PGE 383 Subsurface Machine Learning

Lecture 2: Probability

Lecture outline:

- **Probability Definitions**
- **Venn Diagrams, Probability Operations, Frequentist Concepts**
- **Bayesian Concepts**

Dr. Pyrcz's Unsolicited Interview Advice

Michael Pyrcz, P.Eng., Professor, The University of Texas at Austin

My Experience: 13 years in industry, including the roles of team leader, research program manager and hiring coordinator for my R&D division. During this time I saw a lot of behaviors that impacted students' opportunities in their interviews. *I hope the following comments are helpful during this interview season.*

Internship is an Extended Interview: companies hire former interns; therefore, they are looking for candidates for an entire career! Their questions are attempting to discern your fit over the long-term, not just for the internship. *Understand context for the questions.*

Focus on Plan A: don't discuss your alternative options of going to another field or industry. Let them know that you are excited about the industry. *Don't look uncommitted.*

Professional Life is Stressful: be relaxed, show them that you handle yourself under stress. Work-life balance, regenerative hobbies are good. *Your career is a marathon.*

Always Learning / Flexible: be curious, interested, enthusiastic to learn new stuff, expect many roles over your long career. It'll be fun! *Ready to accept every challenge!*

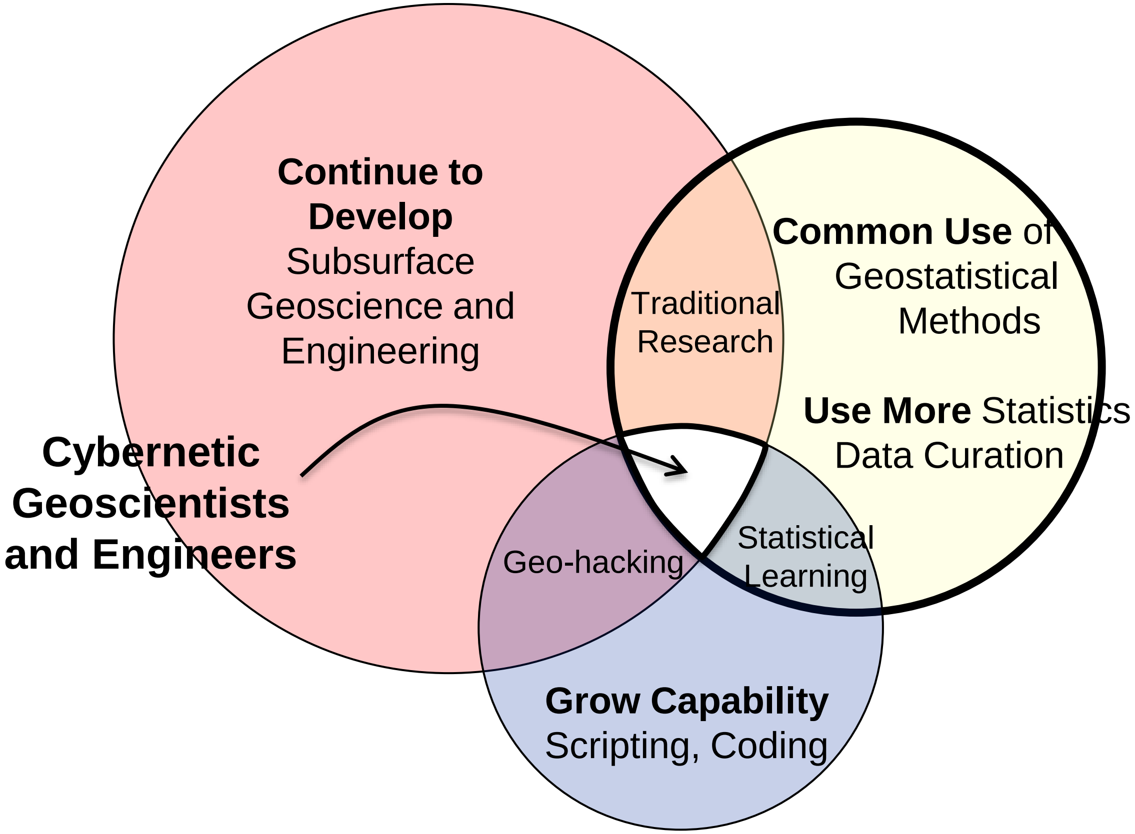
Be Concise: they need to get their job done. *Don't fill the time, attempt to baffle them!*

Be Friendly / Relaxed, but Formal: think about the people you like to work with and learn those traits. Harmonious fit in the team matters, because we all win together. *Synergy!*

Story of You: consider those experiences where you fell in love with the career, benefitted from healthy hobbies, learned new stuff, resolved conflict, worked in a team. *Be ready to tell.*

Motivation

Any machine learning should be driven from expertise on statistics, machine learning is statistical learning!



Proposed Venn diagram for a path forward for growing data science capabilities among engineers and geoscientists.

Probability and Statistics for Machine Learning

Why cover probability and statistics for machine learning?

- Machine learning is statistical learning!
- Machine learning methods make predictions based on:
  - sample data (training data)
  - summarizations / statistics (model parameters)
  - probability models (etc. priors, conditionals, maximum a posteriori estimation)
- Therefore robust use of probability and statistics is a prerequisite to machine learning!

Probability and Statistics

Recorded Lectures

- frequentist probability

https://www.youtube.com/watch?v=NnQeospi6Qg

- Bayesian probability

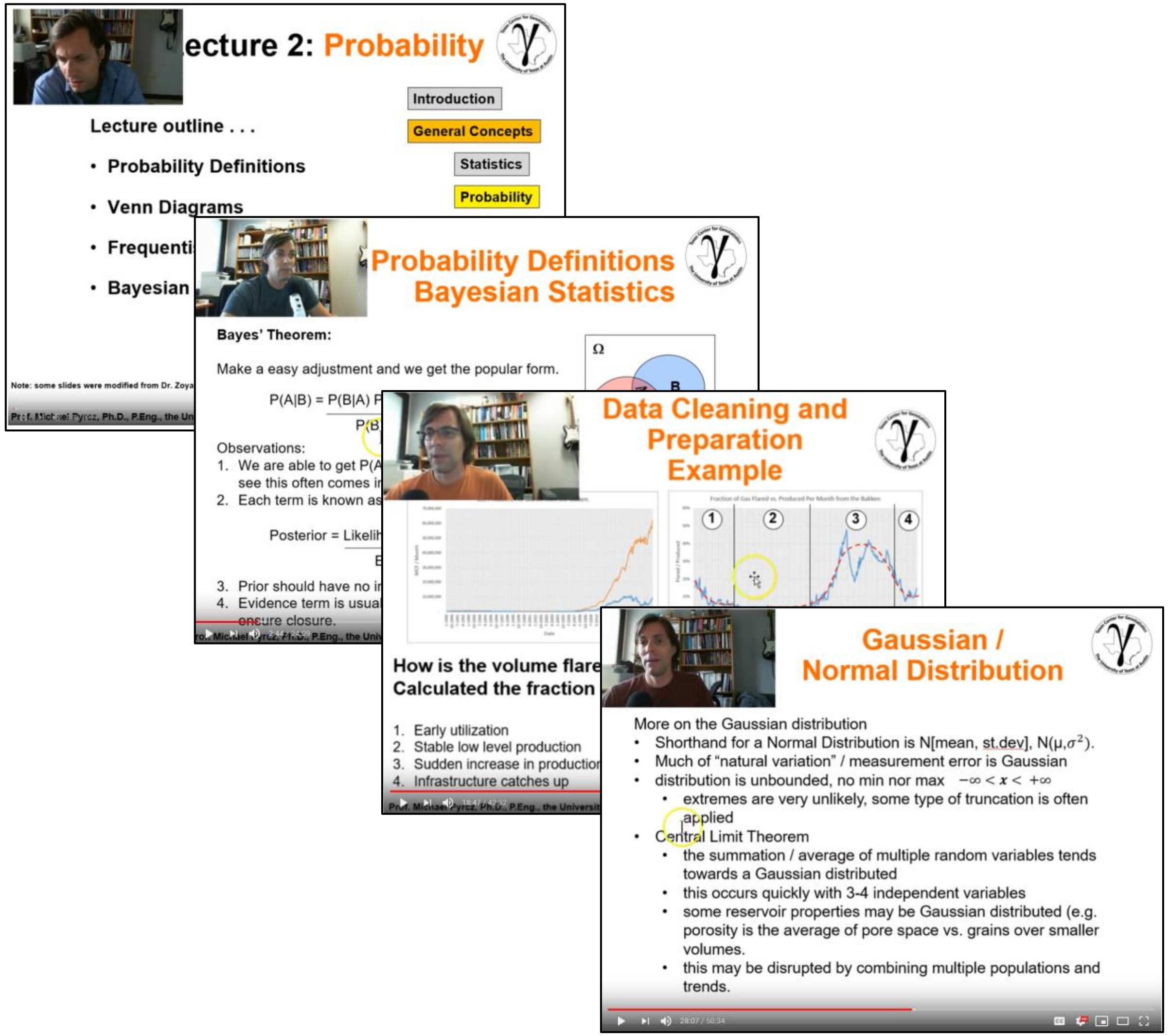
https://www.youtube.com/watch?v=We8wt2qbZ0c

- statistics

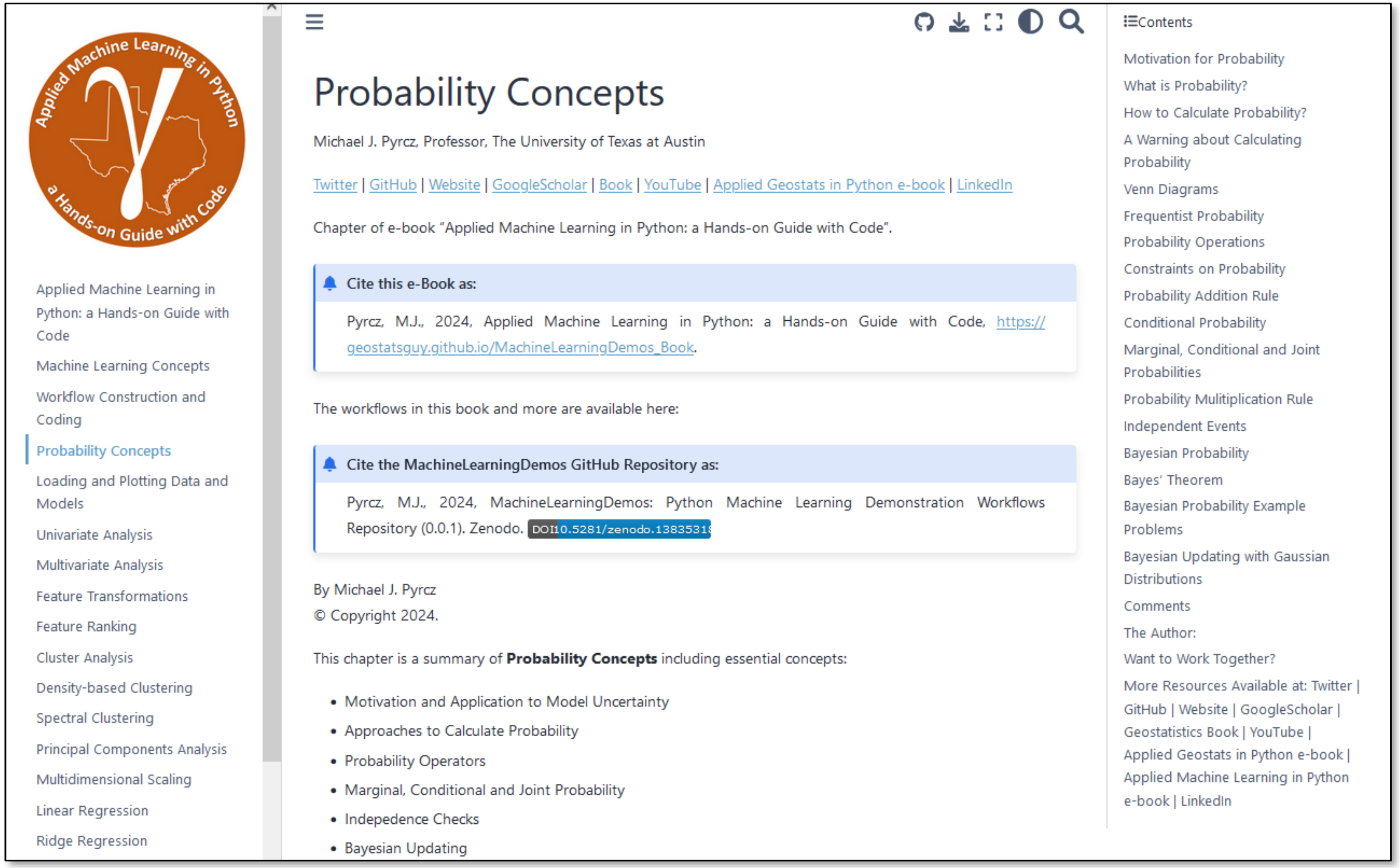
https://www.youtube.com/watch?v=OEvELm66NNo&t=61s

- univariate distributions

https://www.youtube.com/watch?v=U7fGsqCLPHU



Probability and Statistics



Probability Concepts

Michael J. Pyrcz, Professor, The University of Texas at Austin

Twitter | GitHub | Website | GoogleScholar | Book | YouTube | Applied Geostats in Python e-book | LinkedIn

Chapter of e-book "Applied Machine Learning in Python: a Hands-on Guide with Code".

Cite this e-Book as:

Pyrcz, M.J., 2024, Applied Machine Learning in Python: a Hands-on Guide with Code, https://geostatsguy.github.io/MachineLearningDemos_Book.

The workflows in this book and more are available here:

Cite the MachineLearningDemos GitHub Repository as:

Pyrcz, M.J., 2024, MachineLearningDemos: Python Machine Learning Demonstration Workflows Repository (0.0.1). Zenodo. DOI 10.5281/zenodo.1383531

By Michael J. Pyrcz

This chapter is a summary of **Probability Concepts** including essential concepts:

- Motivation and Application to Model Uncertainty
- Approaches to Calculate Probability
- Probability Operators
- Marginal, Conditional and Joint Probability
- Indepedence Checks
- Bayesian Updating

Contents

Applied Machine Learning in Python, Workflow Probability Concepts Chapter.

PGE 383 Subsurface Machine Learning

Lecture 2: Probability

Lecture outline:

- **Probability Definitions**

Big Data Criteria

Big Data, you have big data if your data has a combination of these:

Volume: many data samples, difficult to handle and visualize

Velocity: high rate collection, continuous relative to decision making cycles

Variety: data form various sources, with various types and scales

Variability: data acquisition changes during the project

Veracity: data has various levels of accuracy

Seismic acquisition by MV Geo Coral, capable of towing 8.5 km streamers in a 1 km spread.

"Energy has been big data long before tech learned about big data."

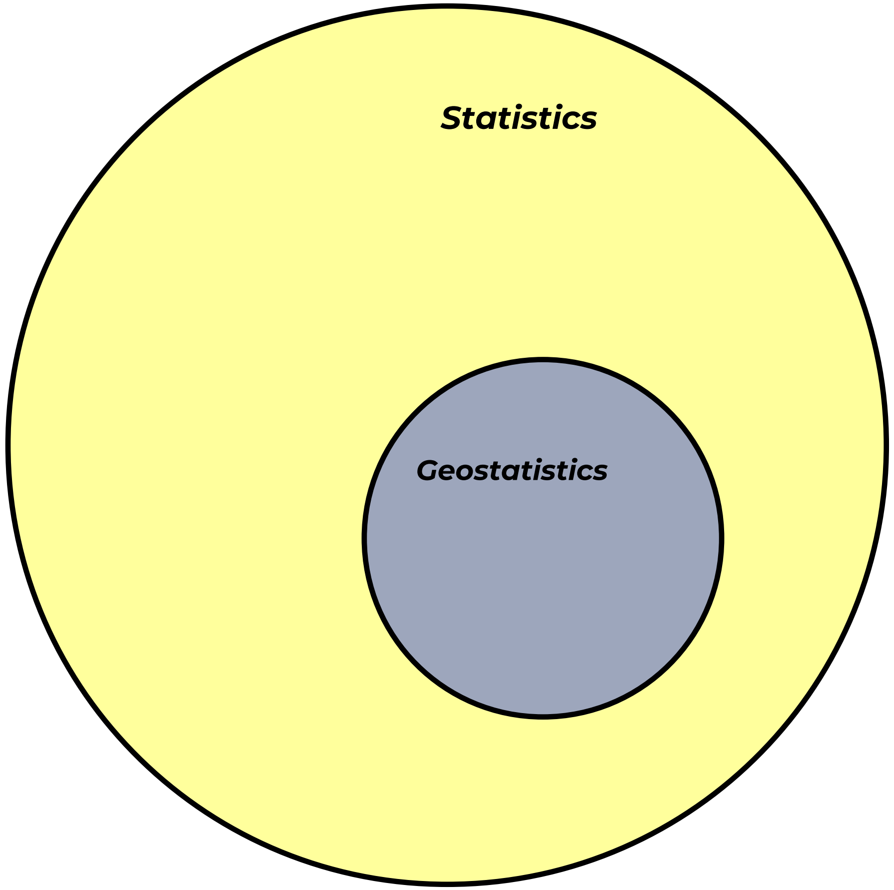
– Michael Pyrcz

Big Data Analytics

Statistics is collecting, organizing, and interpreting data, as well as drawing conclusions and making decisions.

