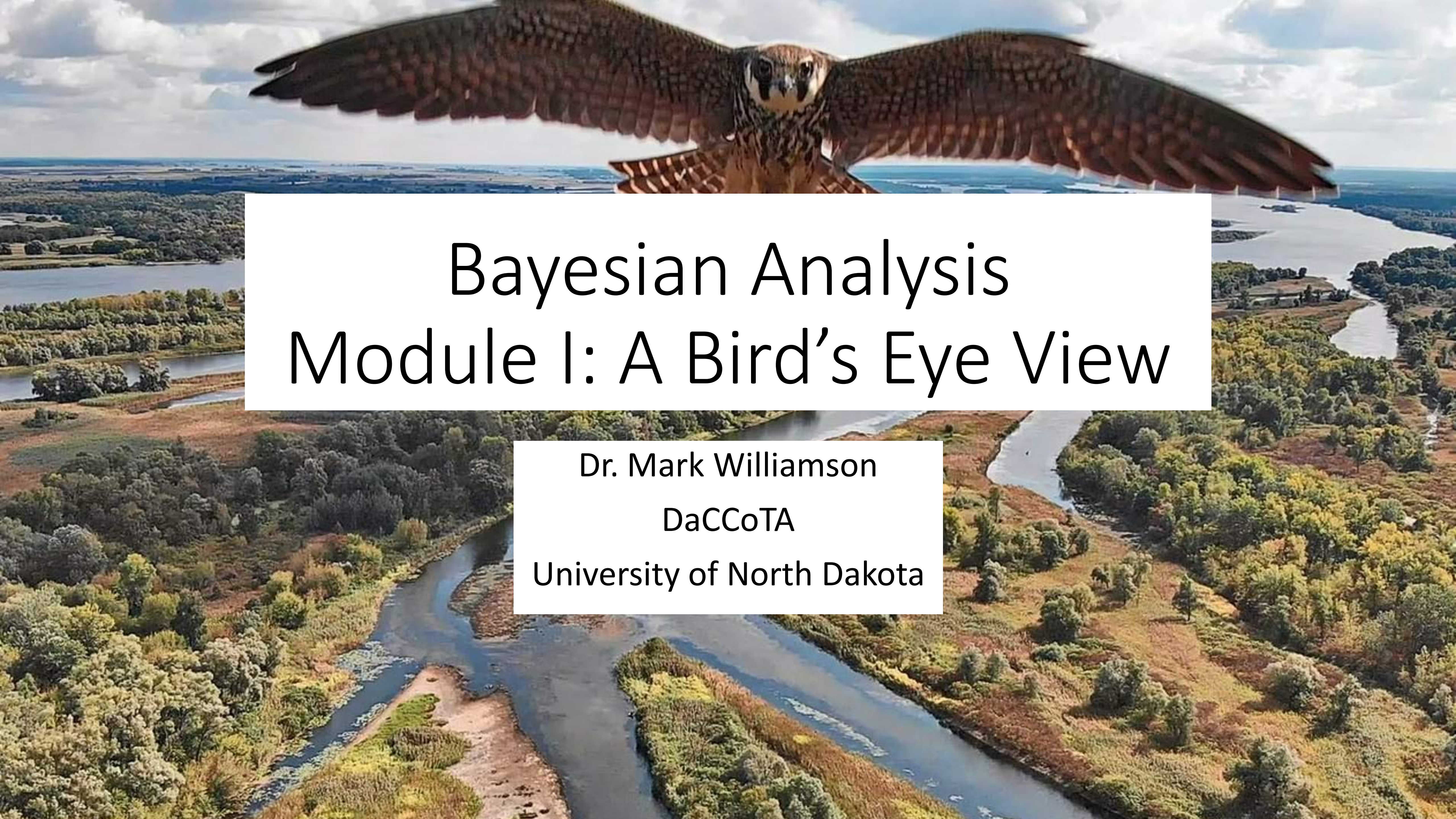


Bayesian Analysis
Module I: A Bird's Eye View

Dr. Mark Williamson

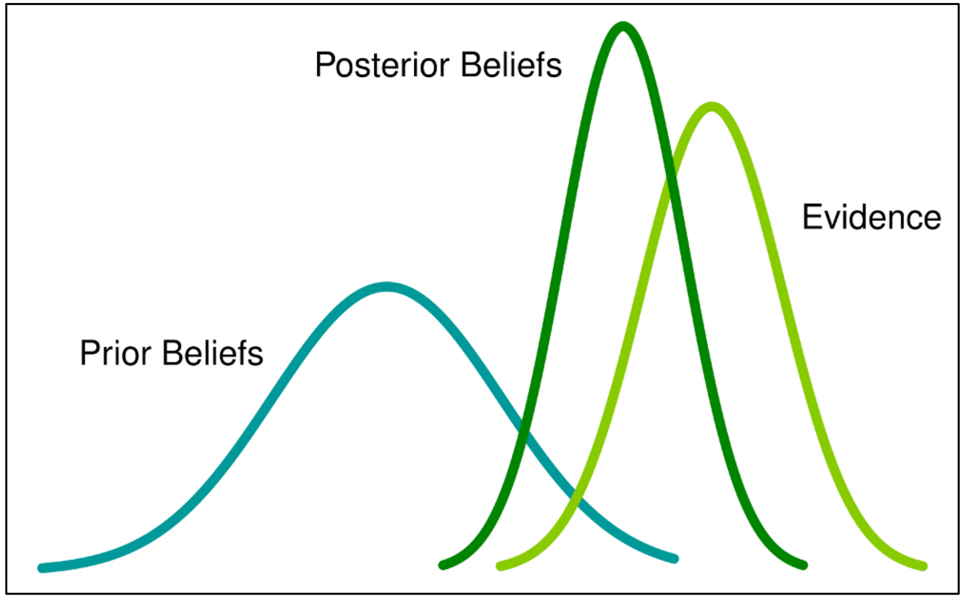
DaCCoTA

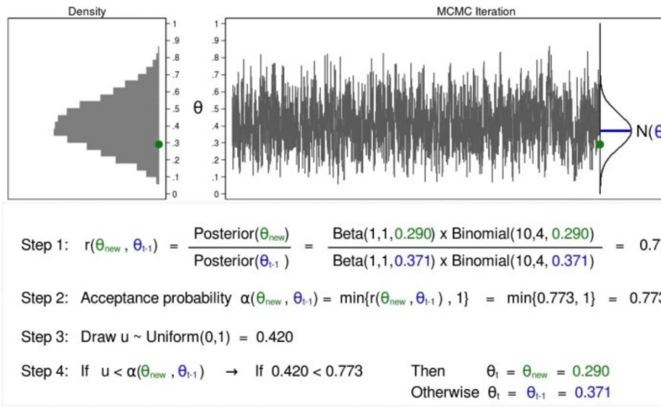
University of North Dakota

Introduction

What is Bayesian analysis?

- Method of assigning probability to events [1]
- Takes in prior beliefs (hypothesis) and evidence (data) to generate posterior beliefs (probability)
- Fundamental concept is Bayes' Theorem
- Characterized by intuitive theory but complicated computation in most actual usage
- Alternative to 'Frequentist' approaches
- Can use Bayesian approach for most Frequentist methods
- Encodes expert opinion and domain-specific knowledge into system [3]
- Predictions are a distribution of likely answers, allowing for risk assessment [3]





[2]

