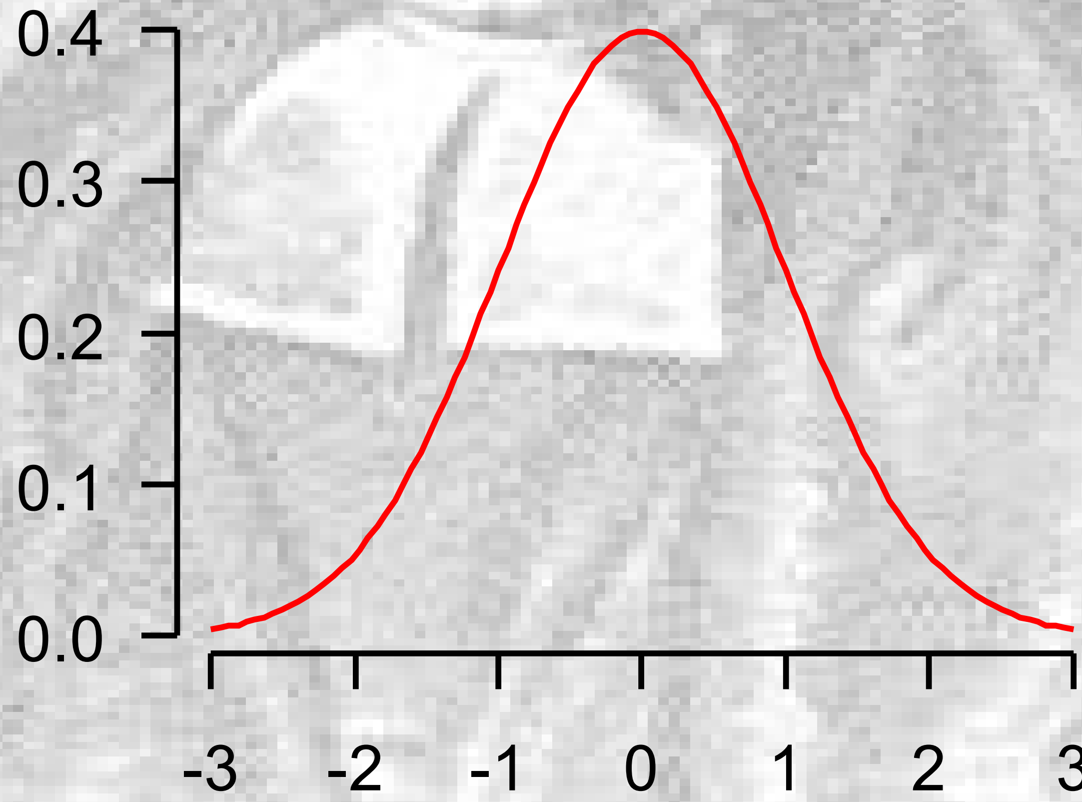
The Binomial Distribution

- Pr(N=3): several ways of happening
  - each one with probability $p^3(1-p)^2$
- To work out how many ways each *N* can happen, use combinatorics
- End up with