Geostatistics is a branch of applied statistics:

1. the spatial (geological) context
2. the spatial relationships
3. volumetric support
4. uncertainty



Proposed Venn diagram for spatial data analytics.

Big Data Analytics

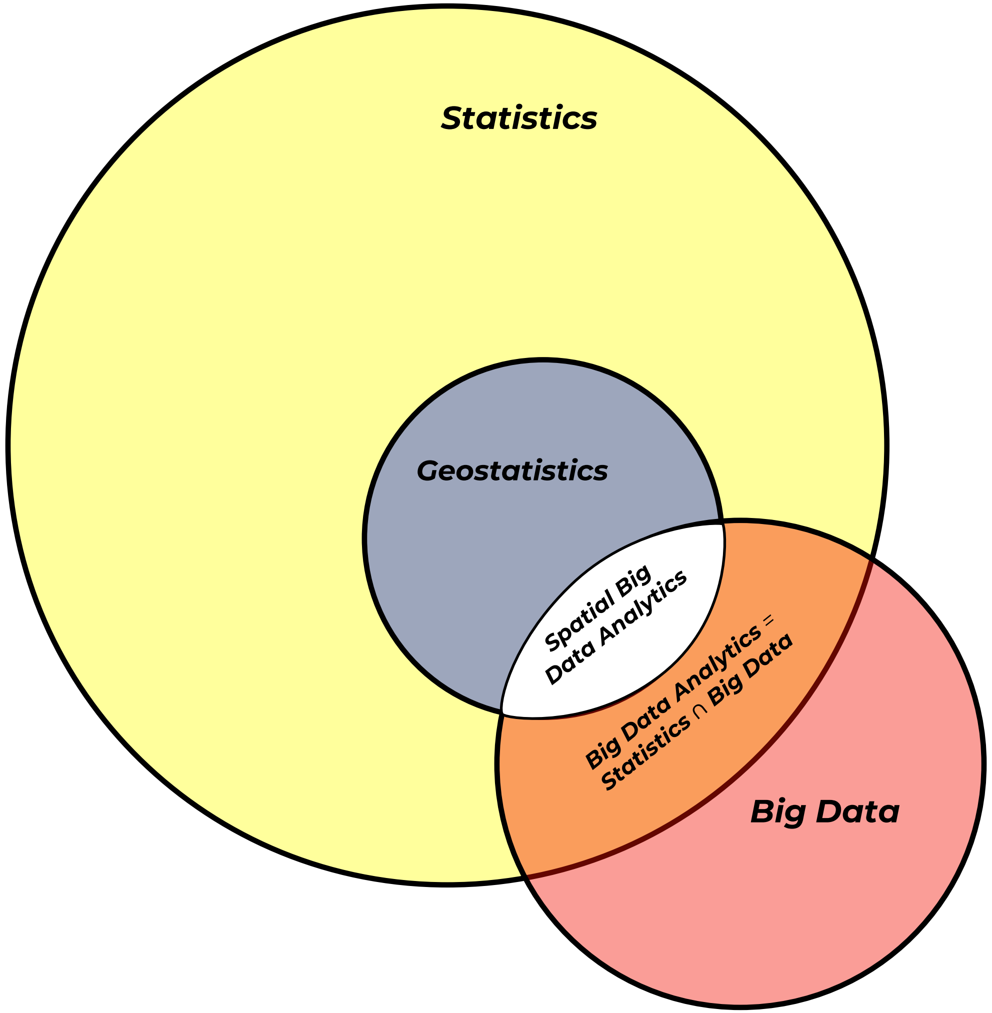
Data Analytics is the analysis of data to support decision making, virtually the same as statistics!

Big Data Analytics is the process of examining large and varied data sets to discover patterns and make decisions.

Spatial Big Data Analytics is expert use of spatial statistics / geostatistics on big data to support decision making.

"Data analytics is the use of statistics and visualization"

– Michael Pyrcz



Proposed Venn diagram for spatial big data analytics.

Probability Helps in Making Decisions

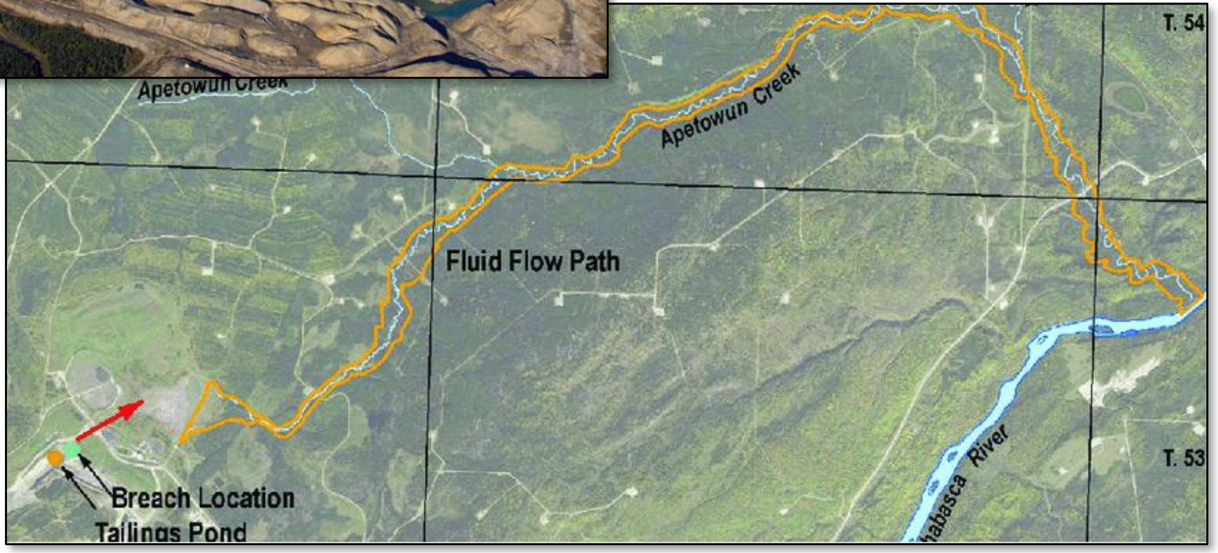
For example:

- What is the probability that a well is a success? – *drill the well*
- What is the probability that a valve has a crack? – *replace the valve*
- What is the probability that a seismic survey finds a reservoir? – *acquire the seismic*
- What is the probability that a reservoir seal will fail? – *inject the CO2*

Most of our decisions involve uncertainty:

- By quantifying probability, we can make better decisions.
- By communicating uncertainty our work is use to support decision making!

Obed Mountain open-pit thermal coal mine near Hinton, Alberta, Canada. (https://www.alamy.com/stock-photo/obed-coal-mine.html?sortBy=relevant/)



Path of contamination due to the Obed Mine tailings dam failure (https://poeschlab.ualberta.ca/2014/10/19/dr-poesch-writes-alberta-outdoorsman-article-on-the-worst-disaster-nobody-has-heard-of-in-regards-to-the-obed-mine-spill/)

What is Probability?

TBD

What is Probability?

A Measure that Honors Kolmogorov's 3 Axioms:

1. Probability of an event is a non-negative number.

$$\text{Prob}(A) \geq 0$$

2. Unit Measure, probability of the entire sample space is one (unity).

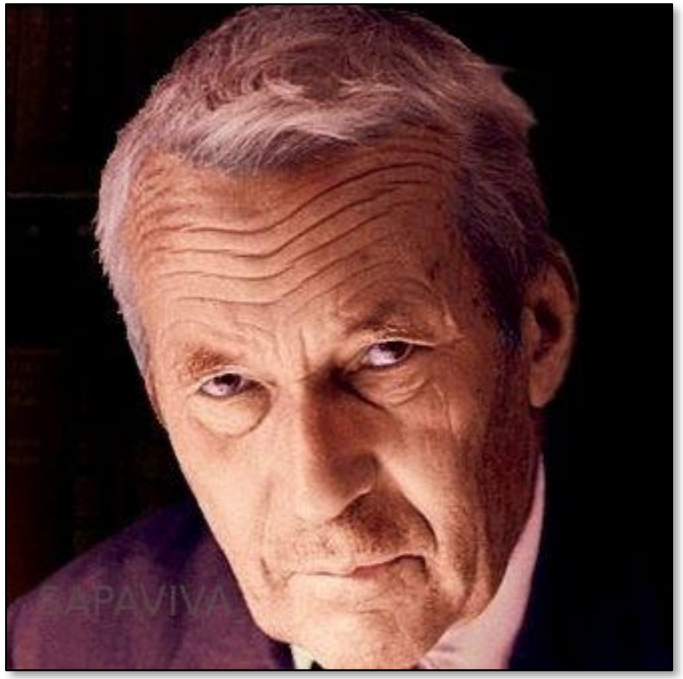
$$\text{Prob}(\Omega) = 1$$

3. Additivity of mutually exclusive events for unions.

$$\text{Prob}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} Prob(A_i)$$

e.g., probability of A_1 and A_2 mutual exclusive events is $Prob(A_1) + Prob(A_2)$

Note, Origin of Probability – gamblers' dispute 1654, 1812 Laplace's definition, and Kolmogorov's Theory in 1933!



Andrey Kolmogorov (1903 –1987), Soviet mathematician.

What is Probability?

The 3 Probability Perspectives:

1. **Long-term frequencies**
   - Probability as ratio of outcomes
   - Requires repeated observations of an experiment

Frequentist Probability

2. **Physical tendencies / propensities**
   - Knowledge about the system
   - Could know the probability of coin toss without the experiment

Engineering & Science

3. **Degrees of belief**
   - Reflect our certainty about a result
   - Very flexible, assign probability to anything, updating with new information

Bayesian Probability

Frequentist Probability Definition

Measure of the likelihood that an event will occur. For random experiments and well-defined settings (such as coin tosses):

$$\text{Prob}(A) = \text{P}(A) = \lim_{n \to \infty} \left(\frac{n(A)}{n(\Omega)} \right)$$

frequentist approach to probability is the limit of relative frequency over a large number of trails.

where:

n(A) = number of times event A occurred
n(Ω) = number of trails, often represented by n.

Example: Possibility of drilling a dry hole for the next well, encountering sandstone at a location ($\mathbf{u}_\alpha$), exceeding a rock porosity of 15% at a location ($\mathbf{u}_\alpha$).

Lecture outline:

- Venn Diagrams, Probability Operations, Frequentist Concepts

Probability Concepts
Venn Diagrams

Venn Diagrams are a tool to communicate probability

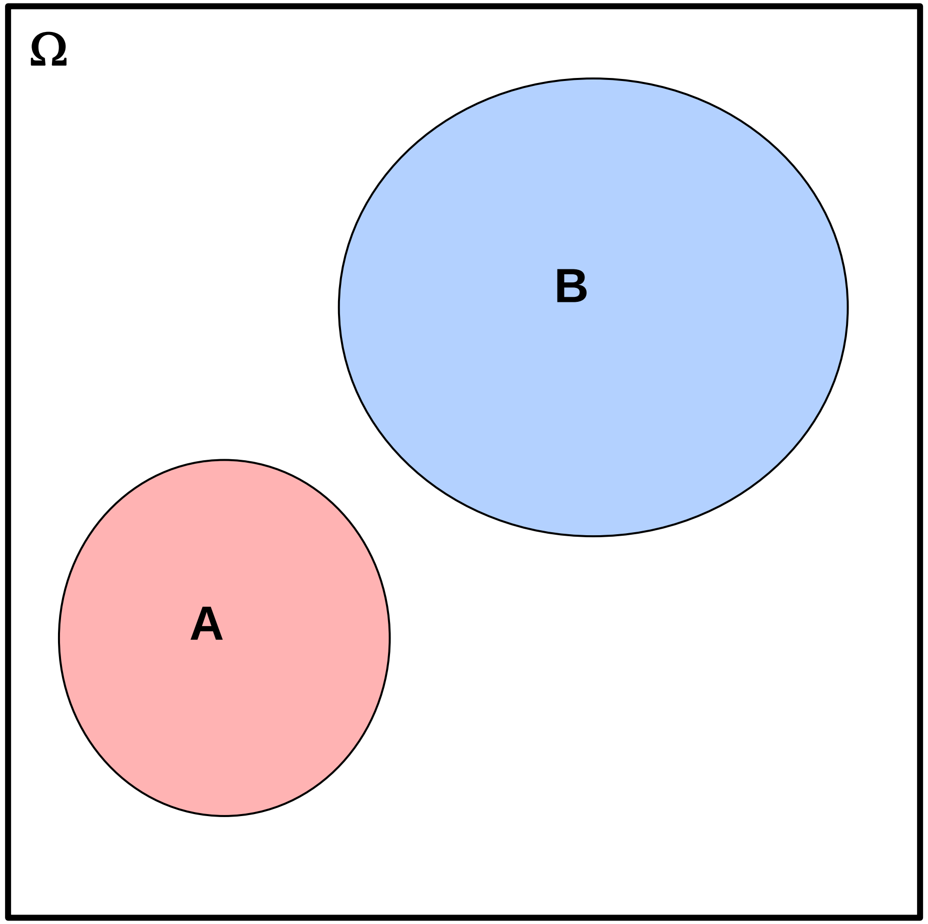
Experiments (Sampling) (J): Establishment of conditions that produce an outcome.

Events (A, B, …): the possible outcomes.

Sample Space (Ω): Collection of all possible events.

What do we learn from a Venn diagram?

- size of regions = probability of occurrence
- overlap = probability of joint occurrence
- excellent tool to visualize marginal, joint and conditional probability.



Venn Diagram – illustration of events and relations to each other.

Probability Concepts
Venn Diagrams Example

Experiments (Sampling):

- Facies determined from a set of well cores ($N = 3{,}000$ measures at 1 foot increments)

Sample Space (Ω):

- Facies for the $N = 3{,}000$ core measures

Events:

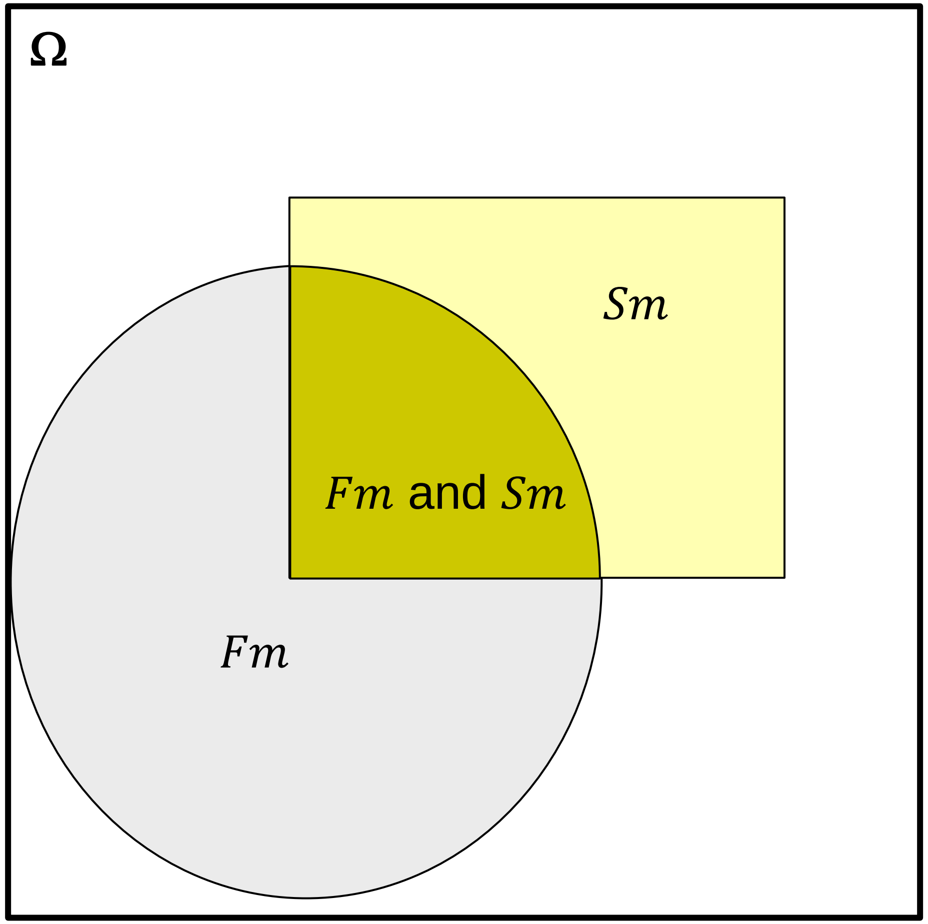
- Facies = {Fm (mudstone), Sm (massive sandstone)}

Venn Diagram Tells Us About Probability, e.g.:

$P(Fm) > P(Sm)$ ⟶ Fm is more likely than Sm

$Prob\{Fm \text{ and } Sm\} > 0.0$ ⟶ Fm and Sm occur together

$Prob\{Sm\} > Prob\{Fm \text{ and } Sm\}$ ⟶ Sm is more like than both Sm and Fm together



Venn Diagram – illustration of events and relations to each other.