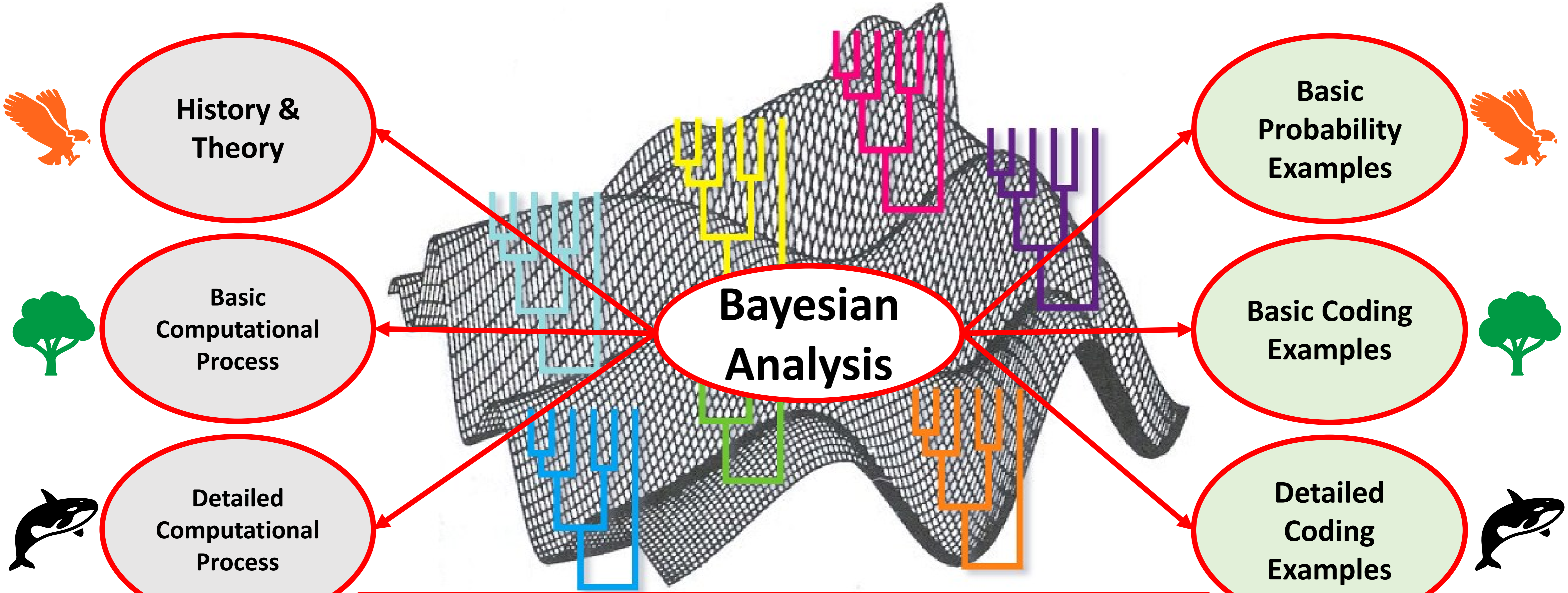
Bayes' Theorem:

$$P(hyp|data) = \frac{P(data|hyp)P(hyp)}{P(data)}$$

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Landscape



1) t-tests, 2) ANOVA, 3) Linear regression, 4) Generalized linear models

R SAS SPSS

History

Thomas Bayes

- English statistician, philosopher and Presbyterian minister [4]
- Born 1701 and died 1761
- Attended University of Edinburgh
- Bayes' Theorem taken from posthumous paper (1763)

Afterwards

- Pierre-Simon Laplace developed the Bayesian interpretation of probability
- Not commonly used unto the 1950's (philosophical and practical considerations)

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

History Cont.

Bayesians vs. Frequentists

Frequentist

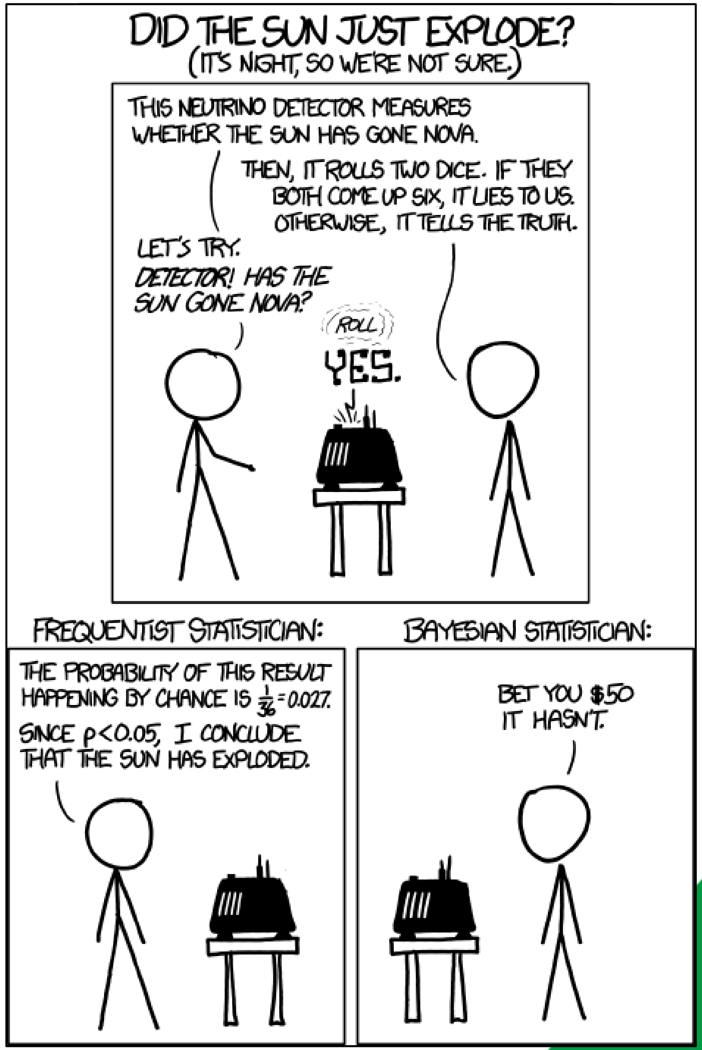
- probability of events in long term [5]
- depends on number coin flips, times experiment is run, sample size, etc.
- parameters are fixed, data is random [6]
- run to get p-value to test probability of test statistic, given H0[6]

