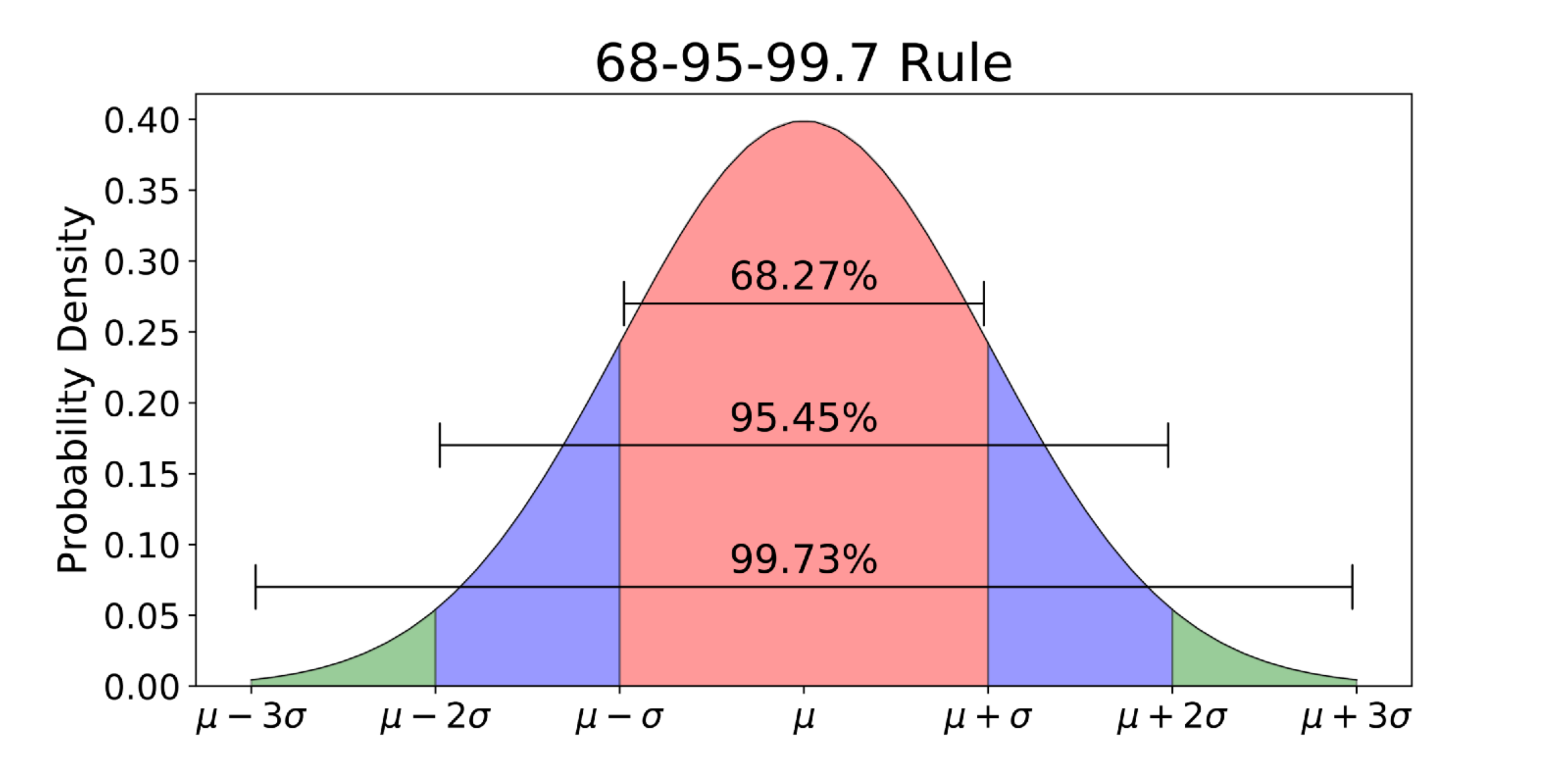
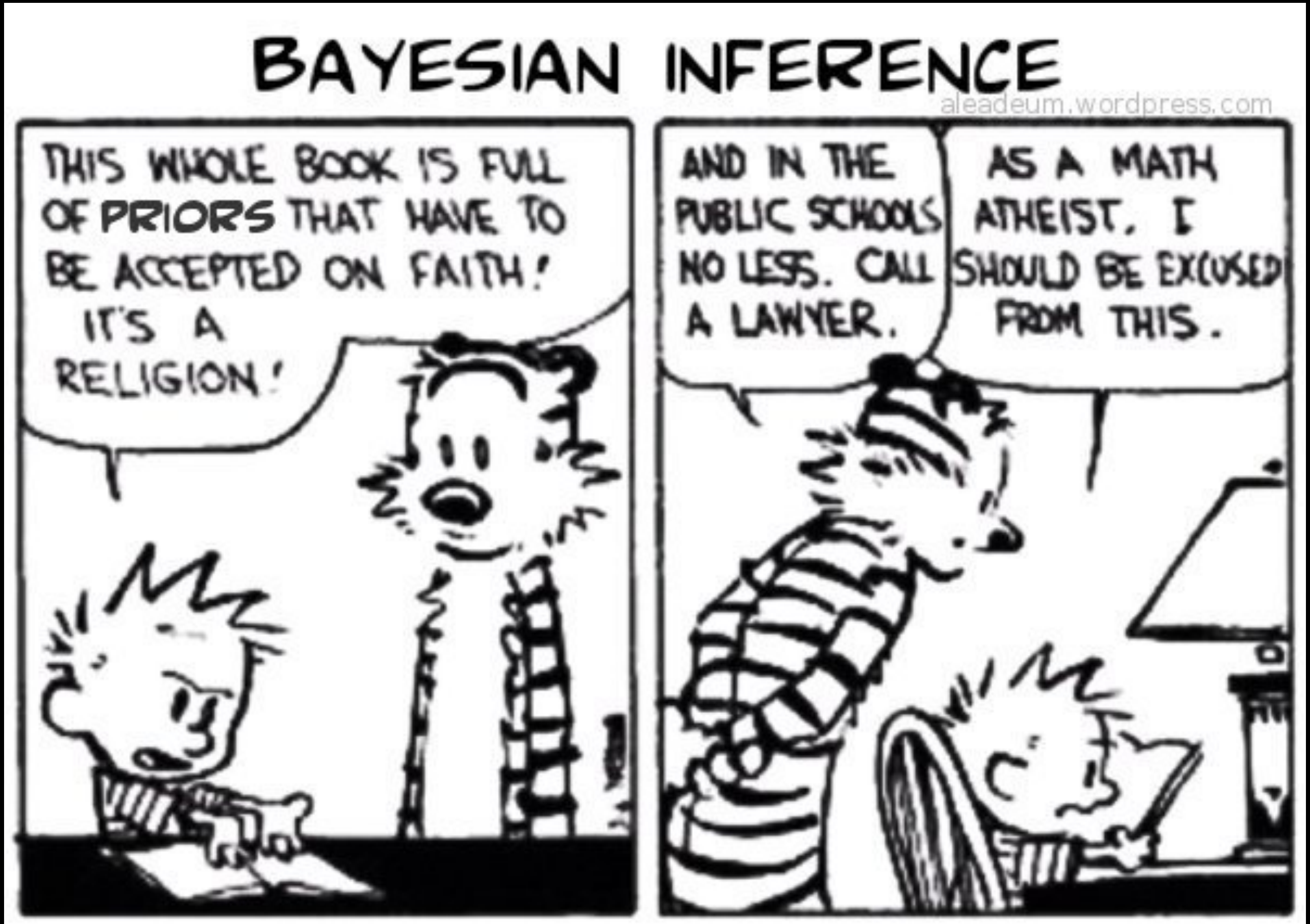