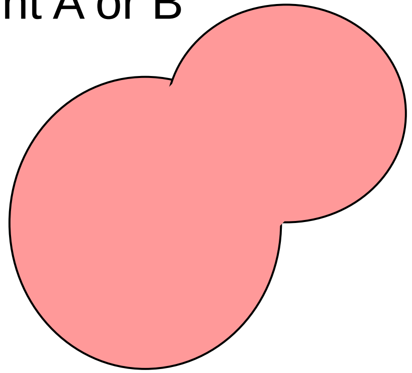
Probability Definitions
Probability Operators

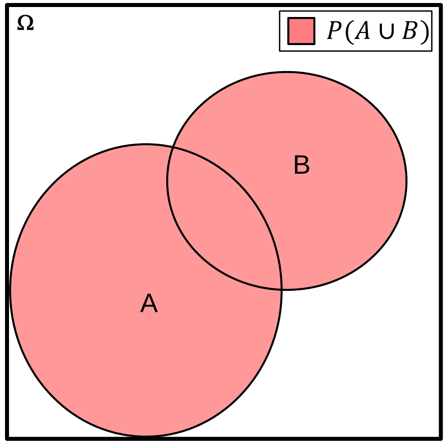
Union of Events:

- All outcomes in the sample space that belong to either event A or B

Set notation for samples in union → $\boldsymbol{A \cup B = \{x: x \in A \text{ or } x \in B\}}$

Probability notation for union → $\boldsymbol{P(A \cup B)}$





Venn Diagram – illustrating union.

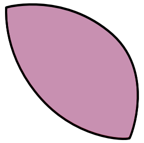
Intersection of Events:

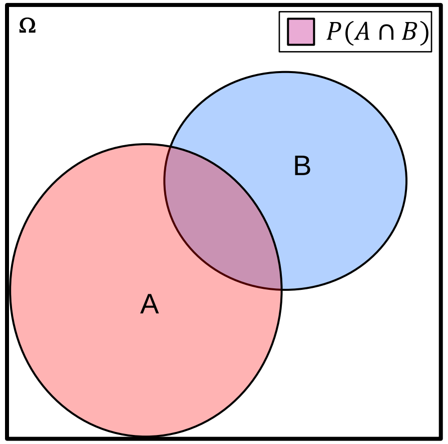
- All outcomes in the sample space that belong to both events A and B

- We will call this a joint probability later, $P(A,B)$

Set notation for samples in intersection → $\boldsymbol{A \cap B = \{x: x \in A \text{ and } x \in B\}}$

Probability notation for intersection → $\boldsymbol{P(A \cap B)}$ or $\boldsymbol{P(A,B)}$





Venn Diagram – illustrating intersection.

Probability Definitions
Probability Operators

Complementary Events:

- All outcomes in the sample space that do not belong to A

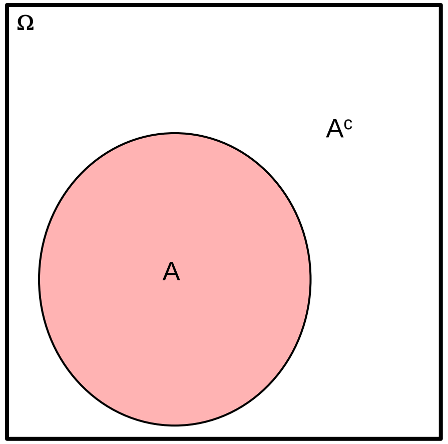
Set notation for samples in compliment → $\boldsymbol{A^c = \{x : x \notin A\}}$

Probability notation for compliment → $\boldsymbol{P(A^c)}$

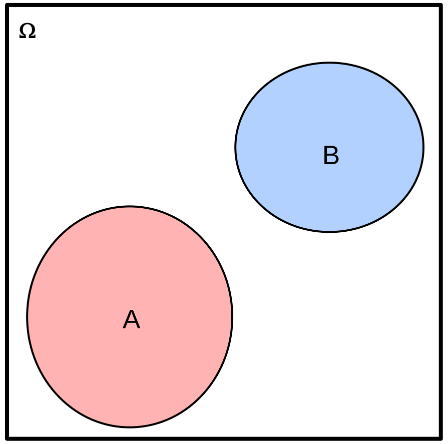
Mutually Exclusive Events:

- The events that do not intersect or do not have any common outcomes

$$A \cap B = \emptyset \rightarrow \text{Null Set}$$



Venn Diagram – illustrating complimentary events.



Venn Diagram – illustrating mutual exclusive events.

Probability Definitions
Probability Operators

Exhaustive, Mutually Exclusive Sequence of Events:

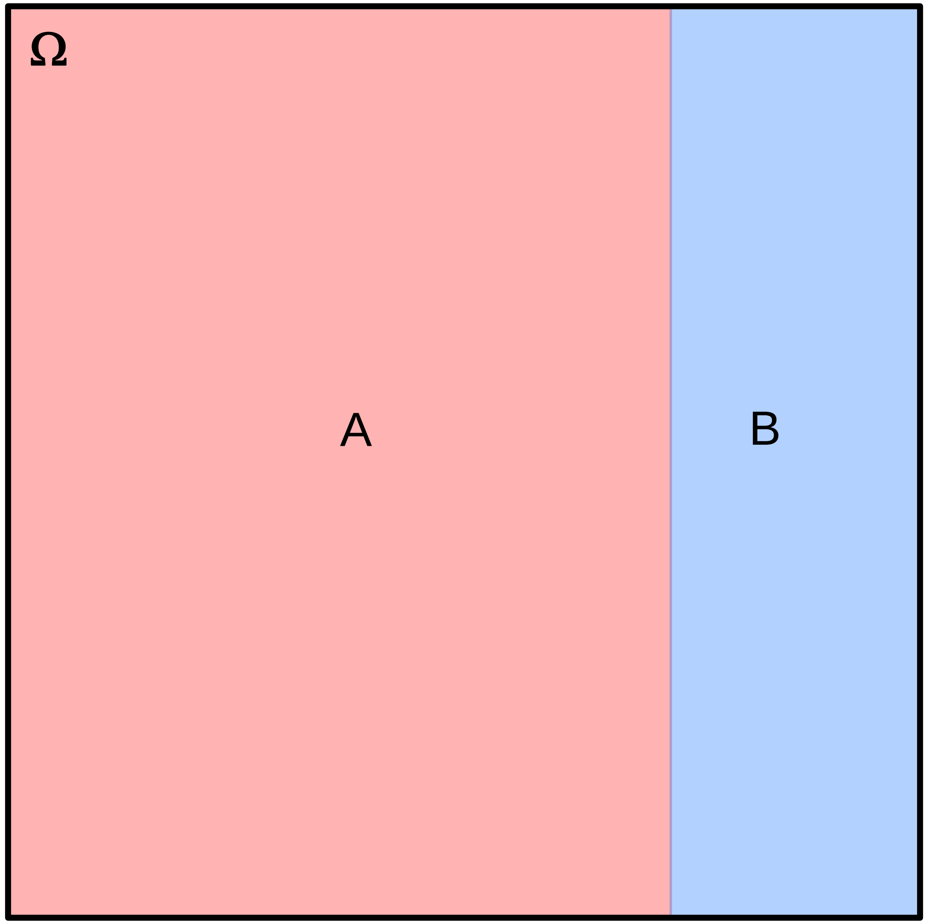
The sequence of events whose union is equal to the sample space, all-possible events (**Ω**):

$$A_1 \cup A_2 \cup \; \ldots \cup A_n = \Omega$$

and there is no intersection between events:

$$A_i \cap A_j = \emptyset \,, \forall \; i \neq j$$

For example, all the samples are either A or B



Venn Diagram – illustrating exhaustive, mutually exclusive events.

Probability Concepts Venn Diagrams Example

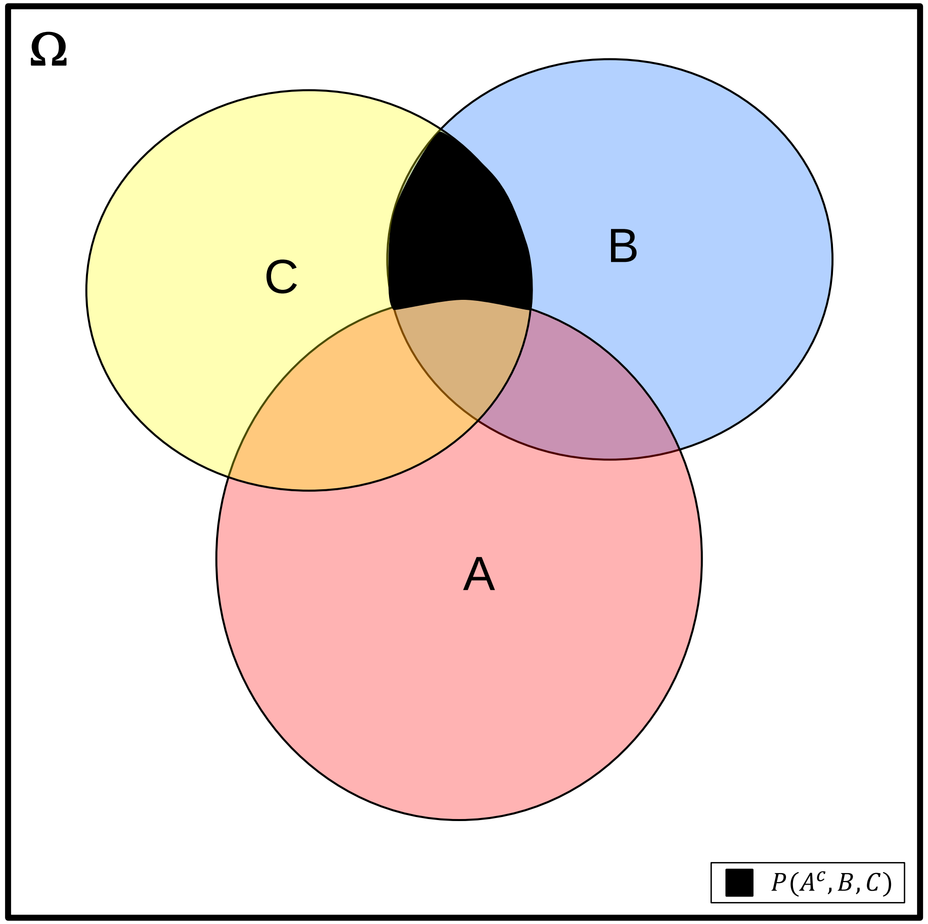
We can get more complicated, with multiple events:

Define:

- **A**: oil present (**A^C**: dry hole)
- **B**: Sm (**B^C**: Fm)
- **C**: porosity $\geq$ 15% (**C^C**: porosity < 15%)

What is the probability of dry hole with massive sandstone (Sm) and porosity $\geq$ 15%?

$$\text{Prob}(\boldsymbol{A^C} \cap \boldsymbol{B} \cap \boldsymbol{C}) = \text{Area}(\boldsymbol{A^C} \cap \boldsymbol{B} \cap \boldsymbol{C}) \,/\, \text{Area}(\Omega)$$



Venn Diagram – illustrating a more complicated 3 event probability.

Frequentist Probability Hands-on

Example: A petrophysicist has measured porosity of 7 core samples from a carbonate formation in the laboratory as follows:

Sample Space: {0.10, 0.12, 0.14, 0.15, 0.17, 0.19, 0.25}

A **C** **B**

We would like to investigate the following events, **find the samples for each event**:

- Event A: Porosity values of less than 0.15, {0.10, 0.12, 0.14}
- Event B: Porosity values of greater than 0.20, {0.25}
- Event C: Porosity values of between 0.14 and 0.17 {0.14, 0.15, 0.17}

Union of Events: $A \cup B$ $B \cup C$ $A \cup C$

Intersection of Events: $A \cap B$ $B \cap C$ $A \cap C$

Complementary Events: A^c B^c C^c

All Events: $A \cup B \cup C$

Find the sets (group of samples) that satisfy these conditions.

Frequentist Probability Hands-on

Example: A petrophysicist has measured porosity of 7 core samples from a carbonate formation in the laboratory as follows:

Sample Space: {0.10, 0.12, 0.14, 0.15, 0.17, 0.19, 0.25}

A **C** **B**

We would like to investigate the following events, **find the samples for each event**:

- Event A: Porosity values of less than 0.15, {0.10, 0.12, 0.14}
- Event B: Porosity values of greater than 0.20, {0.25}
- Event C: Porosity values of between 0.14 and 0.17 {0.14, 0.15, 0.17}

Union of Events: $A \cup B = \{0.10, 0.12, 0.14, 0.25\}$ $B \cup C = \{0.14, 0.15, 0.17, 0.25\}$ $A \cup C = \{0.10, 0.12, 0.14, 0.15, 0.17\}$

Intersection of Events: $A \cap B = \phi$ $B \cap C = \phi$ $A \cap C = \{0.14\}$

Complementary Events: $A^c = \{0.15, 0.17, 0.19, 0.25\}$ $B^c = \{0.10, 0.12, 0.14, 0.15, 0.17, 0.19\}$ $C^c = \{0.10, 0.12, 0.19, 0.25\}$

All Events: $A \cup B \cup C = \{0.10, 0.12, 0.14, 0.15, 0.17, 0.25\}$

Find the sets (group of samples) that satisfy these conditions.

Frequentist Probability Hands-on

Example: A petrophysicist has measured porosity of 7 core samples from a carbonate formation in the laboratory as follows:

Sample Space: {0.10, 0.12, 0.14, 0.15, 0.17, 0.19, 0.25}

A C B

We would like to investigate the following events, **estimate the probabilities of each event**:

- Event A: Porosity values of less than 0.15, {0.10, 0.12, 0.14} $P(A)$ = 3/7
- Event B: Porosity values of greater than 0.20, {0.25} $P(B)$ = 1/7
- Event C: Porosity values of between 0.14 and 0.17 {0.14, 0.15, 0.17} $P(C)$ = 3/7

Union of Events: $A \cup B = \{0.10, 0.12, 0.14, 0.25\}$ $B \cup C = \{0.14, 0.15, 0.17, 0.25\}$ $A \cup C = \{0.10, 0.12, 0.14, 0.15, 0.17\}$

Intersection of Events: $A \cap B = \phi$ $B \cap C = \phi$ $A \cap C = \{0.14\}$

Complementary Events: $A^c = \{0.15, 0.17, 0.19, 0.25\}$ $B^c = \{0.10, 0.12, 0.14, 0.15, 0.17, 0.19\}$ $C^c = \{0.10, 0.12, 0.19, 0.25\}$

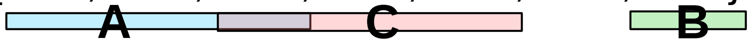
All Events: $A \cup B \cup C = \{0.10, 0.12, 0.14, 0.15, 0.17, 0.25\}$

Find the sets (group of samples) that satisfy these conditions.

Frequentist Probability Hands-on

Example: A petrophysicist has measured porosity of 7 core samples from a carbonate formation in the laboratory as follows:

Sample Space: {0.10, 0.12, 0.14, 0.15, 0.17, 0.19, 0.25}



We would like to investigate the following events, **estimate the probabilities of each event**:

- Event A: Porosity values of less than 0.15, {0.10, 0.12, 0.14} $P(A)$ = 3/7
- Event B: Porosity values of greater than 0.20, {0.25} $P(\mathrm{B})$ = 1/7
- Event C: Porosity values of between 0.14 and 0.17 {0.14, 0.15, 0.17} $P(C)$ = 3/7

Union of Events:

$A \cup B = \{0.10, 0.12, 0.14, 0.25\}$ $P(A \cup B) = 4/7$

$B \cup C = \{0.14, 0.15, 0.17, 0.25\}$ $P(B \cup C) = 4/7$

$A \cup C = \{0.10, 0.12, 0.14, 0.15, 0.17\}$ $P(A \cup C) = 5/7$

Intersection of Events:

$A \cap B = \phi$, $\mathrm{P}(A \cap B) = 0$

$B \cap C = \phi$, $\mathrm{P}(B \cap C) = 0$

$A \cap C = \{0.14\}$, $P(A \cap C) = 1/7$

Complementary Events: $A^c = \{0.15, 0.17, 0.19, 0.25\}$ P = 4/7 $B^c = \{0.10, 0.12, 0.14, 0.15, 0.17, 0.19\}$ P = 6/7 $C^c = \{0.10, 0.12, 0.19, 0.25\}$ P = 4/7

All Events: $A \cup B \cup \mathrm{C} = \{0.10, 0.12, 0.14, 0.15, 0.17, 0.19, 0.25\} = \Omega$, $\mathrm{P}(A \cup B \cup \mathrm{C}) = 6/7$

Use the ratios to calculate the frequentist probabilities for each event.