Bayesian

- applies probabilities to statistical problems and provides updates based on evidence of new data
- does not depend on number of experiments
- parameters are random, data is fixed [6]
- run to get probability of parameter values, given observed data [6]

Comparisons

- Charitable: different tools [8]
- Neutral: reasonable preference [9]
- Spicy: others are wrong [10]



https://xkcd.com/1132/

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

Theory

Bayes' Theorem

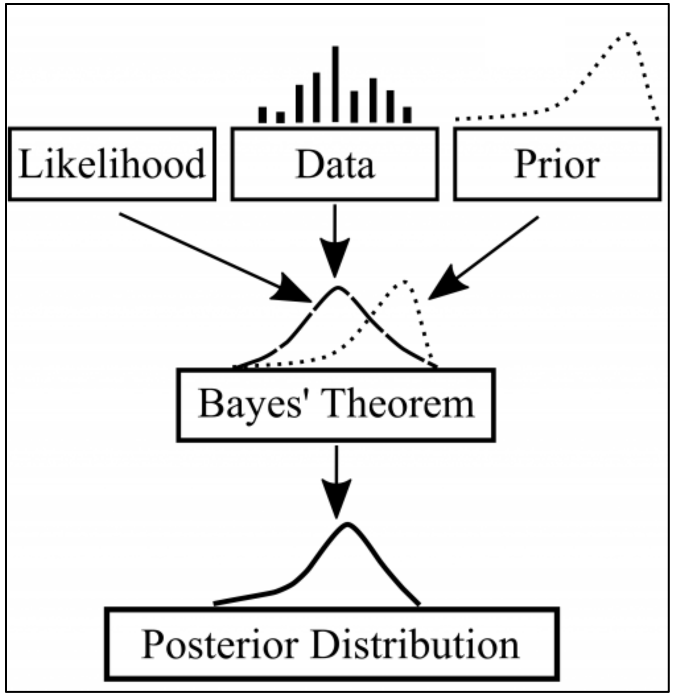
- Updates probabilities (degrees of belief) after obtaining new data
- Probability of A occurring given that B has occurred is equal to the probability that they have both occurred, relative to the probability that B has occurred [11]

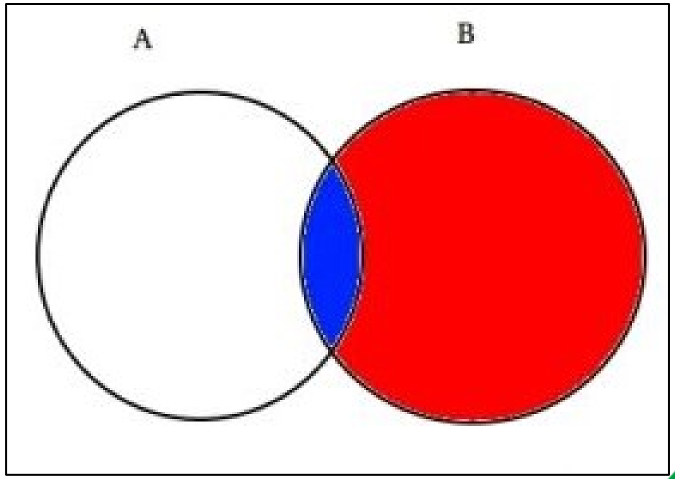
Bayes Theorem:

$$P(hyp|data) = \frac{P(data|hyp) \times P(hyp)}{P(data)}$$

Bayes Theorem:

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





Theory Cont.