$$Pr(N=n)=\frac{n!}{N!(N-n)!}p^n(1-p)^{N-n}$$

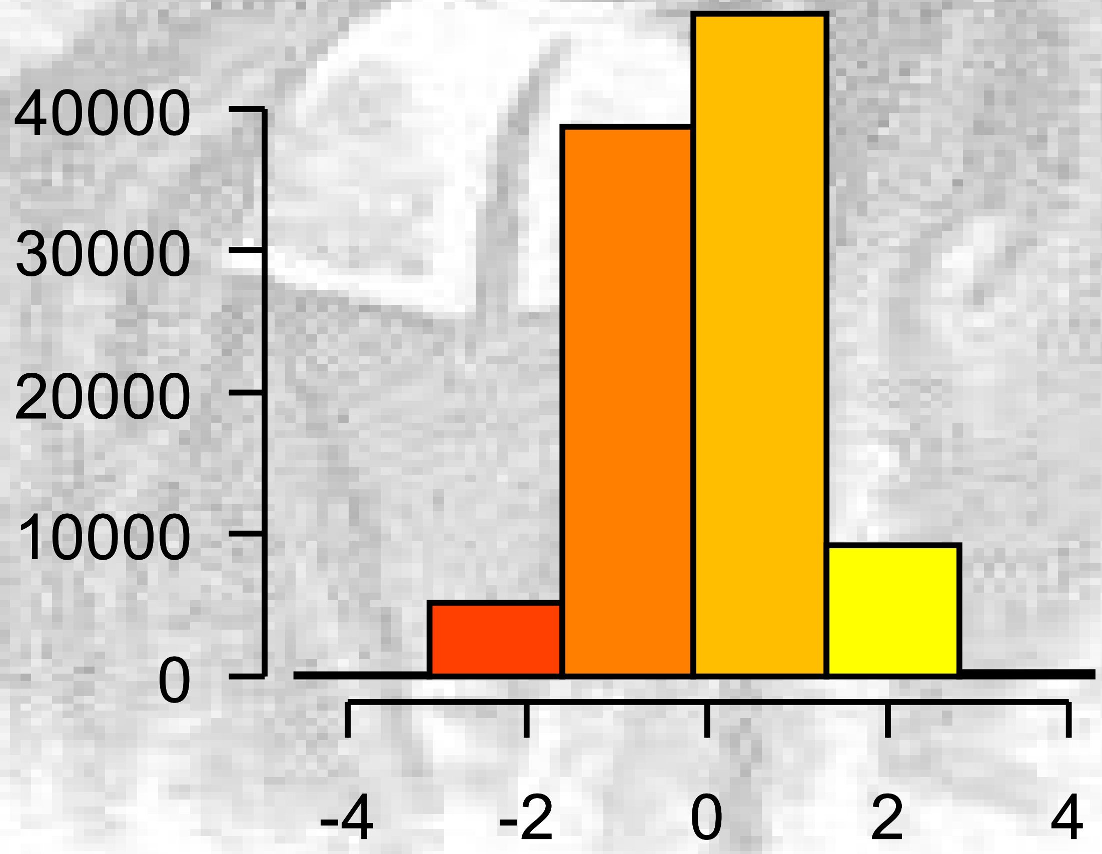
Discrete to Continuous

- So far, all probabilities have been discrete
- For continuous distributions – use densities