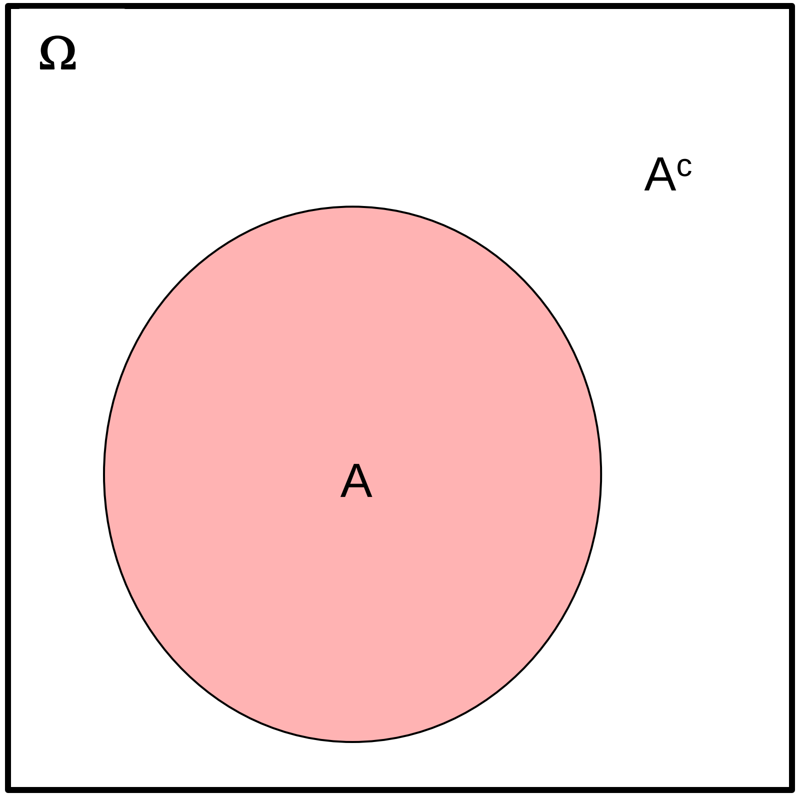
Probability Definitions
Probability Concepts

Now that we have defined probability notation and operators, let's return to Kolmogorov's probability axioms and complete them with these probability requirements.

Requirements for a valid probability:

- Bounded $0 \leq P(A) \leq 1$ probability between 0.0 and 1.0
- Closure $P(\Omega) = 1$ probability of any event is 1.0

  $P(A^c) + P(A) = 1$ probability of A or not A is 1.0
- Null Sets $P(\phi) = 0$ probability of nothing happens is zero



Venn Diagram – illustrating complementary events.

Probability Definitions
Probability Concepts

The Addition Rule:

$$P(A \cup B) = \mathrm{P(A)} + \mathrm{P(B)} - \mathrm{P}(A \cap B)$$

Must account for the intersection!

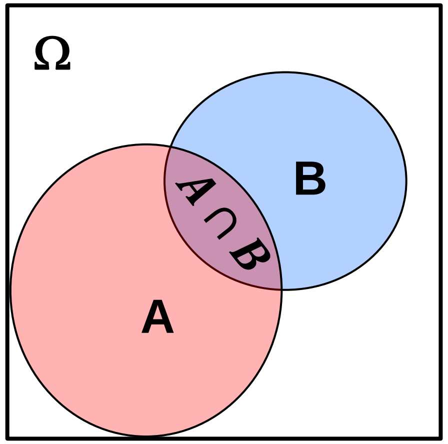
If mutually exclusive events:

$$A_i \cap A_j = \emptyset, \forall\, i \neq j$$

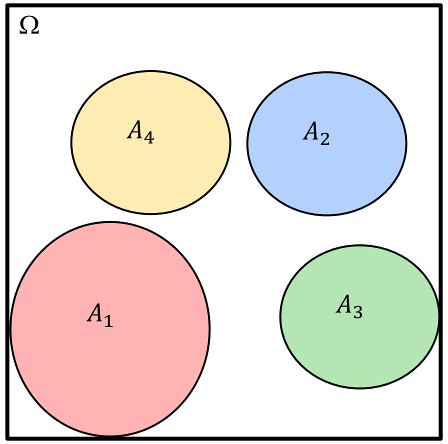
then,

$$\mathrm{P}\left(\bigcup_{i=1}^{k} \mathrm{A}_i\right) = \sum_{i=1}^{k} \mathrm{P}(\mathrm{A}_i)$$

no intersections to account for.



Venn Diagram – illustrating intersection.

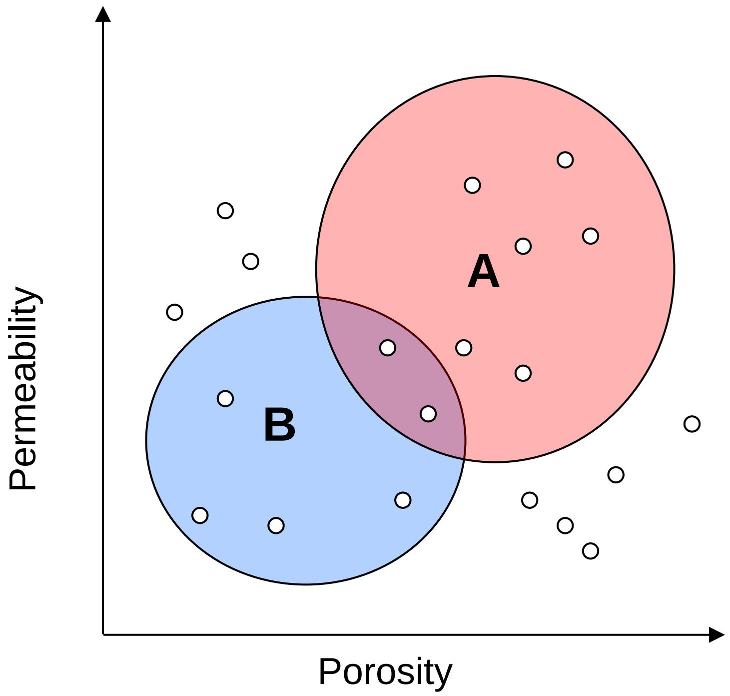


Venn Diagram – with mutually exclusive events.

Probability Definitions Addition Rule Example

Calculate the following probabilities for event A and B:

Note Event A: Sandstone and Event B: Shale

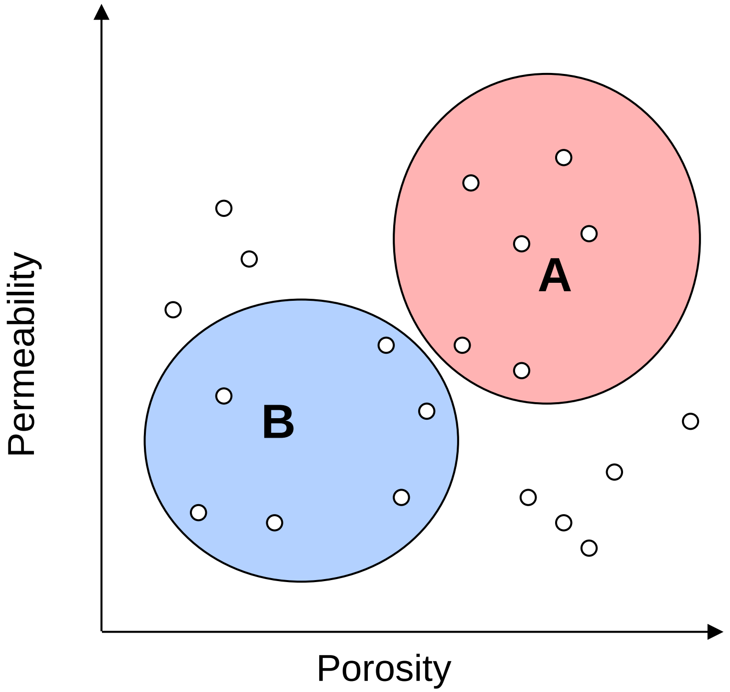


$P(A) =$

$P(B) =$

$P(A \cap B) =$

$P(A \cup B) =$



$P(A) =$

$P(B) =$

$P(A \cap B) =$

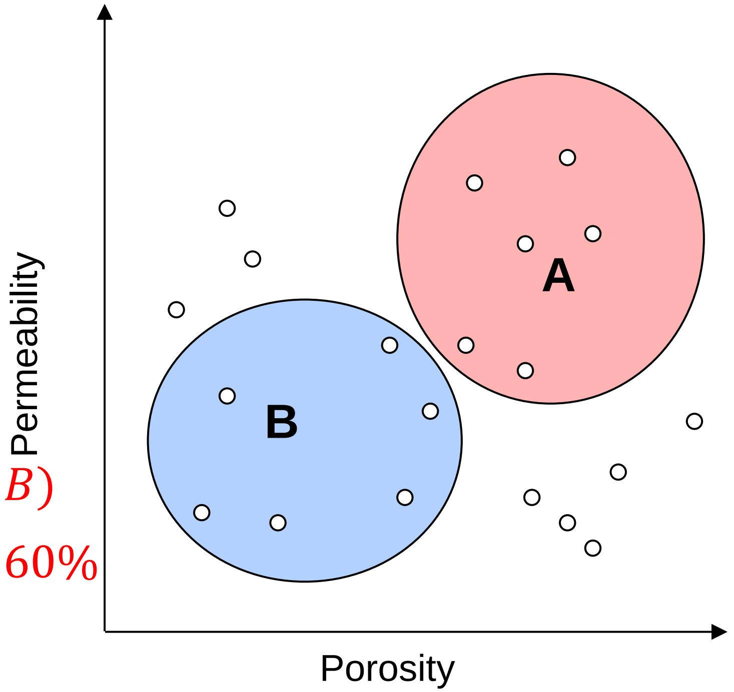
$P(A \cup B) =$

Hint, this is just counting the points!

Probability Definitions
Addition Rule Example

Calculate the following probabilities for event A and B:

Note Event A: Sandstone and Event B: Shale



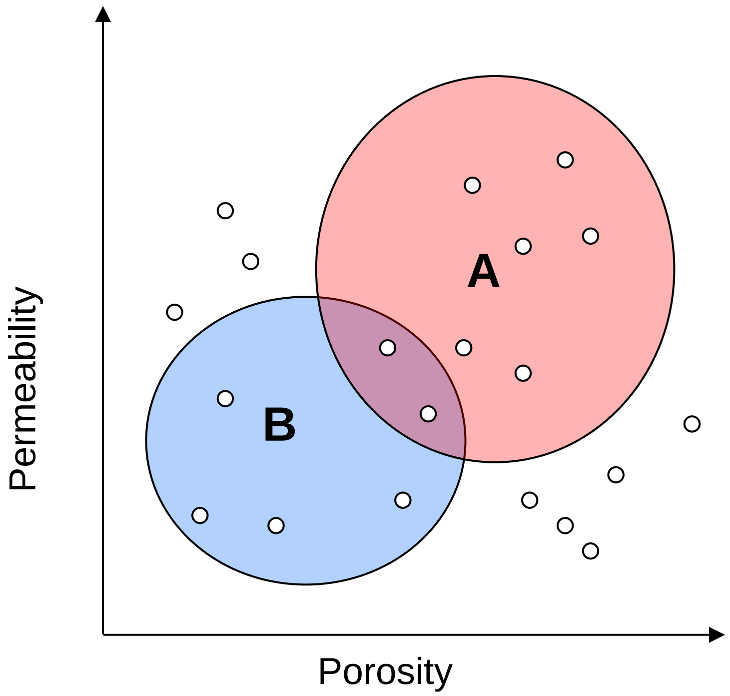
$$\mathrm{P}(A) = {}^{6}/_{20} = 30\%$$

$$\mathrm{P}(B) = {}^{6}/_{20} = 30\%$$

$$\mathrm{P}(A \cap B) = {}^{0}/_{20} = 0\%$$

$$\mathrm{P}(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$30\% + 30\% - 0\% = 60\%$$



$$\mathrm{P}(A) = {}^{8}/_{20} = 40\%$$

$$\mathrm{P}(B) = {}^{6}/_{20} = 30\%$$

$$\mathrm{P}(A \cap B) = {}^{2}/_{20} = 10\%$$

$$\mathrm{P}(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$40\% + 30\% - 10\% = 60\%$$

Hint, this is just counting the points!

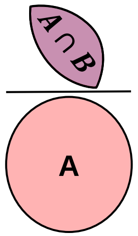
Probability Definitions
Conditional Probability

Probability of B given A occurred?

P(B | A) , read as $P(B \; given \; A)$

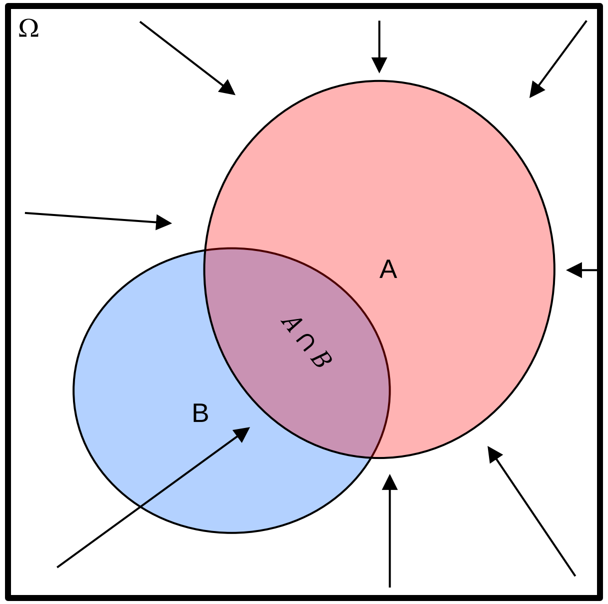
- What is the probability of B given A as already occurred?

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



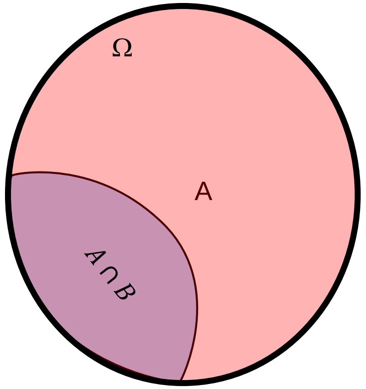
P(A ∩ B)

P(A)



Conceptually we shrink space of possible outcomes.

A occurred so we shrink our universe (**Ω**) to only event A.