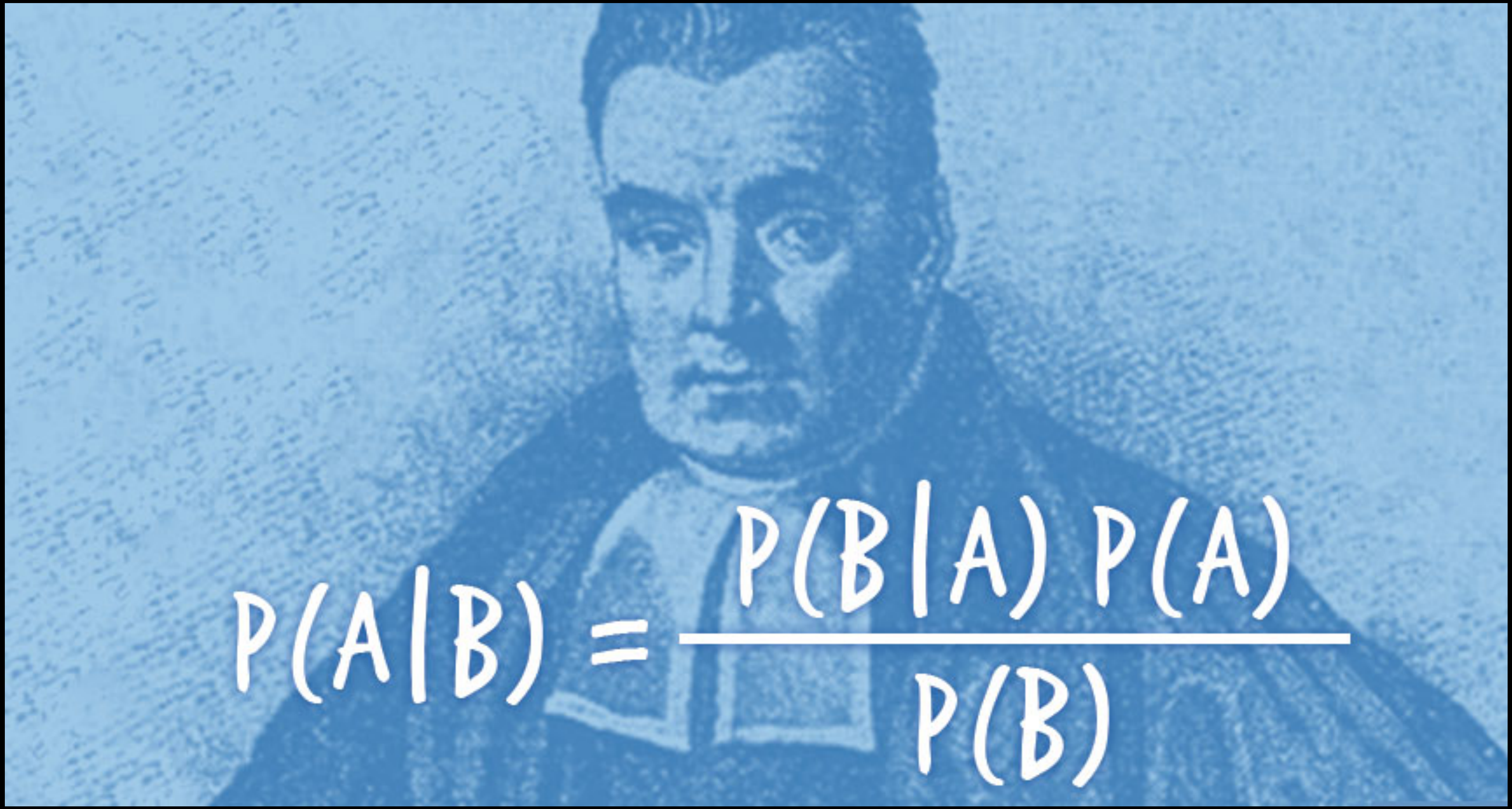
Standard deviations and probability of the population mean