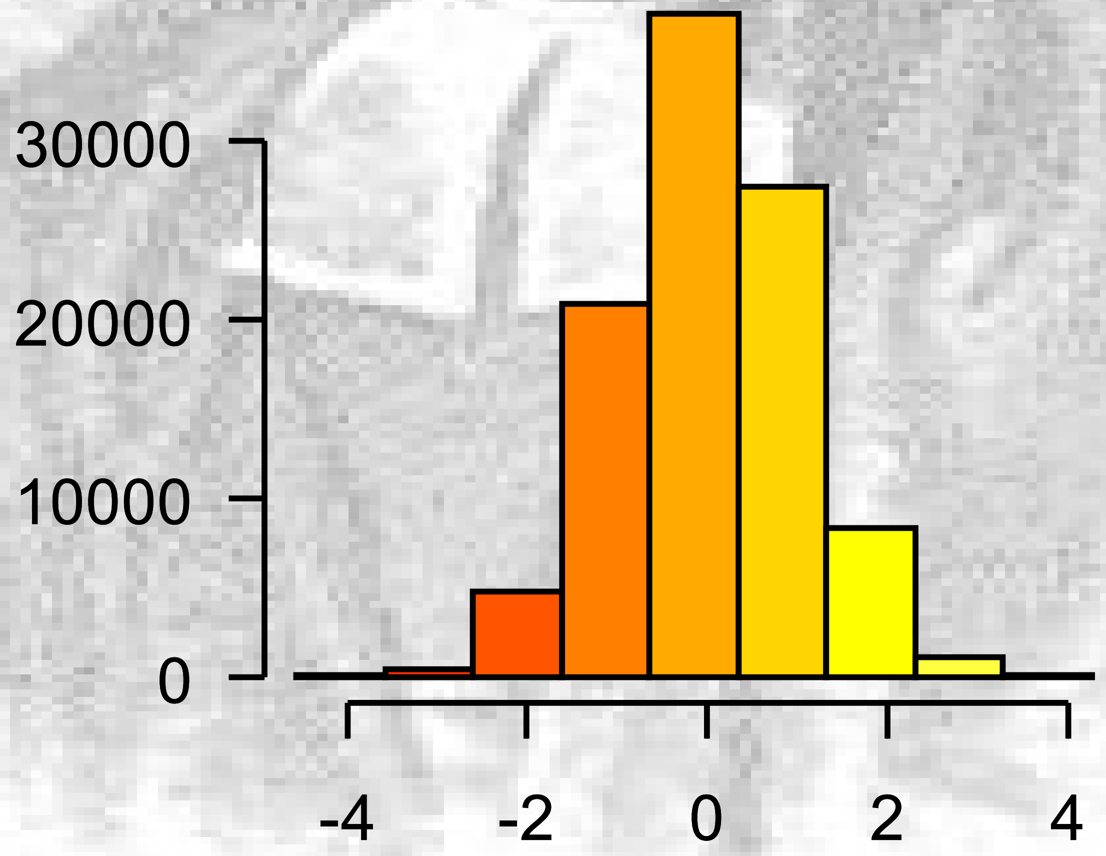
For example

- Split into classes
- count number in each class



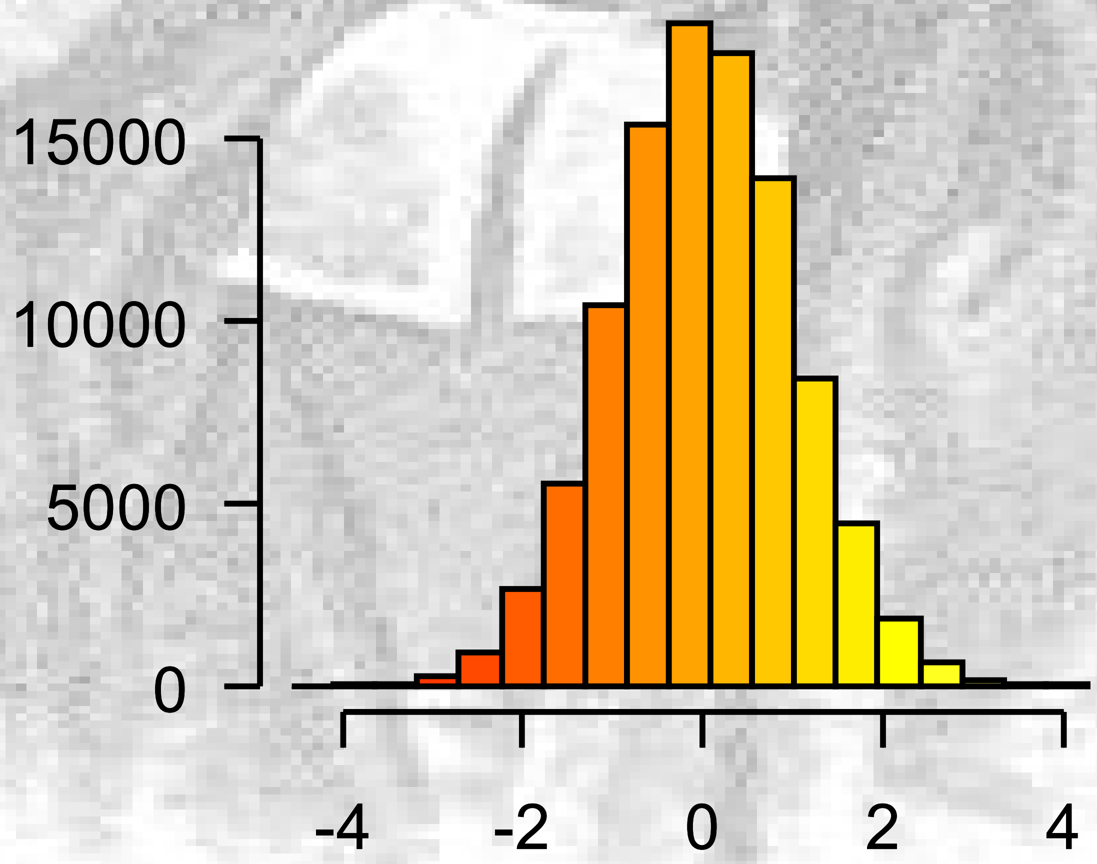
For example

- Now make class widths smaller



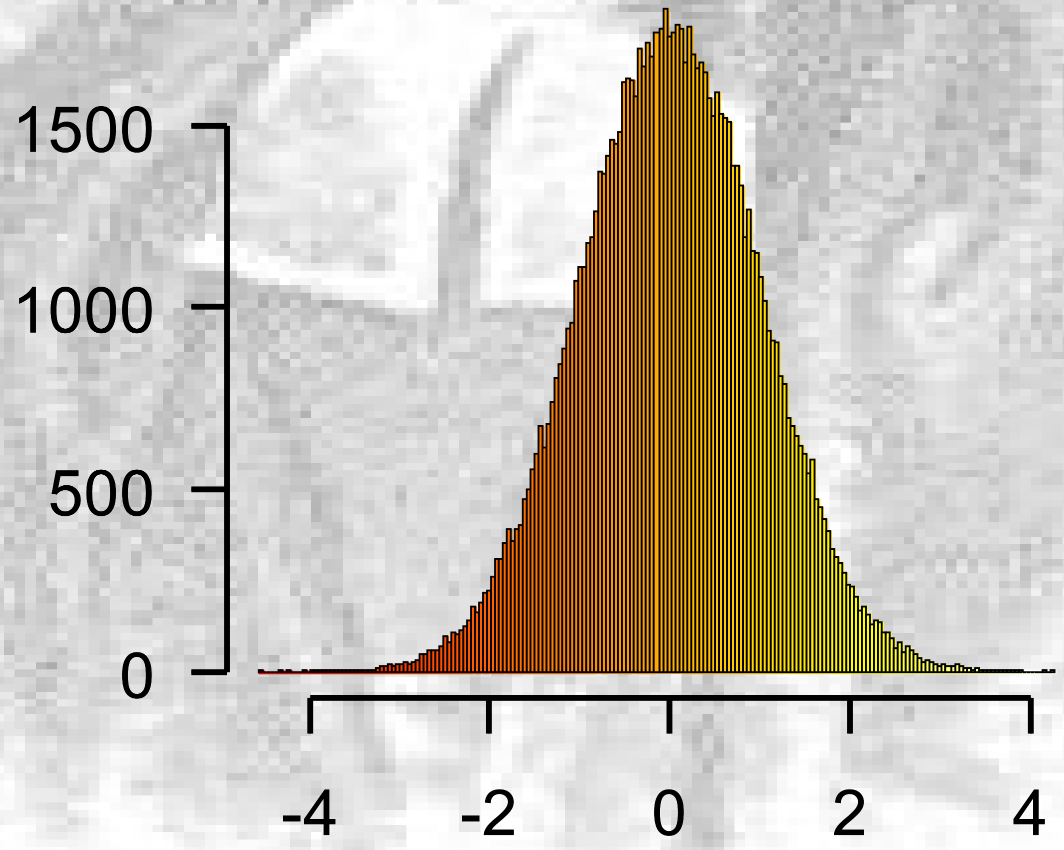
For example