Probability Definitions

Conditional, Marginal and Joint Probability

Now let's define three cases of probability and provide notation:

Marginal Probability: Probability of an event, irrespective of any other event

$$P(X), P(Y)$$

Conditional Probability: Probability of an event, given another event has already occurred

$$P(X \text{ } given \text{ } Y), P(Y \text{ } given \text{ } X)$$

$$P(X \mid Y), P(Y|X)$$

Joint Probability: Probability of multiple events occurring together.

$$P(X \text{ } and \text{ } Y), P(Y \text{ } and \text{ } X)$$

$$P(X \cap Y), P(Y \cap X)$$

$$P(X, Y), P(Y, X)$$

Probability Definitions
Conditional Probability

General Form for Conditional Probability

Recall:

$$P(C \mid B, A) = \frac{P(A \cap B \cap C)}{P(A \cap B)}$$

Substitute:

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

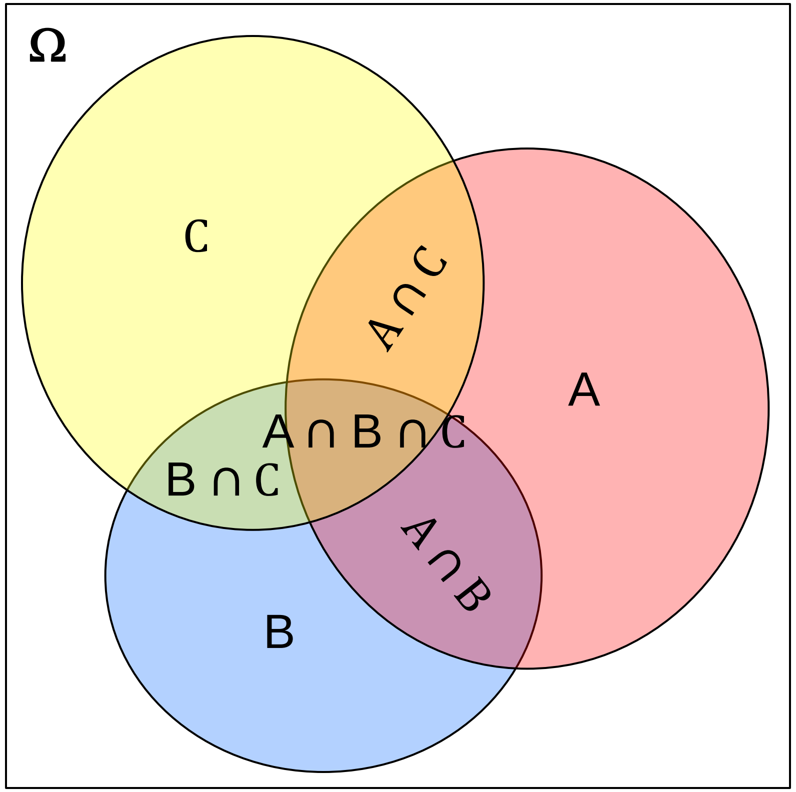
Now we have:

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

General Form, Recursion of Conditionals

$$P(A \cap B \cap C) = P(C \mid B, A)P(B|A)P(A)$$

$$P(A_1 \cap \cdots \cap A_n) = P(A_n|A_{n-1}, \ldots, A_1)P(A_{n-1}|A_{n-2}, \ldots, A_1) \ \ldots \ P(A_1)$$

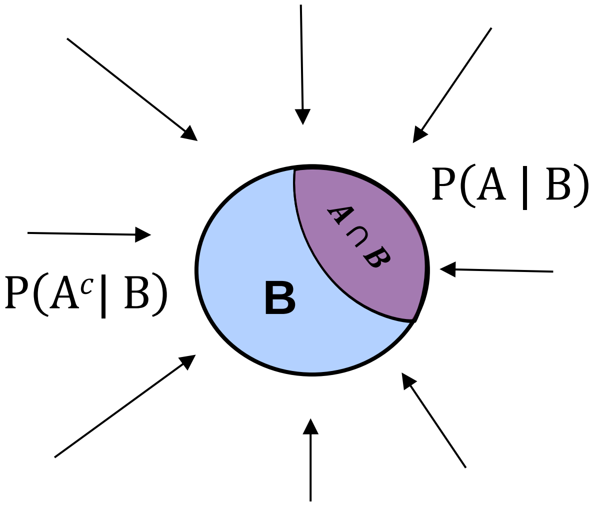


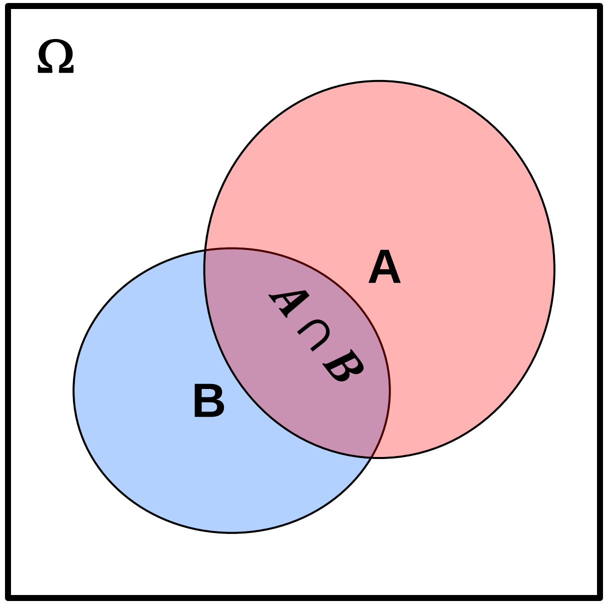
Venn Diagram – illustrating a more complicated 3 event probability with intersections labeled.

Probability Definitions
Conditional Probability

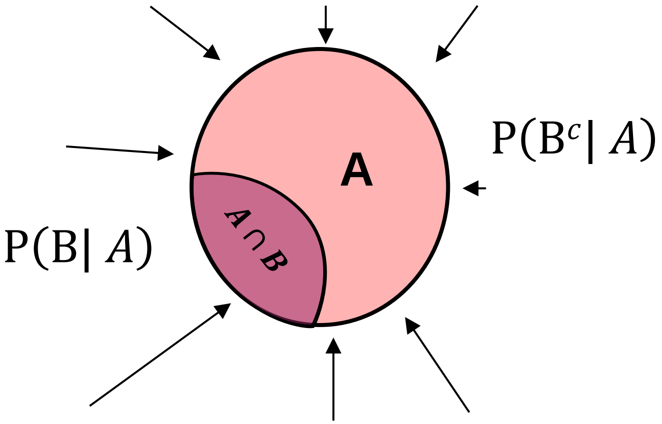
Closure with Conditional Probability

Closure with conditional probabilities:





$$P(A\mid B) + P(A^c\mid B) = 1$$



$$P(B\mid A) + P(B^c\mid A) = 1$$

Probability Fundamentals
Conditional Probability Examples

Recall:

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

For Case 1 calculate:

$P(A \mid B) =$

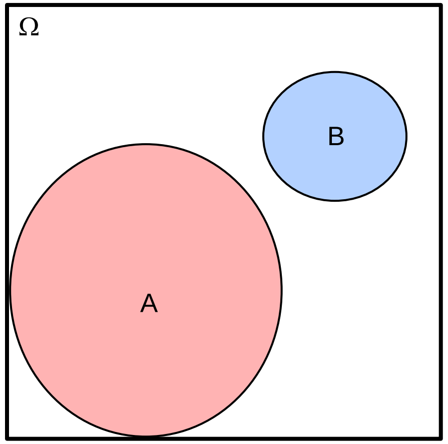
$P(B \mid A) =$

For Case 2 calculate:

$P(A \mid B) =$

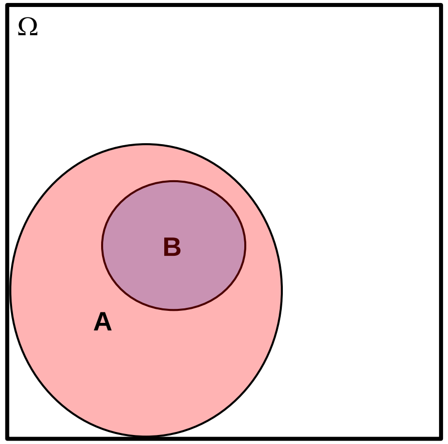
$P(B \mid A) =$

Case 1:



Venn Diagram – case 1.

Case 2:



Venn Diagram – case 2.

Probability Fundamentals
Conditional Probability Examples

Recall:

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

For Case 1 calculate:

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

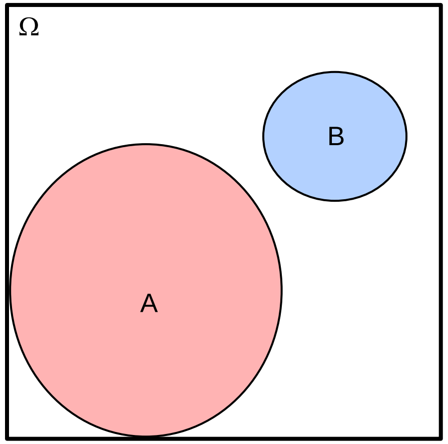
$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\emptyset}{P(A)} = \emptyset$$

For Case 2 calculate:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(B)} = 1, \quad \text{since } P(A \cap B) = P(B)$$

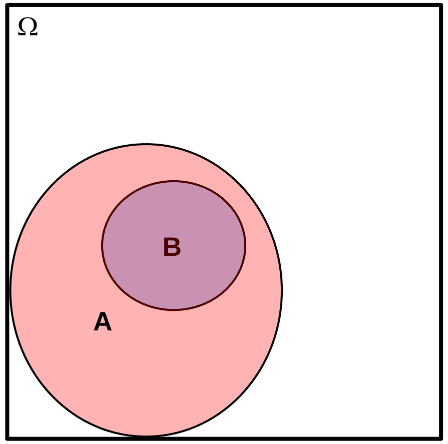
$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B)}{P(A)}, \quad \text{since } P(A \cap B) = P(B)$$

Case 1:



Venn Diagram – case 1.

Case 2:



Venn Diagram – case 2.

Probability Definitions
Conditional Probability Example

Question: Calculate the following probabilities for events A and B:

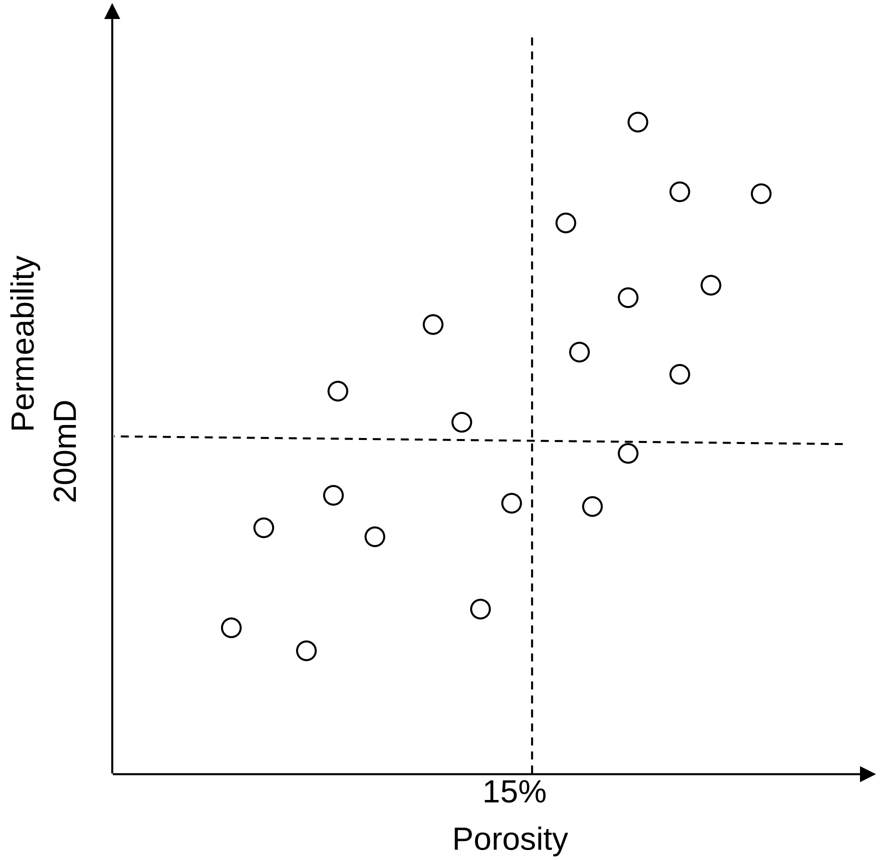
Event A: Porosity $> 15\%$
Event B: Permeability > 200 mD

For Case 1 calculate:

$\mathrm{P(A \mid B)} =$

$\mathrm{P(B \mid A)} =$

Bonus Question: How much information does event B tell you about event A and visa versa?



Hint, more point counting.

Probability Definitions
Conditional Probability Example

Question: Calculate the following probabilities for events A and B:

Event A: Porosity $> 15\%$
Event B: Permeability > 200 mD

For Case 1 calculate:

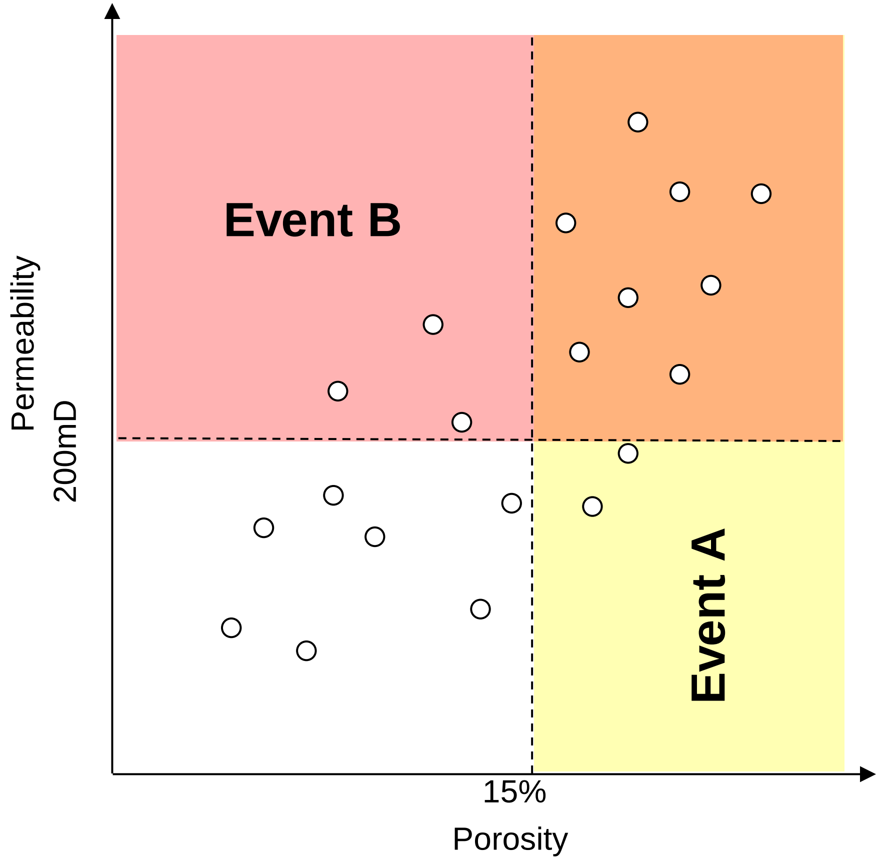
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{8/20}{11/20} = 8/11$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{8/20}{10/20} = 8/10$$

Bonus Question: How much information does event B tell you about event A and visa versa?

$P(A) = 10/20, P(A|B) = 8/11$ Probability from $50\% \rightarrow 73\%$
$P(B) = 11/20, P(B|A) = 8/10$ Probability from $55\% \rightarrow 80\%$

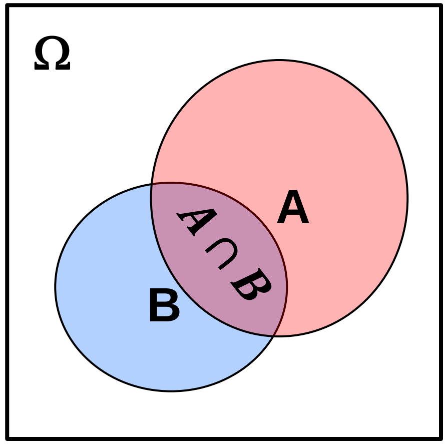
We cannot work with A and B independently; they provide information about each other.



Probability Definitions

Conditional, Marginal and Joint Probability

Let's put this all together now, schematic then exercise.



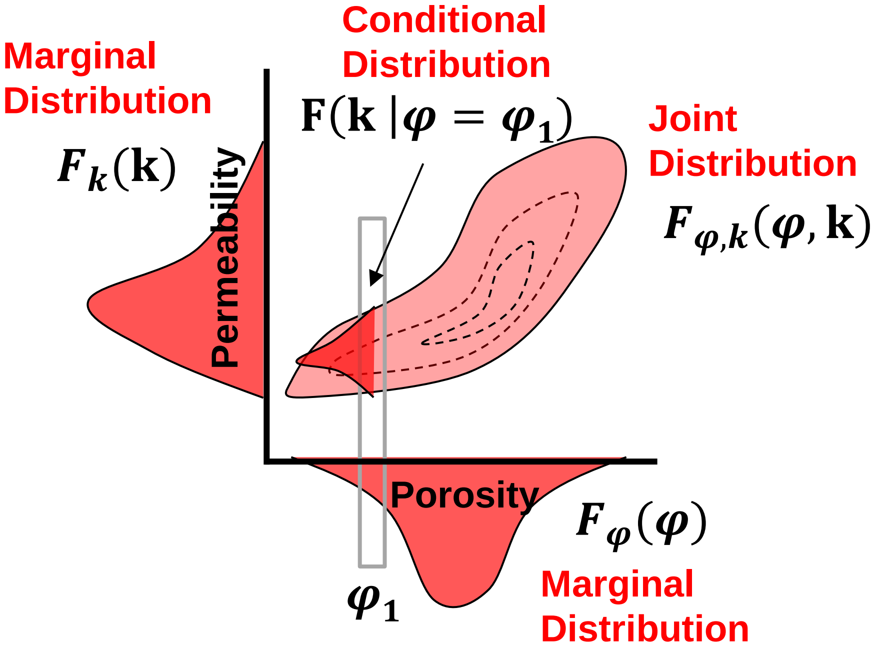
Probability of B given A occurred? $\mathrm{P}(\mathrm{B} \mid \mathrm{A})$

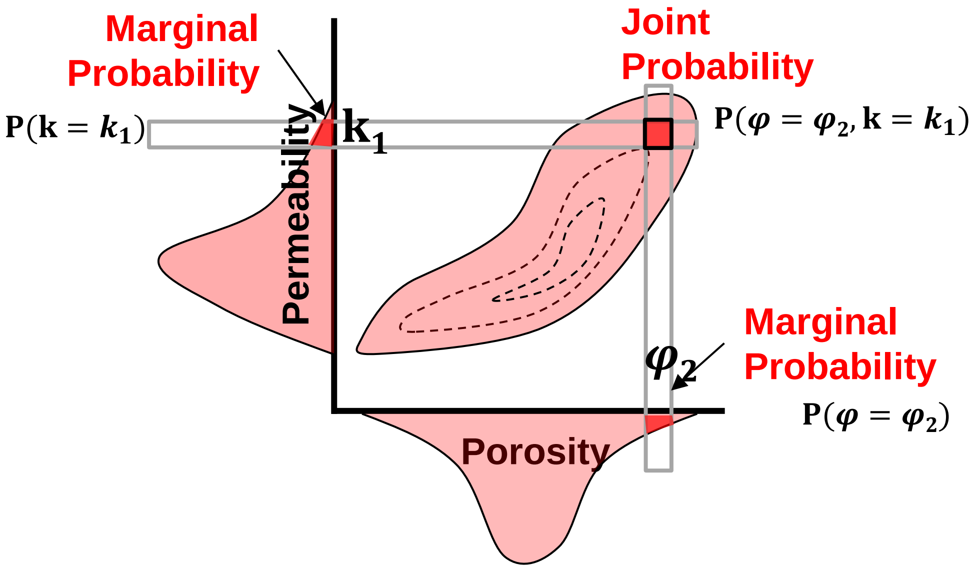
Conditional Probability

Joint Probability

$$\mathrm{P}(\mathrm{B} \mid \mathrm{A}) = \frac{P(\mathrm{A} \cap \mathrm{B})}{\mathrm{P}(\mathrm{A})} = \frac{P(\mathrm{A}\ and\ \mathrm{B})}{\mathrm{P}(\mathrm{A})}$$

Marginal Probability





Probability Definitions

Conditional, Marginal and Joint Probability

Calculating joint, marginal and conditional probability from a table of frequencies.