Bayesian Inferencing



Thomas Bayes



Basic Probabilities

- The probability of something occurring is the number of ways that thing can occur divided by the total number of things that can occur.
- Say you flip a “fair” coin. What’s the probability of heads?
- Ways you can have heads = 1
- Total possible outcomes = 2
- Probability of heads = P(h) = ½ = 50%

Conditional Probabilities

- Conditional probabilities: If I know something ahead of time (or before), then what is the probability of event x? Conditional probabilities are calculated just like basic probabilities. What is P(drawing an Ace from a full deck) = 4/52. Ok. After drawing that ace what are probabilities of getting an Ace again? P(Ace|Ace) = 3/51. This is read what is the probability of getting an Ace given=| we've pulled one already. I.e., P(A|B) = The probability of A given B has already occurred.
- Conditionals = adjust all possibilities

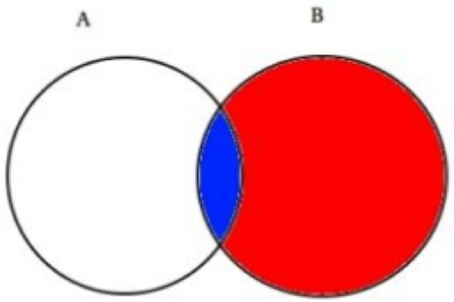
Joint Probabilities

- What is probability of pulling an Ace of (Hearts or Diamonds)?
- P(Ace) = P(a) 4/52
- P(Hearts or Diamonds) = P(hd) =13*2/52 = 26/52 = ½
- P(a and hd) = 4/52 * ½ = 4/104 = 2/52
- We calculate this in our heads but are actually doing some quick multiplication.
- Joint = multiply

Marginal Probabilities

- What is P of pulling any heart or diamond?
- P(heart) = 13/52
- P(diamond) = 13/52
- P(any red) = P(heart) + P(diamond)
- Marginal = add

Set A represents one set of events and Set B represents another. We wish to calculate the probability of A given B has already happened. Lets represent the happening of event B by shading it with red.