- ... and smaller



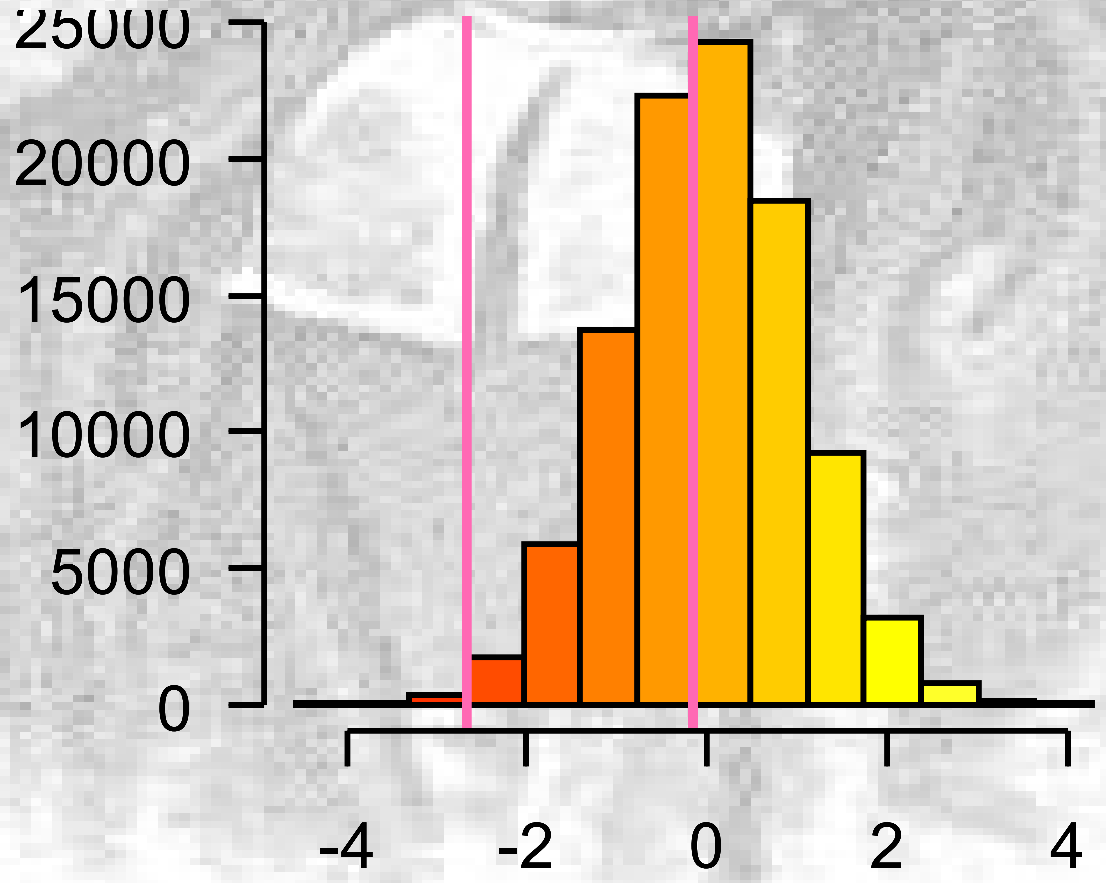
For example

- Eventually look continuous



- To work out the probability that an observation is between 2 values

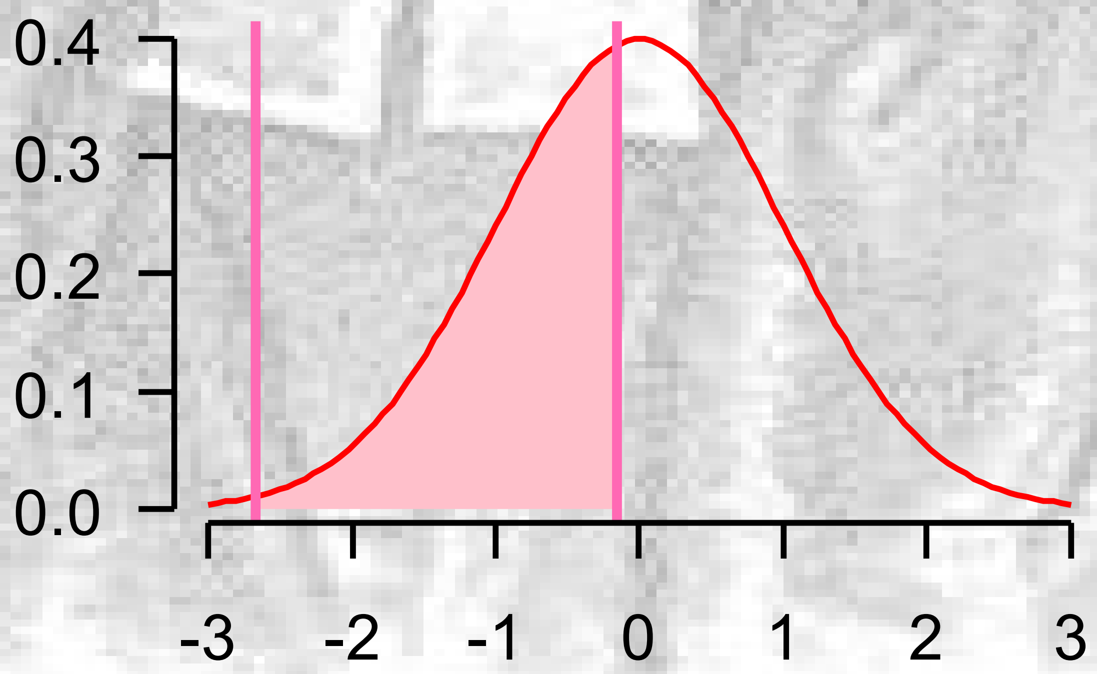
$$Pr(a < x < b) = \sum_{a}^{b} Pr(x)$$



- Now make x $\rightarrow$ 0
- Probability becomes an integral

$$Pr\left(a<x<b\right)=\int_{a}^{b} f\left(x\right)dx$$

- f(x) is called a *probability density function*



Some Distributions

- Several distributions are commonly used
- Here are some...