Continuous / Function

Joint Distribution:

$$f_{XY}(x,y)$$

Marginal Distribution:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy$$

Conditional Distribution:

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

Discrete Bins / Samples

Joint Probability:

$$\mathrm{P}(\mathrm{x}, y) = \frac{n(x,y)}{n(\Omega)}$$

Marginal Probability:

$$\mathrm{P}(\mathrm{x}) = \frac{\sum_{i=1}^{bins} n(x,y_i)}{n(\Omega)} = \frac{n(x)}{n(\Omega)}$$

Conditional Probability:

$$\mathrm{P}(\mathrm{x}|y) = \frac{n(x,y)}{n(\mathrm{y})}$$

Table of Frequencies $n(\varphi, v_{sh})$

Porosity, φ (%)	10%	30%	50%	70%	90%
25%	1	1	0	0	0
20%	2	3	2	0	0
15%	1	2	2	1	0
10%	0	0	2	3	2
5%	0	0	1	1	1

Fraction Shale, v_{sh} (%)

Samples (○) with Porosity and Fraction of Shale
Scatter plot binned with frequencies, n = 25.

Probability Definitions
Conditional, Marginal and Joint Probability

Joint Probability:

$$\boxed{P(\mathrm{x}, y) = \frac{n(x, y)}{n(\Omega)}}$$

Table of Joint Probabilities $P(\varphi, v_{sh})$

Porosity, φ (%)	10%	30%	50%	70%	90%
25%	4%	4%	0	0	0
20%	8%	12%	8%	0	0
15%	4%	8%	8%	4%	0
10%	0	0	8%	12%	8%
5%	0	0	4%	4%	4%

Fraction Shale, v_{sh} (%)

Porosity and Fraction of Shale binned joint probabilities, n = 25.

Marginal Probability:

$$\mathrm{P(x)} = \frac{\sum_{i=1}^{ybins} n(x, y_i)}{n(\Omega)} = \frac{n(x\ )}{n(\Omega)}$$

Marginalization
Integrate over y to remove it

Conditional Probability:

$$\mathrm{P(x}|y) = \frac{n(x, y)}{n(\mathrm{y})}$$

Standardization
Divide by $n(\mathrm{y})$ to shrink the sample space to occurrences of y

Probability Definitions

Conditional, Marginal and Joint Probability

Given these joint probabilities calculate the:

Marginal Probabilities:

v_{sh}	10%	30%	50%	70%	90%
$Prob(v_{sh}) =$					

φ	5%	10%	15%	20%	25%
$Prob(\varphi) =$					

Conditional Probabilities:

v_{sh}	10%	30%	50%	70%	90%
$Prob(v_{sh} \mid \varphi = 15\%) =$					

Table of Joint Probabilities

Porosity, φ (%) \ Fraction Shale, v_{sh} (%)	10%	30%	50%	70%	90%
25%	4%	4%	0	0	0
20%	8%	12%	8%	0	0
15%	4%	8%	8%	4%	0
10%	0	0	8%	12%	8%
5%	0	0	4%	4%	4%

Fraction Shale, v_{sh} (%)

Porosity and Fraction of Shale
binned joint probabilities

Probability Definitions

Conditional, Marginal and Joint Probability

Given these joint probabilities calculate the:

Marginal Probabilities:

$Prob(v_{sh}) =$

v_{sh}	10%	30%	50%	70%	90%
	16%	24%	28%	20%	12%

$Prob(\varphi) =$

φ	5%	10%	15%	20%	25%
	12%	28%	24%	28%	8%

Conditional Probabilities:

$Prob(v_{sh} | \varphi = 15\%) =$

v_{sh}	10%	30%	50%	70%	90%
	1/6	1/3	1/3	1/6	0

Table of Joint Probabilities

Porosity, φ (%) \ Fraction Shale, v_{sh} (%)	10%	30%	50%	70%	90%
25%	4%	4%	0	0	0
20%	8%	12%	8%	0	0
15%	4%	8%	8%	4%	0
10%	0	0	8%	12%	8%
5%	0	0	4%	4%	4%

Porosity and Fraction of Shale
binned joint probabilities

Frequentist Hands-on
Joint, Conditional and Marginal

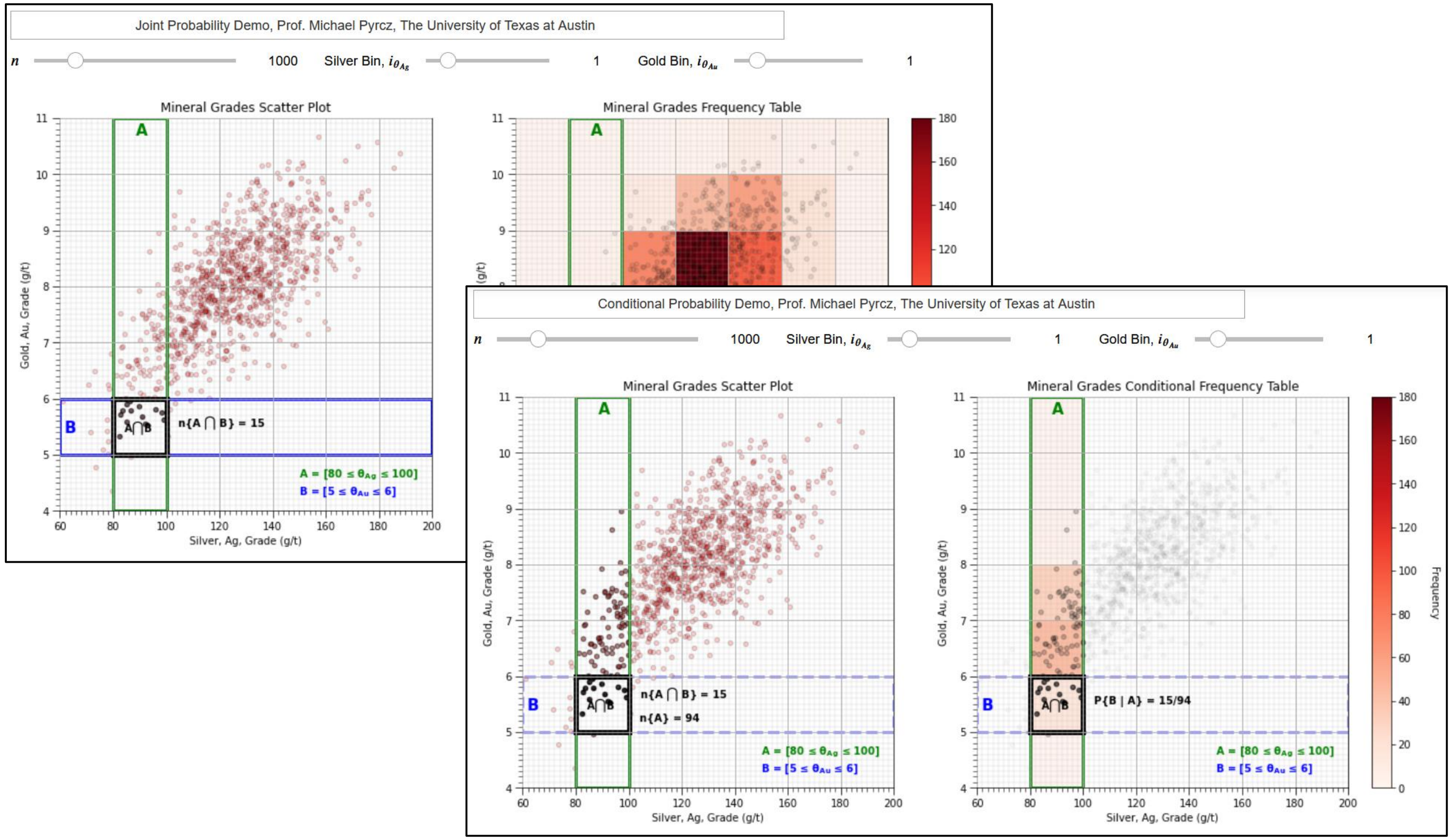
Joint, Conditional and Marginal Probability and Distributions:

Things to try:

1. Change the A and B bins and observe the joint and conditional probabilities.

2. Change the number of data and observe the impact on the observed probabilities.

- how many data samples are needed for stable statistics?



Interactive joint and conditional probabilities with a mining dataset, the file is Interactive_MarginalJointConditional.ipynb.

Frequentist Hands-on
Joint, Conditional and Marginal

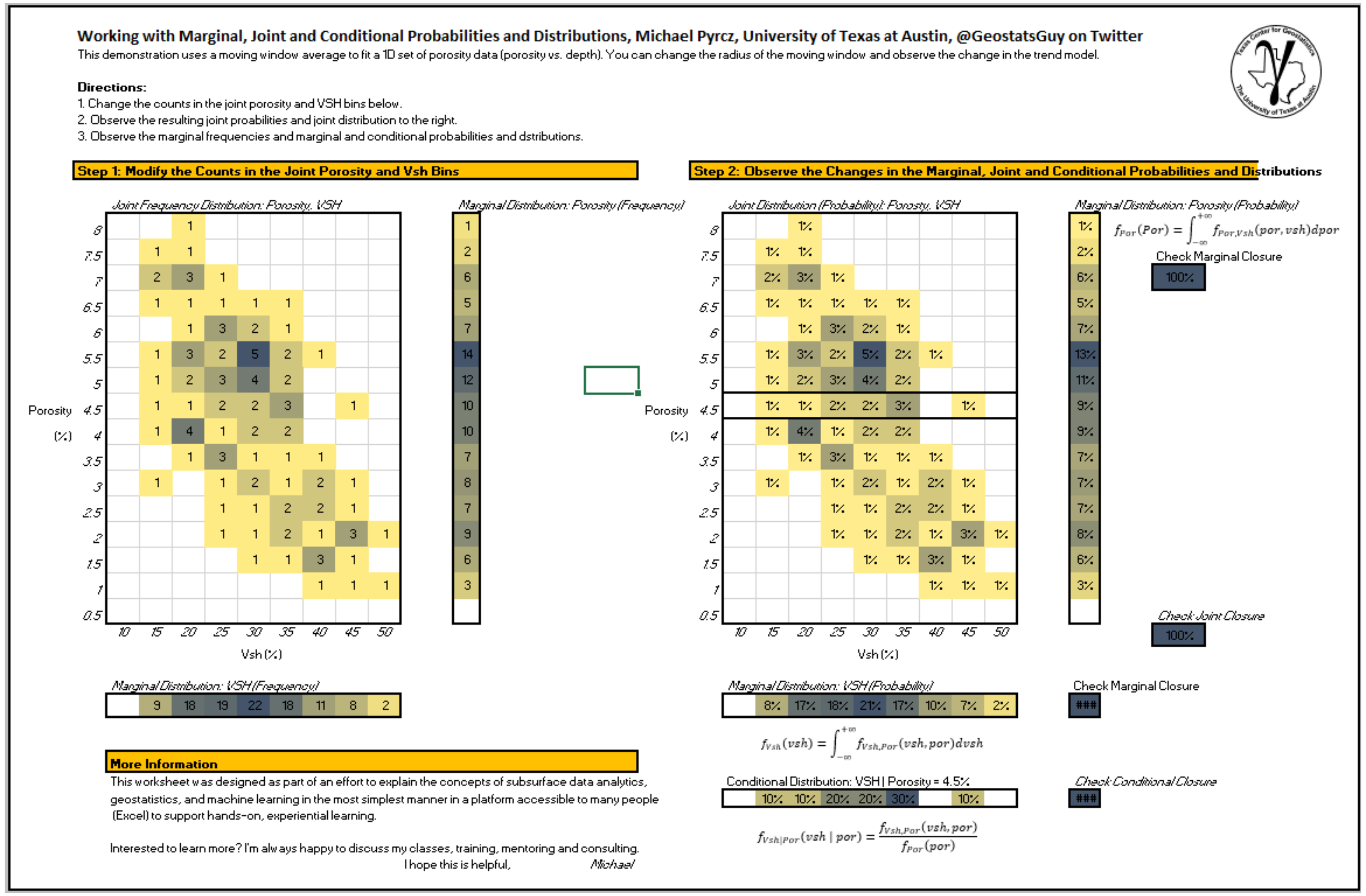
Joint, Conditional and Marginal Probability and Distributions:

Things to try:

1. Add a spike (large frequency) in an outlier location.

2. Systematically increase the frequencies for the low porosity, high vsh region.

Observe the impact on joint, marginal and conditional probabilities and distributions.



The file is Marginal_Joint_Conditional.xlsx The file is at: https://git.io/fhA9X.

Probability Definitions
Multiplication Rule

The Multiplication Rule:

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

We rearranged the definition of conditional probability.

If events A and B are independent:

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

knowing something about A does nothing to help predict B. Then by substitution:

$$\mathrm{P}(A \cap B) = \mathrm{P}(\mathrm{B})\,\mathrm{P}(A)$$

The general form given independence for all cases, $i = 1, \ldots, k$:

$$\mathrm{P}\left(\cap_{\boldsymbol{i=1}}^{\boldsymbol{k}} \mathrm{A_i}\right) = \prod_{\boldsymbol{i=1}}^{\boldsymbol{k}} \mathrm{P}(\mathrm{A_i})$$

e.g., $P(\mathrm{A_1} \cap \mathrm{A_2} \cap \mathrm{A_3}) = \mathrm{P}(\mathrm{A_1})\,P(\mathrm{A_2})\,P(\mathrm{A_3})$

Probability Definitions
Multiplication Rule Example

Given independence between fluid type and porosity:

Event A = Oil
Event B = Porosity > 13%

Given: $\mathrm{P(A)} = 30\%$ and $\mathrm{P(B)} = 50\%$

What is the P($A \cap B$)?

Given independence between fluid type, porosity and saturation:

Event A = Oil
Event B = Porosity > 13%
Event C = S_{oil} > 40%

Given: $\mathrm{P(A)} = 30\%, \mathrm{P(B)} = 50\%, \mathrm{P(C)} = 10\%$

What is the P($A \cap B \cap C$)?