Parts

Priors

- belief
- best-guess for how the data is distributed
- what is already known

Data

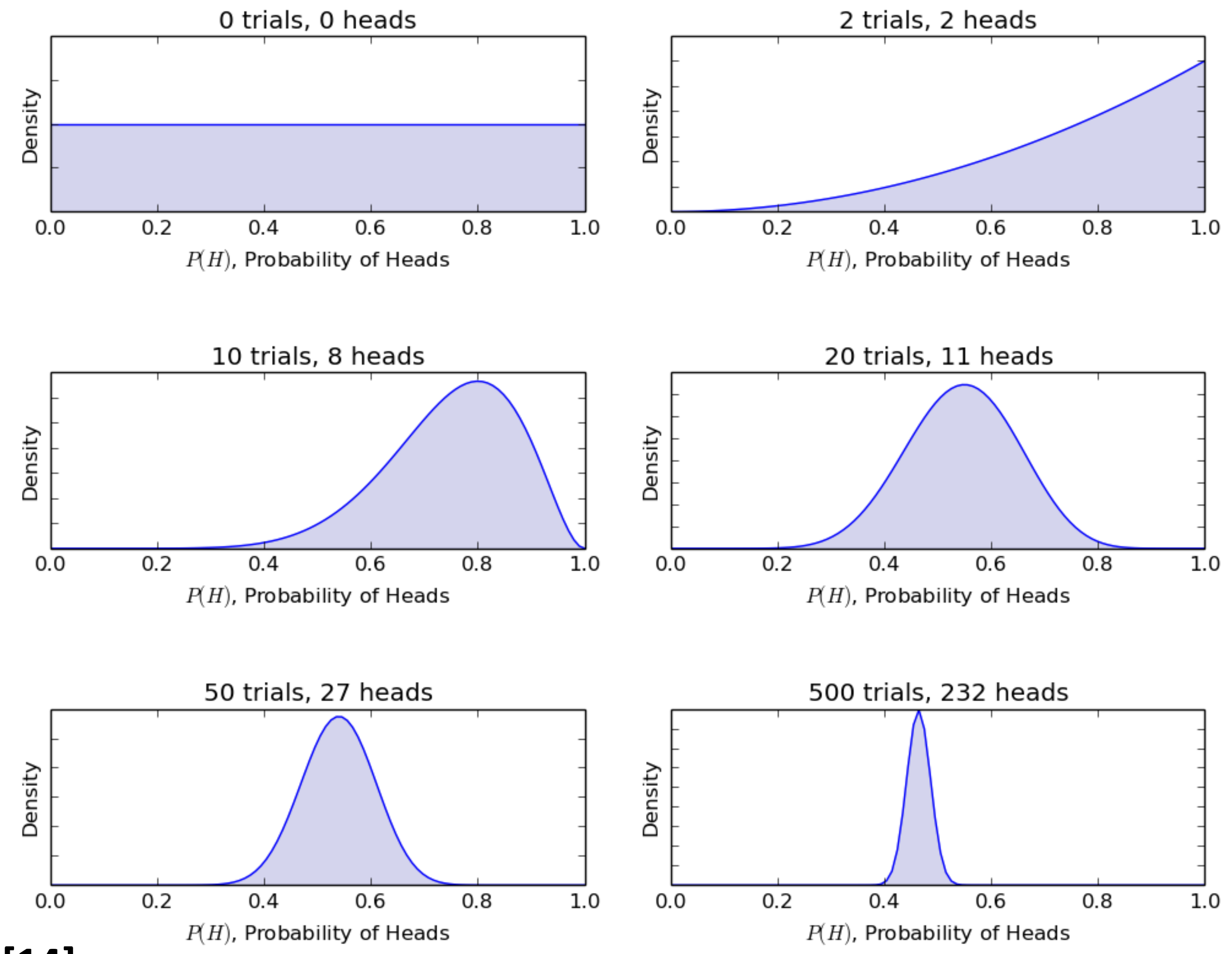
- values from experiment or observation

Likelihood

- conditional density of parameters given the data [13]

Posterior

- combined insights from the prior model and observed data
- iterative process



[14]

Examples

What is the probability of a patient having liver disease (A) if they are alcoholic (B)?