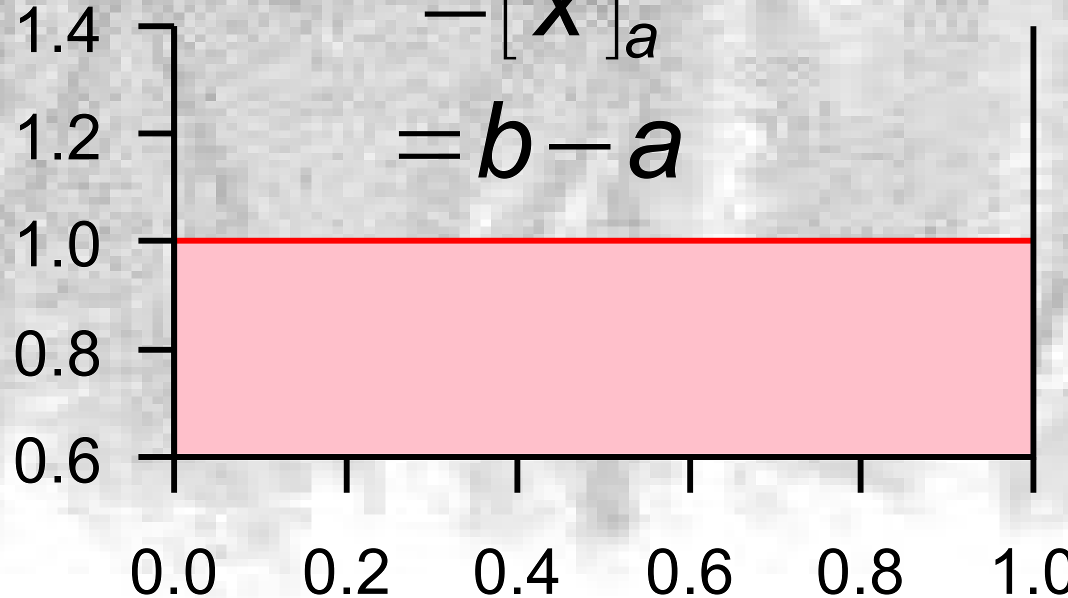
Uniform Distribution

- Probability density constant   f(x) = 1
  - usually defined between 0 and 1

$$Pr(a < x < b) = \int_a^b 1\, dx$$

$$= [x]_a^b$$

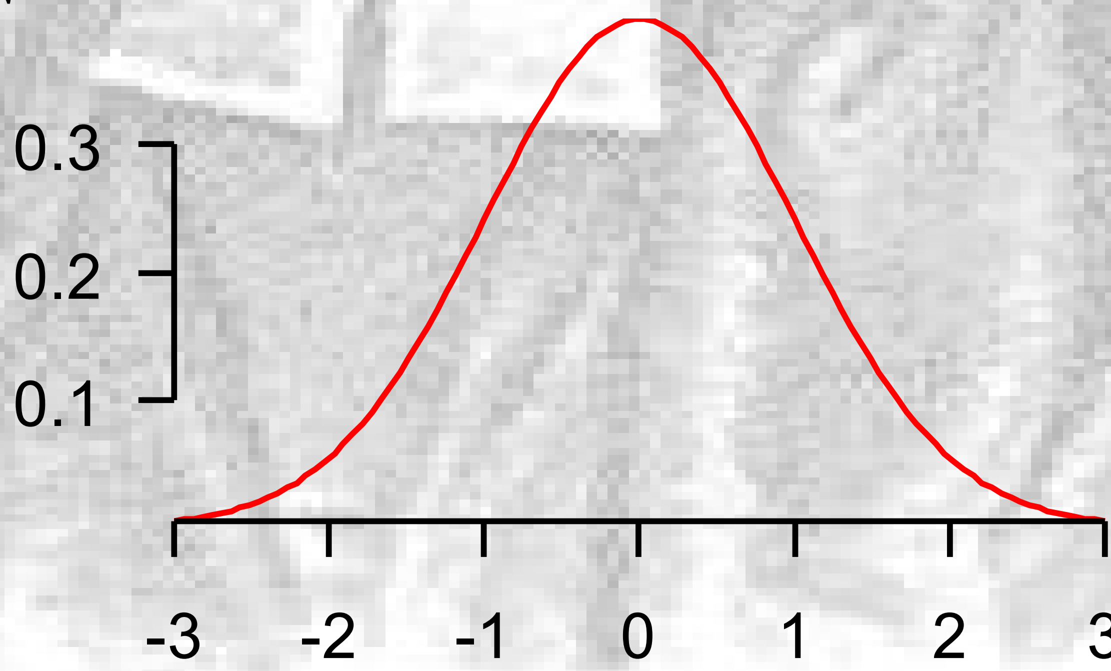
$$= b - a$$



Normal Distribution

- Commonly used
- Probability density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(x-\mu)^2}{2\sigma^2}$$



Expectations

- "Average" values
  - theoretical values
- Definition:

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx$$

- In general:

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

$$E(g(x)) = \sum_{i=-\infty}^{\infty} g(i) Pr(x=i)$$

Expected Value for a Uniform Distribution

- Simple Example

$$
\begin{aligned}
E(X) &= \int_0^1 x.1\,dx \\
&= \left[\frac{x^2}{2}\right]_0^1 \\
&= \frac{1}{2}
\end{aligned}
$$

Variance

- Var(X) = E(X^2) – E^2(X)
- e.g. Uniform:

$$
\begin{aligned}
Var(X) &= \int_0^1 x^{2.1}\,dx - \left(\frac{1}{2}\right)^2 \\
&= \left[\frac{x^3}{3}\right]_0^1 - \frac{1}{4} \\
&= \frac{1}{12}
\end{aligned}
$$