Now since B has happened, the part which now matters for A is the part shaded in blue which is interestingly $A \cap B$. So, the probability of A given B turns out to be:

$$\frac{BlueArea}{RedArea + BlueArea}$$

Therefore, we can write the formula for event B given A has already occurred by:

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

or

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

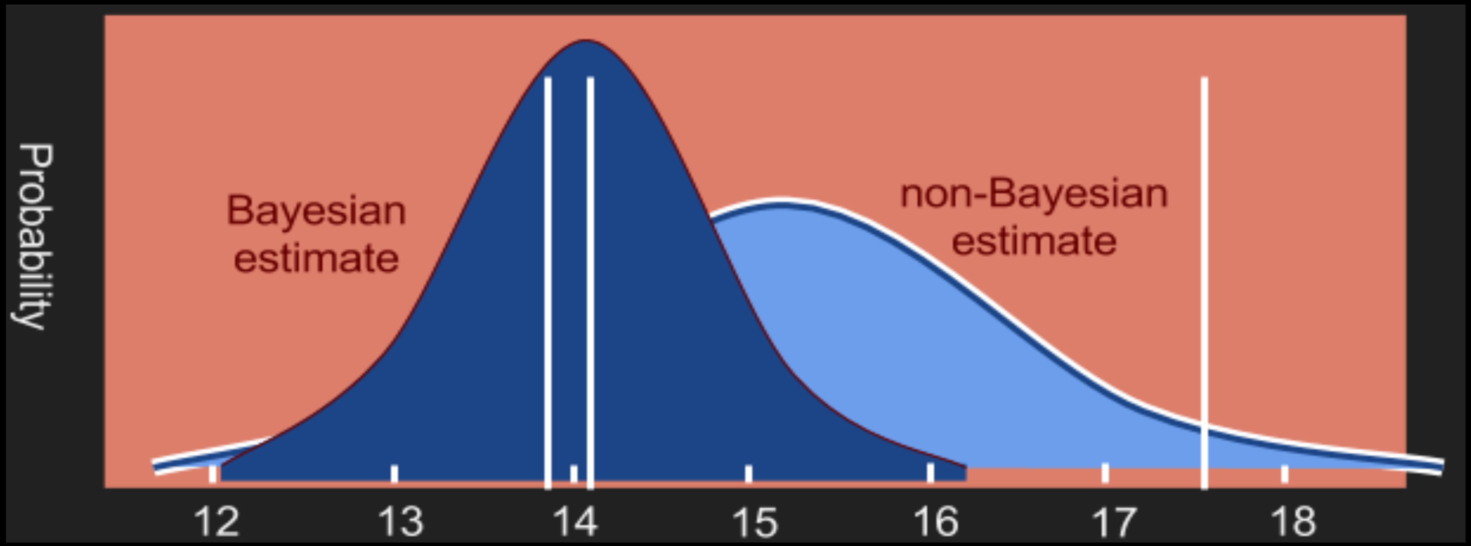
Now, the second equation can be rewritten as :