Clinical records show that 10% of patients entering have liver disease, 5% are alcoholic, and 7% of those with liver disease are diagnosed as alcoholics [15].

P(A)	
P(B)	
P(B|A)	

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

Examples

What is the probability of a patient having liver disease (A) if they are alcoholic (B)?

Clinical records show that 10% of patients entering have liver disease, 5% are alcoholic, and 7% of those with liver disease are diagnosed as alcoholics [15].

P(A)	0.10
P(B)	0.05
P(B|A)	0.07

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

$$\boldsymbol{\frac{0.07 \times 0.10}{0.05} = 0.14 = 14\%}$$

Examples

Dangerous fires are rare (1%), but smoke is common (10%) due to cooking and outdoor activities. 90% of dangerous fires make smoke [15].

What is the probability of dangerous fire when there is smoke?

P(A)	
P(B)	
P(B|A)	

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

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Examples

Dangerous fires are rare (1%), but smoke is common (10%) due to cooking and outdoor activities. 90% of dangerous fires make smoke [15].

What is the probability of dangerous fire when there is smoke?

P(A)	0.01
P(B)	0.10
P(B|A)	0.90

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

$$\frac{0.90 \times 0.01}{0.10} = 0.09 = 9\%$$

Examples

In four Formula 1 races, Niki won 3 while James won 1. However, it rained twice, once when when James won, once when Niki won [5].

What is the probability of James winning the next race (B), given the probability of rain (A)?

P(A)	
P(B)	
P(A|B)	

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

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

Examples

In four Formula 1 races, Niki won 3 while James won 1. However, it rained twice, once when when James won, once when Niki won [5].

What is the probability of James winning the next race (B), given the probability of rain (A)?

P(A)	**0.5 (rained at 2/4 races)**
P(B)	**0.25 (James won 1/4 races**
P(A|B)	**1.00 (rained every time James won)**

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

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

$$\frac{1.00 \times 0.25}{0.50} = 0.50 = 50\%$$

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Examples

There are two boxes. The first box contains 4 red balls and 2 green balls. The second box contains 4 green balls and 2 red balls. By design, the probabilities of selecting the first or second box are random; 1/3 for the first and 2/3 for the second. A box is selected at random, and a ball is selected at random from it [16].

What is the probability the ball was selected for the first box (A) if the ball is red (B)?

P(A)	
P(B)	
P(B|A)	

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

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Examples

There are two boxes. The first box contains 4 red balls and 2 green balls. The second box contains 4 green balls and 2 red balls. By design, the probabilities of selecting the first or second box are random; 1/3 for the first and 2/3 for the second. A box is selected at random, and a ball is selected at random from it [16].

What is the probability the ball was selected for the first box (A) if the ball is red (B)?

P(A)	**1/3 (from design)**
P(B)	**(4/6 x 1/3) + (2/6 x 2/3) = 2/9 + 2/9 = 4/9**
P(B|A)	**4/6 = 2/3 (from design)**

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

$$\frac{2/3 \times 1/3}{4/9} = 1/2 = 50\%$$

Real World Applications

General [17]