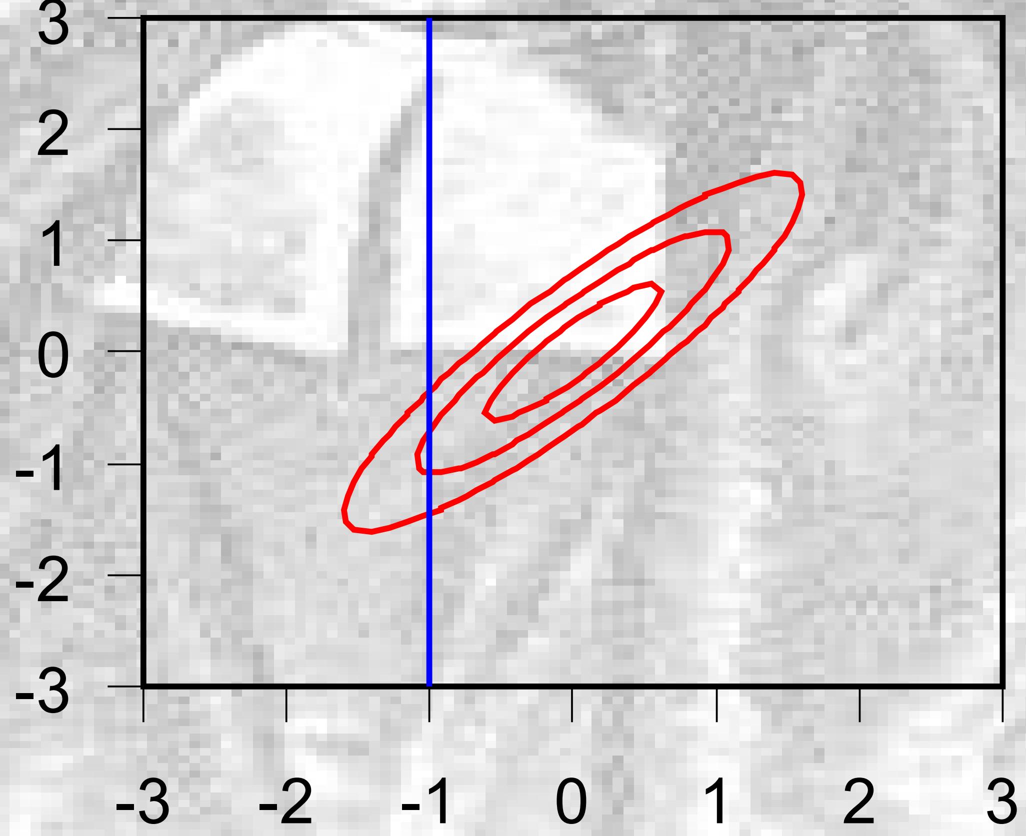
Conditional and Marginal Distributions

- Important distinction in practice
- Conditional distribution: P(x|y)
- Marginal distribution:

$$P(x) = \int_{-\infty}^{\infty} P(x|y)\,P(y)\,dy$$

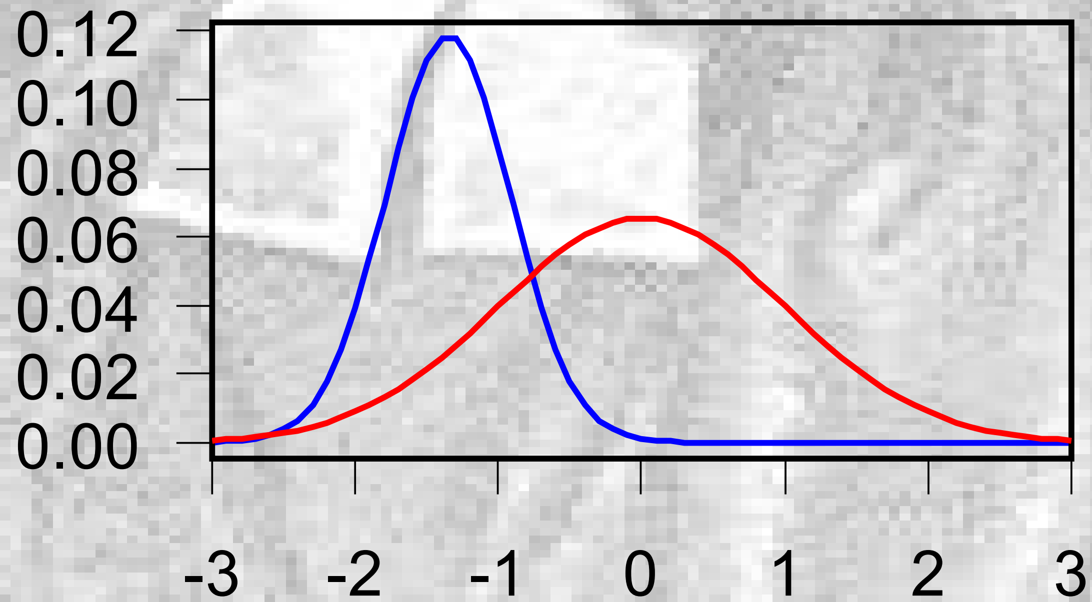
Example: Bivariate Normal

- x and y correlated:

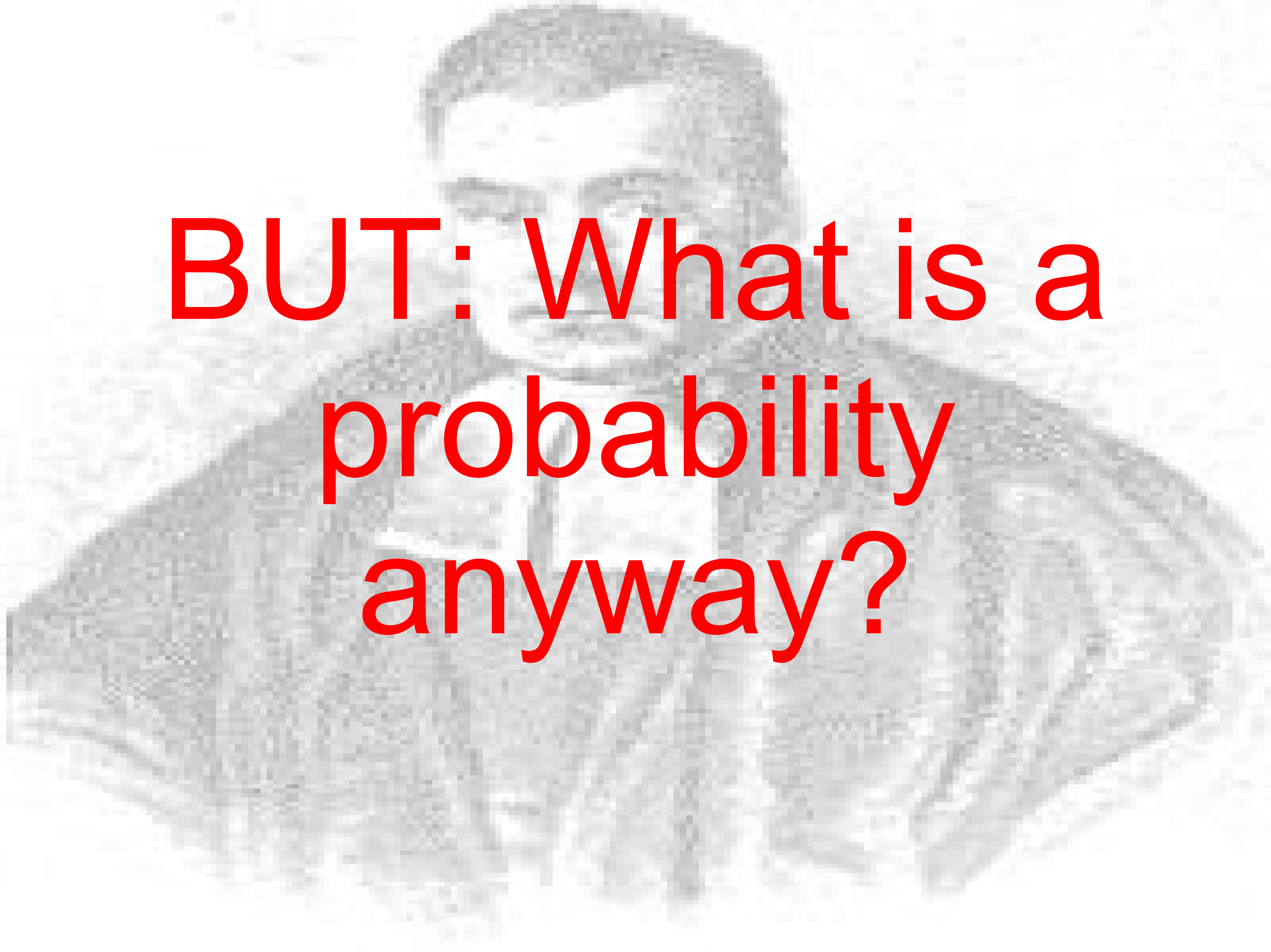


Example: Bivariate Normal

- P(y) and P(y|x=-1)
- Marginal and Conditional



- Marginal distribution has a larger variance
  - summed over all values of *x*



BUT: What is a probability anyway?