Probability Definitions
Multiplication Rule Example

Given independence between fluid type and porosity:

Event A = Oil

Event B = Porosity > 13%

Given: $\mathrm{P}(\mathrm{A}) = 30\%$, and $\mathrm{P}(\mathrm{B}) = 50\%$

What is the $\mathrm{P}(A \cap B)$? = 30% x 50% = 15%

Given independence between fluid type, porosity and saturation:

Event A = Oil

Event B = Porosity > 13%

Event C = S_{oil} > 40%

Given: $\mathrm{P}(\mathrm{A}) = 30\%$, $\mathrm{P}(\mathrm{B}) = 50\%$, and $\mathrm{P}(\mathrm{C}) = 10\%$

What is the $\mathrm{P}(A \cap B \cap C)$? = 30% x 50% x 10% = 1.5%

Probability Definitions
Evaluating Independence

Events A and B are independent if and only if:

$$P(A \cap B) = P(B) \cdot P(A)$$

or

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(\boldsymbol{B})$$

Recall the General Form:

Events $A_1, A_2, \ldots, A_n$ are independent if: $P\left(\cap_{i=1}^{k} A_i\right) = \prod_{i=1}^{k} P(A_i)$

Then We Can Derive:

$$\mathbf{P}(\boldsymbol{A}|\boldsymbol{B}) = \frac{\boldsymbol{P}(\boldsymbol{A},\boldsymbol{B})}{\boldsymbol{P}(\boldsymbol{B})} = \frac{\boldsymbol{P}(\boldsymbol{A})\cdot\cancel{\boldsymbol{P}(\boldsymbol{B})}}{\cancel{\boldsymbol{P}(\boldsymbol{B})}} = \mathbf{P}(\boldsymbol{A})$$

Probability Definitions
Evaluating Independence Example

Example: Facies Fm (mudstone), Sl (laminated sandstone) and Sm (massive sandstone) in 5 wells:

Position	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8	Well 9	Well 10
Top	Sm	Sl	Sl	Fm	Fm	Fm	Sl	Sl	Fm	Fm
Middle	Fm	Fm	Fm	Fm	Sl	Sl	Fm	Sl	Sl	Sl
Bottom	Sl	Sl	Sl	Sm	Sm	Sm	Sm	Sm	Sm	Sl

Event A_1 = middle facies if Fm (mudstone)
Event A_2 = bottom facies is Sm (massive sandstone)

Question: are events A_1 and A_2 independent?

Probability Definitions
Evaluating Independence Example

Example: Facies Fm (mudstone), Sl (laminated sandstone) and Sm (massive sandstone) in 5 wells:

Position	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8	Well 9	Well 10
Top	Sm	Sl	Sl	Fm	Fm	Fm	Sl	Sl	Fm	Fm
Middle	Fm	Fm	Fm	Fm	Sl	Sl	Fm	Sl	Sl	Sl
Bottom	Sl	Sl	Sl	Sm	Sm	Sm	Sm	Sm	Sm	Sl

Event A_1 = middle facies if Fm (mudstone)
Event A_2 = bottom facies is Sm (massive sandstone)

Question: are events A_1 and A_2 independent?

Recall Definition of Independence

$$P(\boldsymbol{A_1} \cap \boldsymbol{A_2}) = P(\boldsymbol{A_1}) \cdot P(\boldsymbol{A_2})$$

$$P(\boldsymbol{A_1} | \boldsymbol{A_2}) = P(\boldsymbol{A_1}) \qquad P(\boldsymbol{A_2} | \boldsymbol{A_1}) = P(\boldsymbol{A_2})$$

Not Independent if One Does not Hold

Pick one condition for independence and check it, $P(\boldsymbol{A_1} \cap \boldsymbol{A_2}) = P(\boldsymbol{A_1}) \cdot P(\boldsymbol{A_2})$

$$P(\boldsymbol{A_1}) = {}^{5}/_{10} = 0.50, P(\boldsymbol{A_2}) = {}^{6}/_{10} = 0.60, P(\boldsymbol{A_1} \cap \boldsymbol{A_2}) = {}^{2}/_{10} = 0.20$$

$$P(\boldsymbol{A_1}) \cdot P(\boldsymbol{A_2}) = \mathbf{0.50 \cdot 0.60 = 0.30} \neq P(\boldsymbol{A_1} \cap \boldsymbol{A_2}) = {}^{2}/_{10} = 0.20$$

A_1 and A_2 are not independent

Probability Definitions
Bayesian Statistics

Recall the Multiplication Rule:

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

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

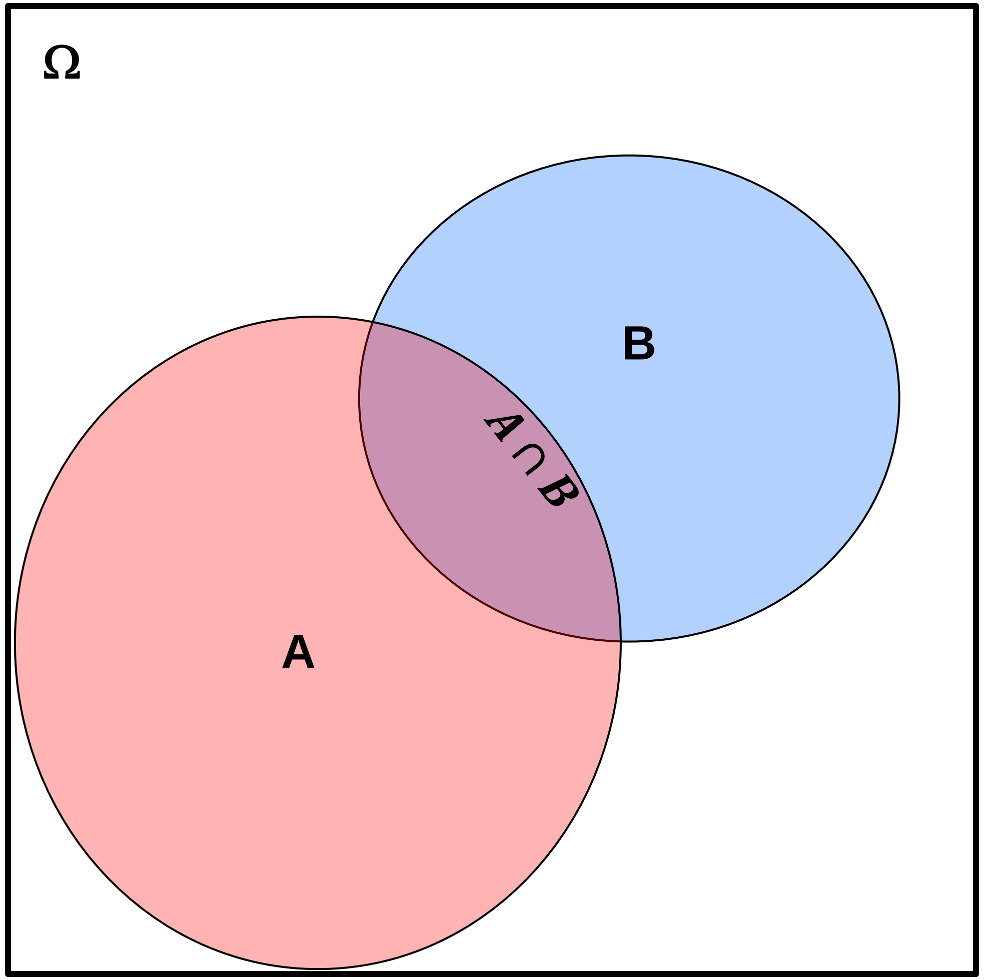
It follows that:

$$P(B \cap A) = P(A \cap B)$$

Therefore, we combine two product rules, substitute:

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

We get Bayes' Theorem!



Venn Diagram – illustrating intersection.

Bayes Theorem Details

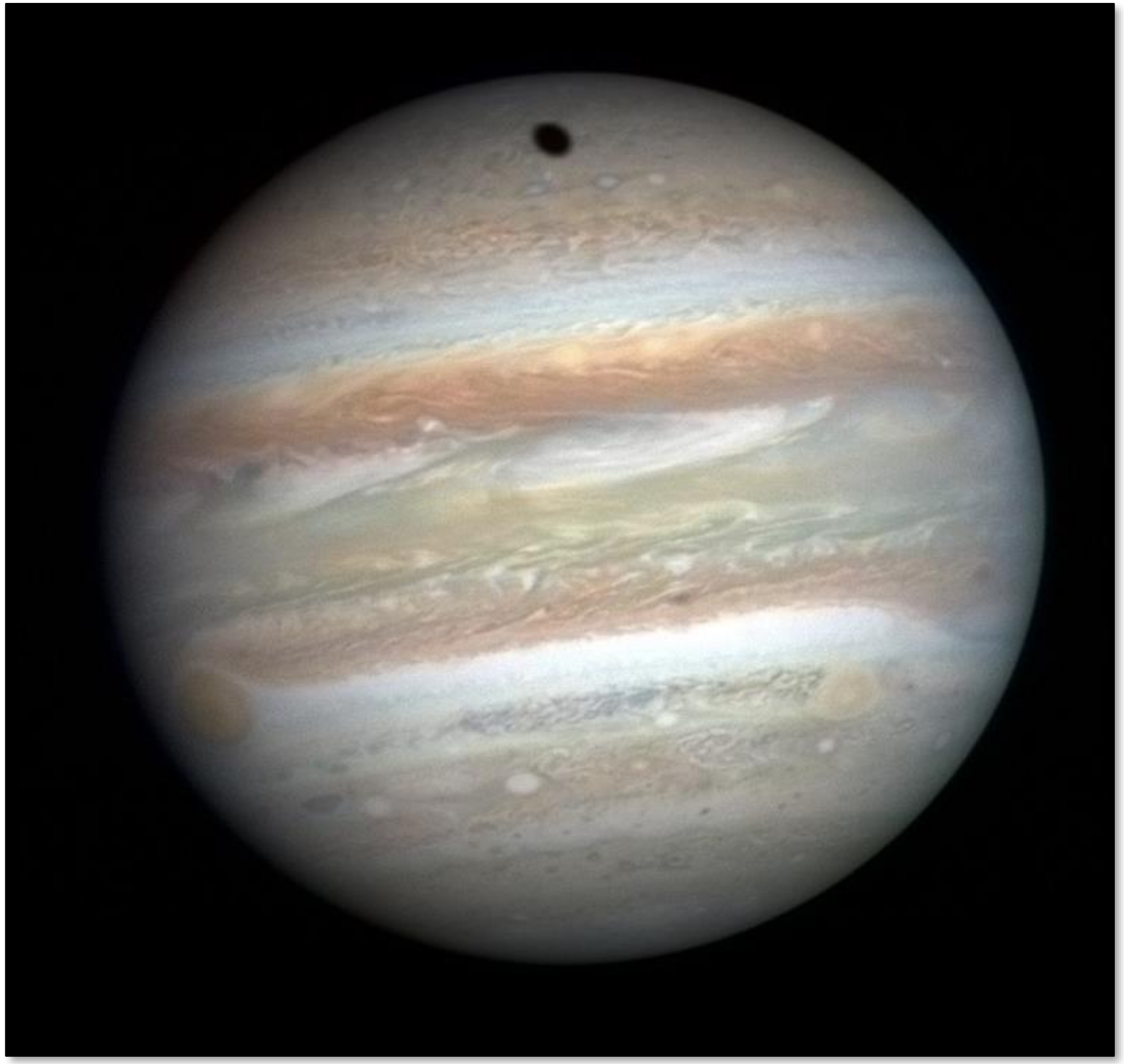
Bayesian Statistical Approaches:

- probabilities based on a degree of belief (expert experience) in an event
- updated as new information is available
- solve probability problems that we cannot use simple frequencies

From Sivia (1996), What is the mass of Jupiter?

Frequentist: measure the mass of enough Jupiter-like planets from multiple solar systems.

Bayesian: form a prior probability and update with any available information.



Jupyter image from New Horizons Long Range Imager (LORRI), taken at 57 million km on January 2007 (https://en.wikipedia.org/wiki/Jupiter#/media/File:Jupiter_New_Horizons.jpg)

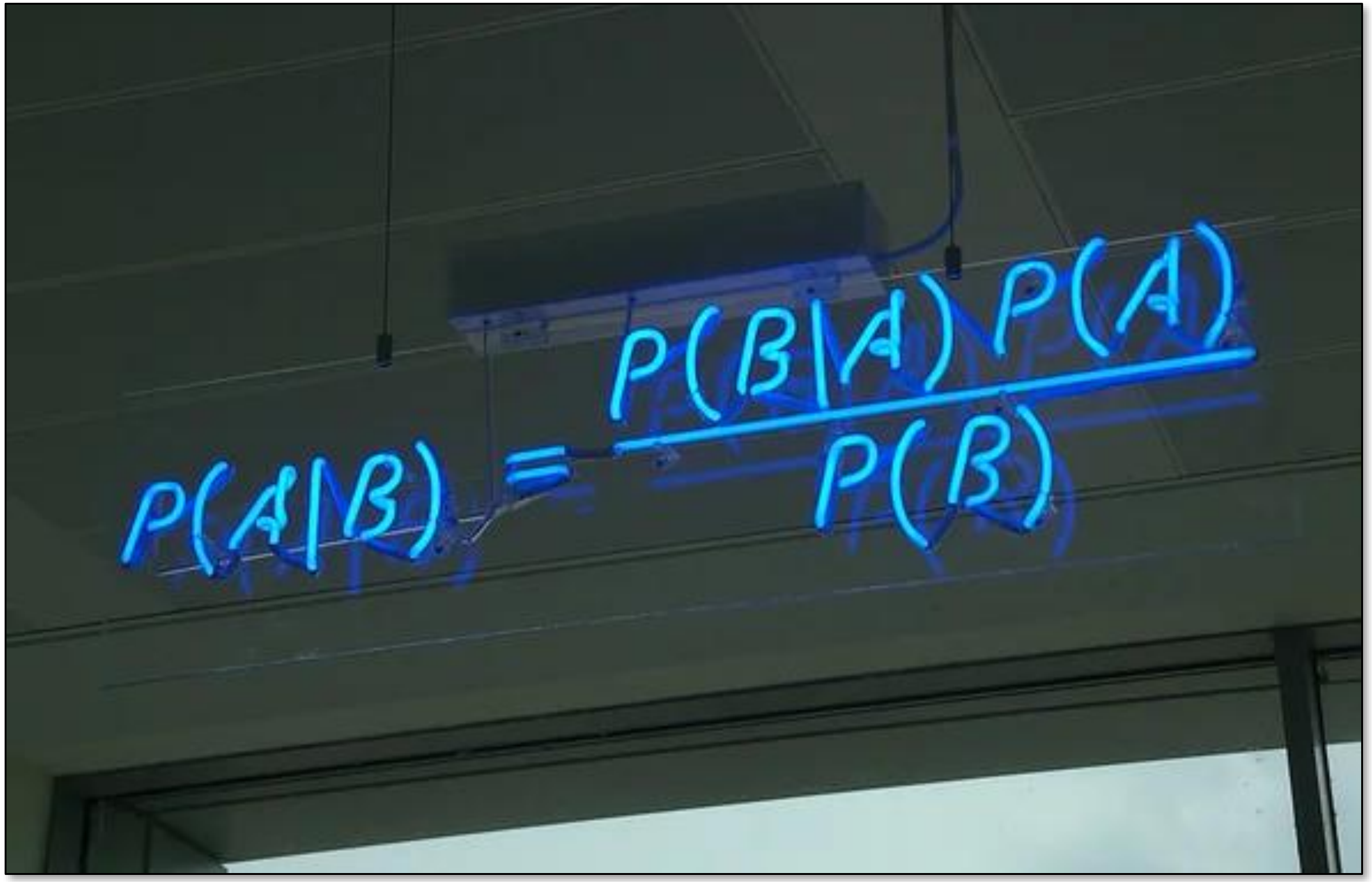
What is Probability? Bayesian Approach

We got here:

$$P(A|B)\,P(B) \;=\; P(B|A)\,P(A)$$

With a simple operation we get to the common, popular form:

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



The Bayes Appreciation Society Bayes' theorem in neon lights at the office of a software company in Cambridge, England (Matt Buck, from https://blogs.scientificamerican.com/roots-of-unity/the-bayes-appreciation-society/).

Comments on Bayesian Statistics

Bayesian Statistical Approaches:

- probabilities based on:
    - state of knowledge
    - degree of belief in an event

- utilize an assessment prior to data collection
- updated as new information is available
- solve probability problems that we cannot use simple frequencies

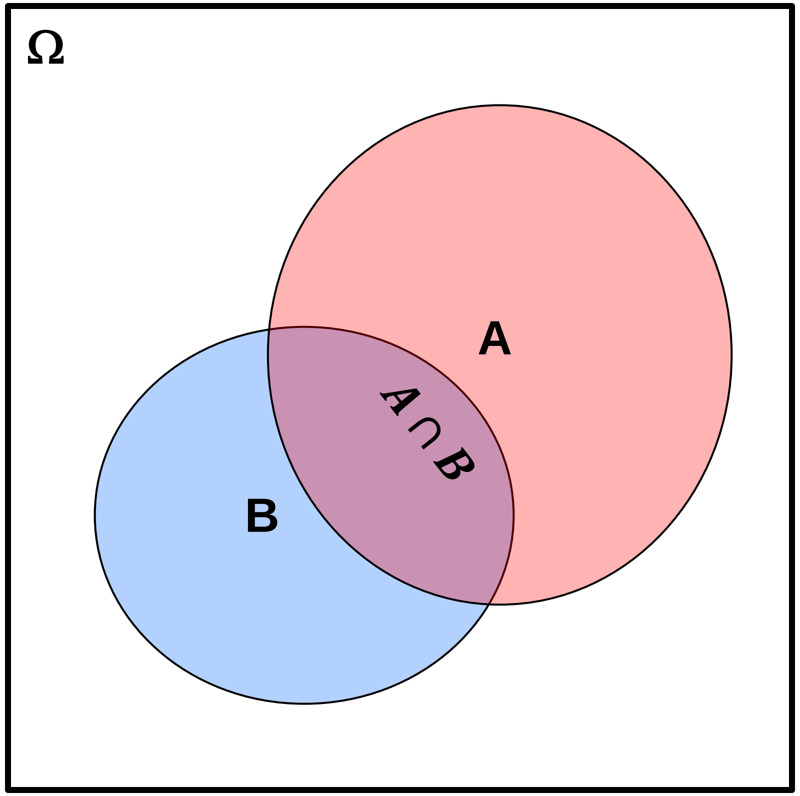
Advanced Concept on Uncertainty Modeling:

- **Bayesian credibility intervals** provide a more intuitive measure of uncertainty than frequentist confidence intervals, more later…

Bayes' Theorem Details

Bayes' Theorem Observations:

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



Venn Diagram – illustration of events and relations to each other.