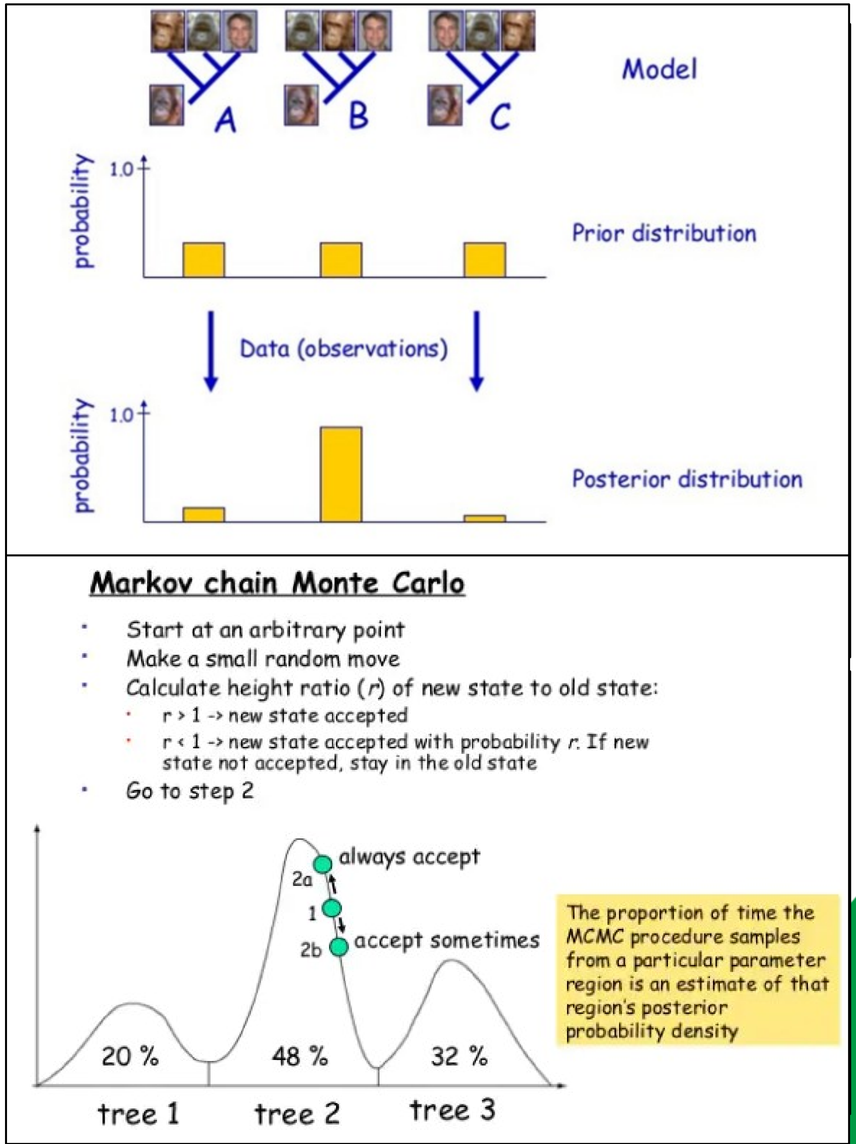
- Any application area where there is high heterogeneity or noisiness
- Any application where you need a clear understanding of your uncertainty
- Applications that require hierarchical models
- Examples: E-commerce, insurance, finance, and healthcare

Bayesian Network [18, 19]

- Probabilistic graphic model
- Natural science: gene regulatory network, medicine, biomonitoring, system biology
- Data science: document classification, semantic search, spam filter, information retrieval, image processing, turbo code

Phylogenetics [20, 21]

- Workhorse is tree reconstruction
- Uses DNA, amino acid, or morphological character alignments



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Assessment

qualtrics XM

UNIVERSITY OF NORTH DAKOTA

https://und.qualtrics.com/jfe/form/SV_1OlXzmd3JroFk90

Summary and Conclusion

- Bayesian analysis takes in prior beliefs (hypothesis) and evidence (data) to generate posterior beliefs (probability)
- A philosophically and computationally different method of statistics when compared to Frequentist methods
- Tune in next time for a more detailed look at Bayesian analysis in Bayesian Analysis Module II: Leaves and Trees

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

References

[1] https://www.quantstart.com/articles/Bayesian-Statistics-A-Beginners-Guide/

[2] https://i.ytimg.com/vi/OTO1DygELpY/maxresdefault.jpg

[3] https://towardsdatascience.com/what-is-bayesian-statistics-used-for-37b91c2c257c

[4] https://en.wikipedia.org/wiki/Bayesian_statistics

[5] https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/

[6] https://campus.datacamp.com/courses/bayesian-regression-modeling-with-rstanarm/introduction-to-bayesian-linear-models?ex=1

[7] https://sites.google.com/site/rbicss/home/seminars/bayesianvsfrequentistapproachinstatistics