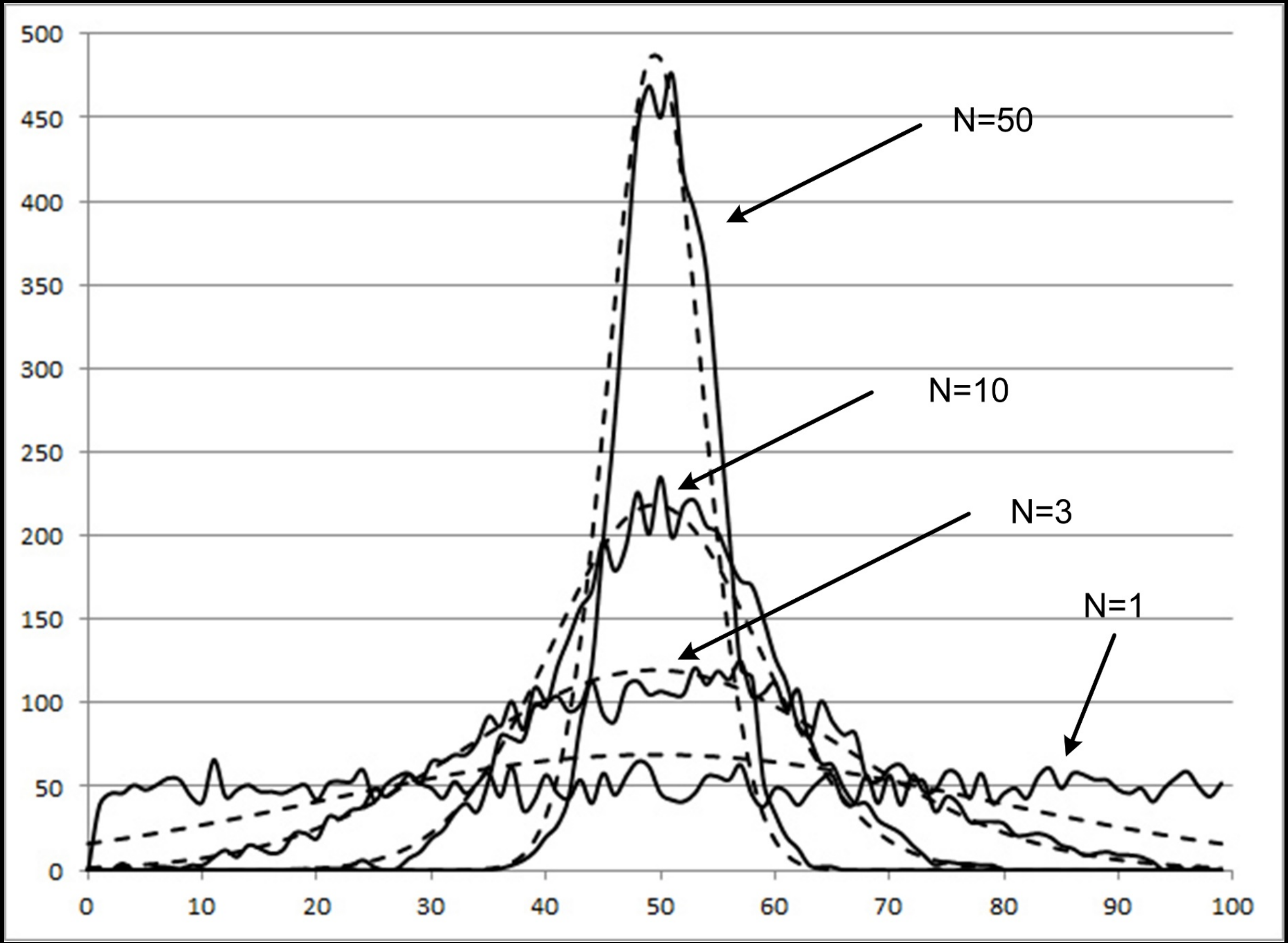
$$P(A|B) = \frac{P(B|A) X P(A)}{P(B)}$$

This is known as **Conditional Probability**.