[8] http://faculty.washington.edu/kenrice/BayesIntroClassEpi2018.pdf

[9] https://link.springer.com/content/pdf/10.3758/s13423-017-1266-z.pdf

[10] https://people.math.umass.edu/~lavine/whatisbayes.pdf

[11] https://www.quantstart.com/articles/Bayesian-Statistics-A-Beginners-Guide/

[12] https://www.vosesoftware.com/riskwiki/Bayesiananalysisexamplegenderofarandomsampleofpeople.php

[13] https://www.image.ucar.edu/pub/TOY07.4/nychka2a.pdf

[14] https://www.quantstart.com/articles/Bayesian-Statistics-A-Beginners-Guide/

[15] https://www.upgrad.com/blog/bayes-theorem-explained-with-example-complete-guide/

[16] https://www.analyzemath.com/probabilities/bayes-theorem.html

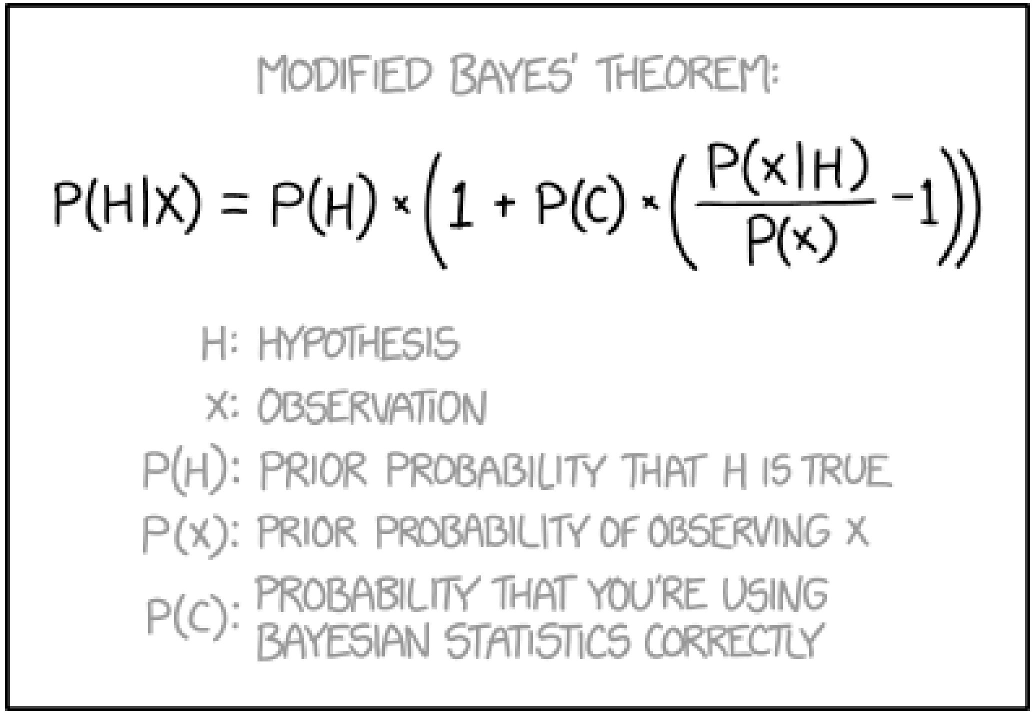
[17] https://towardsdatascience.com/what-is-bayesian-statistics-used-for-37b91c2c257c

[18] http://www.slideshare.net/OrochiKrizalid/bayesian-networks-13646169

[19] https://data-flair.training/blogs/bayesian-network-applications/

[20] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5624502/pdf/emss-73449.pdf

[21] https://www.slideshare.net/FredrikRonquist/bayesian-phylogenetic-inferencebig4ws20161010



https://xkcd.com/2059/

DaCCoTA
DAKOTA CANCER COLLABORATIVE
ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Acknowledgements

- The DaCCoTA is supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729.
- For the labs that use the Biostatistics, Epidemiology, and Research Design Core in any way, including this Module, please acknowledge us for publications. ***"Research reported in this publication was supported by DaCCoTA (the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729)".***

DaCCoTA
DAKOTA CANCER COLLABORATIVE
ON TRANSLATIONAL ACTIVITY