Bayes Theorem

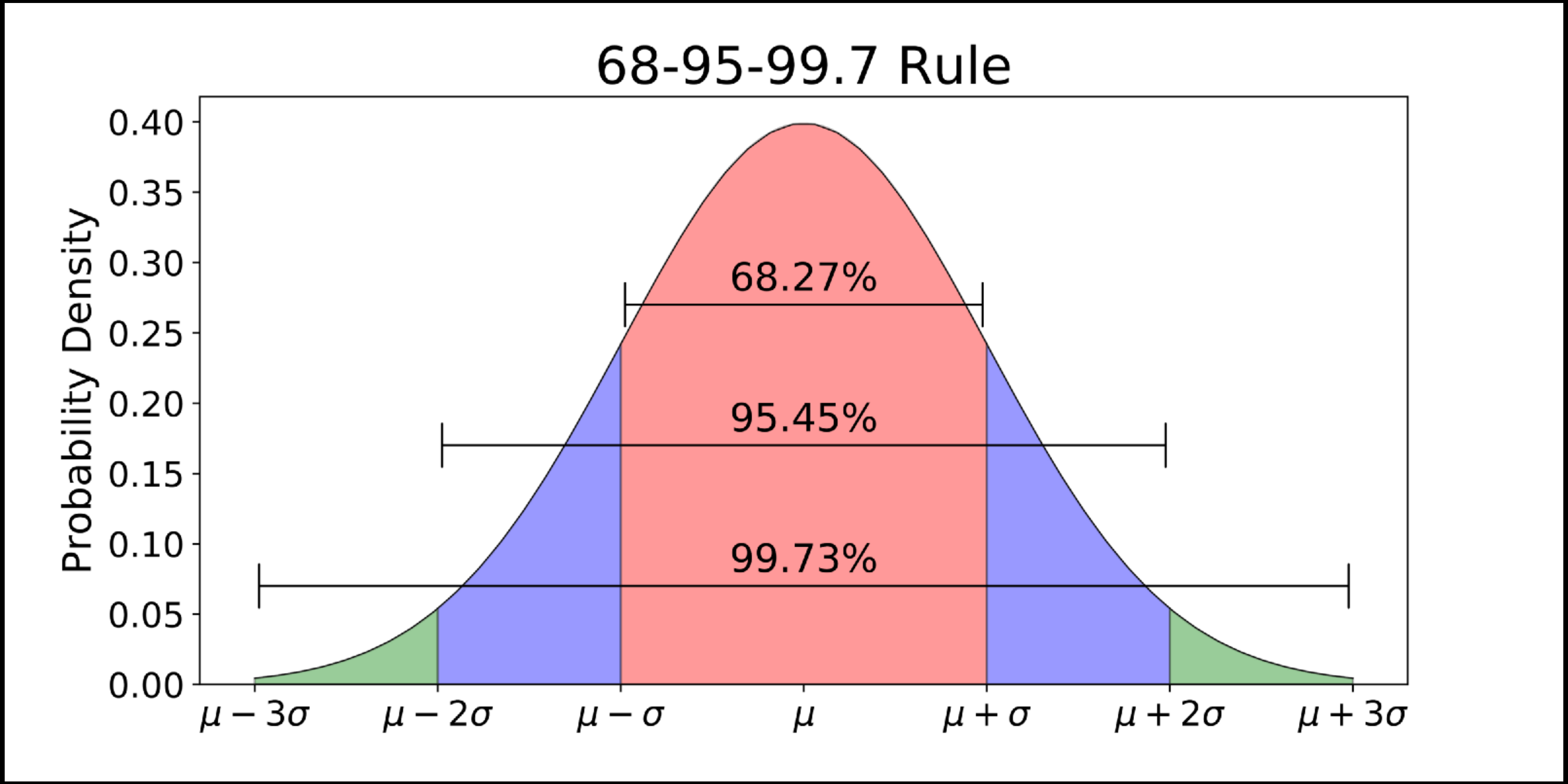
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$





N=50
N=10
N=3
N=1
500
450
400
350
300
250
200
150
100
50
0
0
10
20
30
40
50
60
70
80
90
100

Standard deviations and probability of the population mean



Type I & Type II errors

Null Hypothesis True	Did we reject	Example	Outcome	Probability we're right	Probability wrong
t	t	They say 10 minutes is the best, we say 4 but in fact they're correct (maybe we needed more N)	Type 1 error	99.970%	**0.030%**
t	f	They say 10 minutes is the best, we say 9 but 9 is within 95% so the null hypothesis holds.	Great, should have accepted it	**94.900%**	5.100%
f	f	They say 10 minutes is the best, we say 9 but 9 is within 95% so the null hypothesis holds. We are both wrong proven with higher N or another study is conducted and finds we sampled poorly	Type 2 error	94.900%	**5.100%**
f	t	They say 10 minutes is the best, we say 4 and we're correct (we increased N later on and verified)	Great, should have rejected it	**99.970%**	0.030%

Some examples

- The average time it takes to find Waldo at Comic Con by a random group of people over age 6 is 7 minutes. We surveyed 30 random people over age six and they found Waldo in an average time of 5 minutes. Who is right?
- Null hypo = mean = 7
- New hypo = mean = 5. N = 30, calculated sdev = 2.08, Z-Score = -.96
- Z-table look up gives = .33*2 → p = 66%. So there’s a 66% chance that the population mean is in this range. Margin of error = 5+-.36 66% confidence. Should we reject?

Time permitting in class examples with Waldo