Observations:

1. We can get $P(A \mid B)$ from $P(B \mid A)$, as you will see this often comes in handy.

2. Each term is known as:

$$\text{Posterior} = \frac{\text{Likelihood x Prior}}{\text{Evidence}}$$

3. Prior should have no information from likelihood.

4. Evidence term is usually just a standardization to ensure closure.

Bayes' Theorem Details

Bayes' Theorem Observations:

Observations:

5. If the prior distribution is naïve, no information before the data is collected, then $P(A) = P(A^c)$.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)\,P(A^c)}$$

Substitute

$P(A^c) = P(A)$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)\,P(A)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{[P(B|A) + P(B|A^c)]P(A)} = \frac{P(B|A)\cancel{P(A)}}{[P(B|A) + P(B|A^c)]\cancel{P(A)}}$$

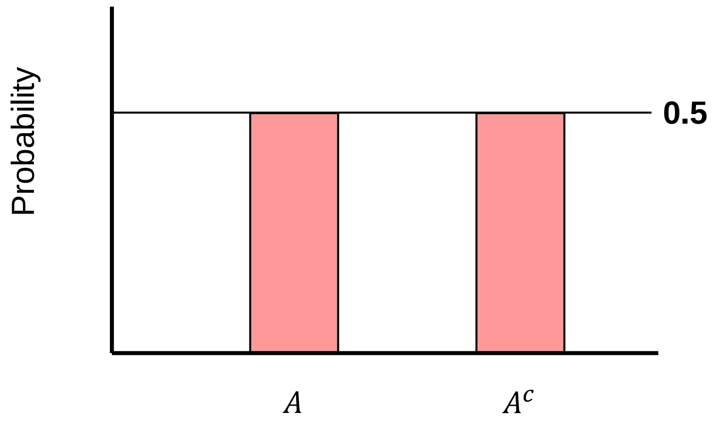
A naïve prior cancels out!

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

Evidence just ensures closure

Naïve = no information.

Naïve prior



Naïve prior, before data we have maximum uncertainty between the outcomes, all equally likely.

Posterior is equal to the likelihood.

Bayes' Theorem Details

Bayes' Theorem Observations:

Observations:

6. If the likelihood is naïve, the data provides no information, then $P(B|A) = c$. **Likelihood is constant.**

Substitute

$P(B|A) = c$

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

Calculate Evidence by Marginalization

$$p(x) = \int p(B|A)p(A)dA = \int c \cdot p(A)dA = c\int p(A)dA = c$$

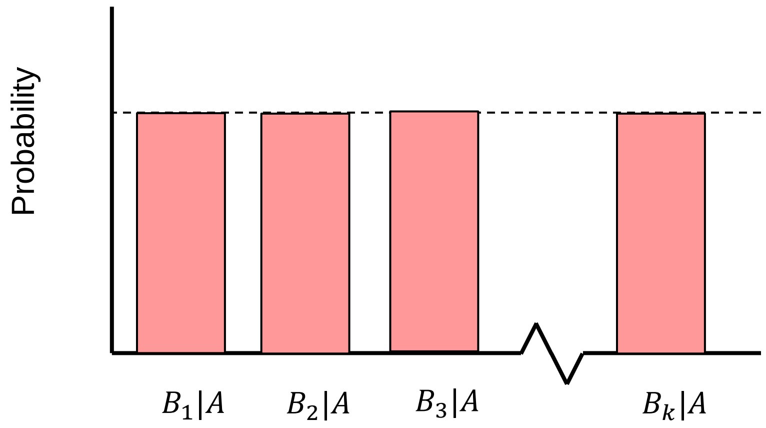
From closure

$$\int \boldsymbol{p(A)dA = 1}$$

Substitude into Bayes Theorem

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

Naïve likelihood



Naïve likelihood, the new data has no information about A, all conditionals are equally likely.

Posterior is equal to the prior.

Bayes' Theorem
Common Approach

Bayes' Theorem:

A common example of Bayes' Theorem for modeling the subsurface.

Let's change the labels to communicate model updating with a new data source:

Posterior **Likelihood** **Prior**

$$P(\text{Model} \mid \text{New Data}) = \frac{P(\text{New Data} \mid \text{Model})\, P(\text{Model})}{P(\text{New Data})}$$

Evidence

Bayes' Theorem Alternative Form

Bayes' Theorem:

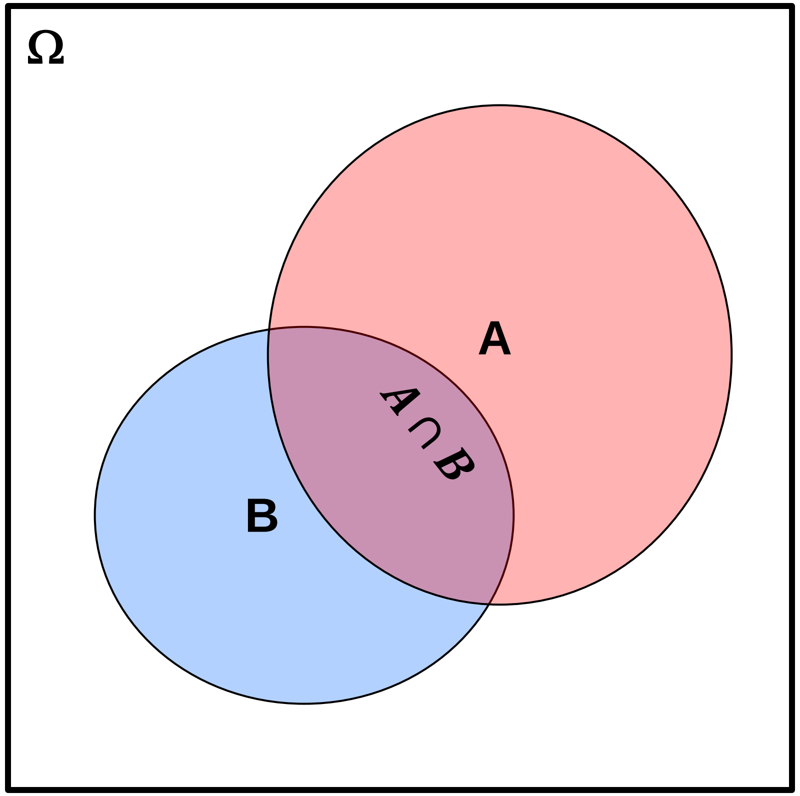
Alternative form, symmetry:

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

Alternative form to calculate evidence term:

Given: $\mathrm{P(A)} = \underbrace{P(A|B)P(B)}_{P(A\ and\ B)} + \underbrace{P(A|B^c)\,P(B^c)}_{P(A\ and\ B^c)}$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)\,P(B^c)}$$



Venn Diagram – illustration of events and relations to each other.

Since evidence term is often not readily available, we derive it by probability summation (recall, *marginalization*) over all possible outcomes, $(A, B)$and (A, B^c).

Applications of Bayes' Theorem

Bayesian approaches allow you to solve statistical problems that are not possible otherwise. Consider:

Something is Happening	Looks like it's happening
You have a disease	You test positive for the disease
There is fault compartmentalization	Geologist says there's a fault
Low permeability of a sample	The laboratory measure is low
A valve will fail	X-ray test is positive
You drill a dry well	Seismic AVO response looks poor

In all these cases you need to calculate:

$$P\left(\text{Something is Happening} \mid \text{Looks like it's happening}\right) = \frac{P\left(\text{Looks like it's happening} \mid \text{Something is Happening}\right) P\left(\text{Something is Happening}\right)}{P\left(\text{Looks like it's happening}\right)}$$

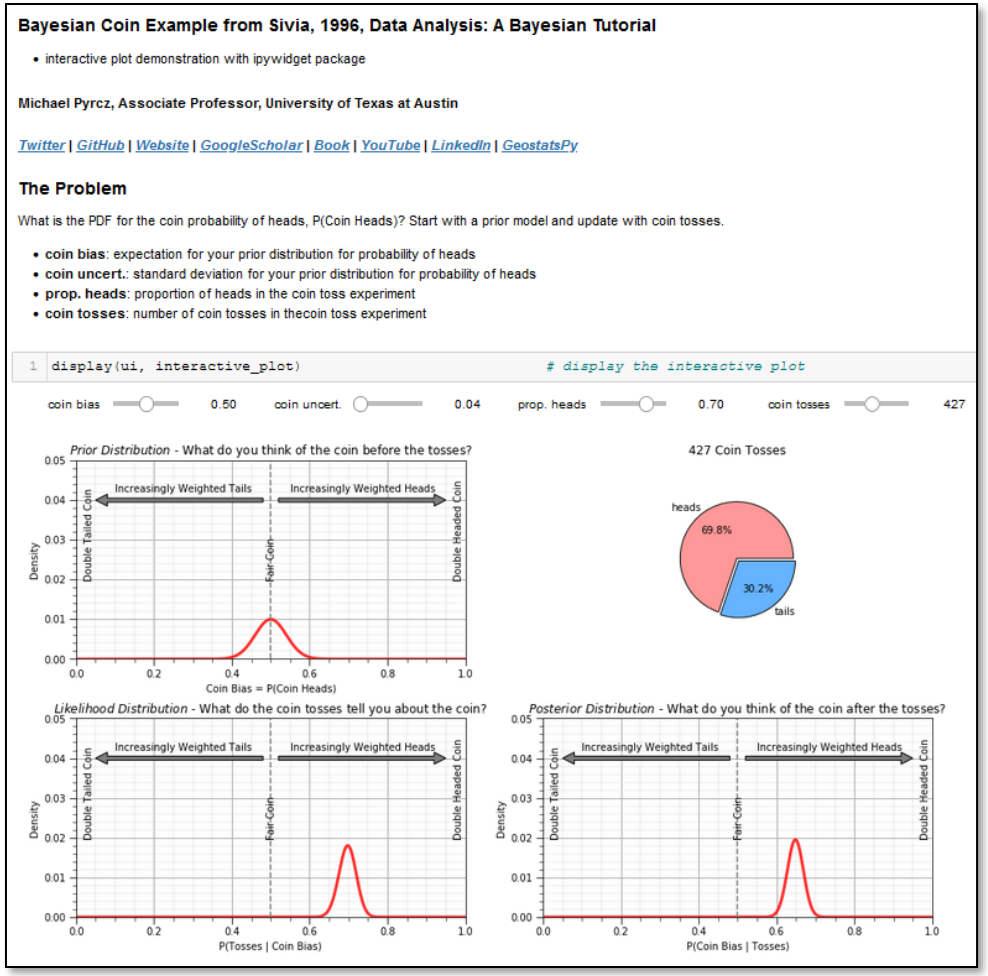
Bayesian Hands-on
The Sivia Coin Example in Python

Is Dr. Pyrcz's coin a fair coin?

Jupyter Notebook Python Demonstration

Things to try:

1. Try a naïve prior, I know nothing about Dr. Pyrcz's coin.
2. Try of very specific prior, I'm sure Dr. Pyrcz's coin is fair.
3. Try few and many coin tosses.
4. Contradiction between prior and likelihood.



The file is Interactive_Sivia_Coin_Toss.ipynb. An Excel version is available as Bayesian_Demo.xlsx.

Applications of Bayes' Theorem

Bayesian approaches allow you to solve statistical problems that are not possible otherwise. Consider:

True Positive Probability

Correct Detection Rate x Occurrence Rate

$$P\left(\begin{matrix}\text{Something is}\\ \text{Happening}\end{matrix} \;\middle|\; \begin{matrix}\text{Looks like}\\ \text{it's happening}\end{matrix}\right) = \frac{P\left(\begin{matrix}\text{Looks like}\\ \text{it's happening}\end{matrix} \;\middle|\; \begin{matrix}\text{Something is}\\ \text{Happening}\end{matrix}\right) P\left(\begin{matrix}\text{Something is}\\ \text{Happening}\end{matrix}\right)}{P\left(\begin{matrix}\text{Looks like}\\ \text{it's happening}\end{matrix}\right)}$$

All Detection Probability (included true and false positives)

Often these terms are much easier to collect:

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

Let's try this out next.

Bayesian Example #1

Example: Prior information at a site suggests a deepwater channel reservoir exists at a given location with probability of 60%. We consider further investigation with a 3D seismic survey.

3D seismic survey will indicate a channelized reservoir:
- is present with 90% probability if it really is present
- is not present with a probability 70% if it really is not

***A*=The deepwater channel is present**
***B* =Seismic shows a deepwater channel**
***A^c^* =The deepwater channel not present**
***B^c^* =Seismic does not show a deepwater channel**

$$P(A) = 0.6$$
$$P(B|A) = 0.9$$
$$P(B^c|A^c) = 0.7$$
$$P(A^c) =$$
$$P(B|A^c) =$$

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

Will a 3D seismic survey be useful?

Sesimic horizon of late Pleistocene leveed channels offshore Nigeria (Posamentier and Kolla, 2003).

Bayesian Example #1

Example: Prior information at a site suggests a deepwater channel reservoir exists at a given location with probability of 60%. We consider further investigation with a 3D seismic survey.

3D seismic survey will indicate a channelized reservoir:
- is present with 90% probability if it really is present
- is not present with a probability 70% if it really is not

$$P(A) = 0.6$$
$$P(B|A) = 0.9$$
$$P(B^c|A^c) = 0.7$$
$$P(A^c) = 1 - P(A) = 0.4$$
$$P(B|A^c) = 1 - P(B^c|A^c) = 0.3$$

***A*=The deepwater channel is present**
***B* =Seismic shows a deepwater channel**
***A*^{c} =The deepwater channel not present**
***B*^{c} =Seismic does not show a deepwater channel**

Will a 3D seismic survey be useful?

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{\overbrace{(0.9)(0.6)}^{\text{True Positive}}}{\underbrace{(0.9)(0.6)}_{\text{True Positive}} + \underbrace{(0.3)(0.4)}_{\text{False Positive}}} = 0.82$$

Bayesian Example #2

Example: One in every thousand blow out preventers (BOPs) has a serious crack. X-ray analysis has a 99% chance of detecting the crack if present. If the BOP does not have a crack, there is a 2% chance that the X-ray detects a crack. The rate of BOP cracks is 0.1%.

A BOP has been X-rayed and the result is positive. What is the probability that the BOP has a crack?

Solution:

A = **BOP has cracks**

B = **BOP tests positive**

A^c = **BOP does not have cracks**

B^c = **BOP did not test positive**

$P(A|B) =$ **?**

$P(A) = 0.001$ ← crack rate

$P(A^c) =$

$P(B|A) = 0.99$ ← true positive

$P(B|A^c) = 0.02$ ← false positive

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

Blow out preventer

Image from https://www.arescotx.com/photos-videos/what-is-a-blowout-preventer/

Bayesian Example #2

Example: One in every thousand blow out preventers (BOPs) has a serious crack. X-ray analysis has a 99% chance of detecting the crack if present. If the BOP does not have a crack, there is a 2% chance that the X-ray detects a crack. The rate of BOP cracks is 0.1%.

A BOP has been X-rayed and the result is positive. What is the probability that the BOP has a crack?

Solution:

A = **BOP has cracks**

B = **BOP tests positive**

A^c = **BOP does not have cracks**

B^c = **BOP did not test positive**

$P(A|B)$ = **?**

$P(A) = 0.001$ ← crack rate

$P(A^c) = 1.0 - P(A^c) = 0.999$

$P(B|A) = 0.99$ ← true positive

$P(B|A^c) = 0.02$ ← false positive

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{\overbrace{(0.99)(0.001)}^{\textbf{True Positive}}}{\underbrace{(0.99)(0.001)}_{\textbf{True Positive}} + \underbrace{(0.02)(0.999)}_{\textbf{False Positive}}} = 0.047$$

Is the test useful? Probability of a crack in the BOP given a positive crack test is only 0.047! Why?
Cracks are very unlikely + high false positive rate (0.02)! Could be a screening tool?

Bayes' Theorem Hands-on
Prob(Event | Indicator of the Event)

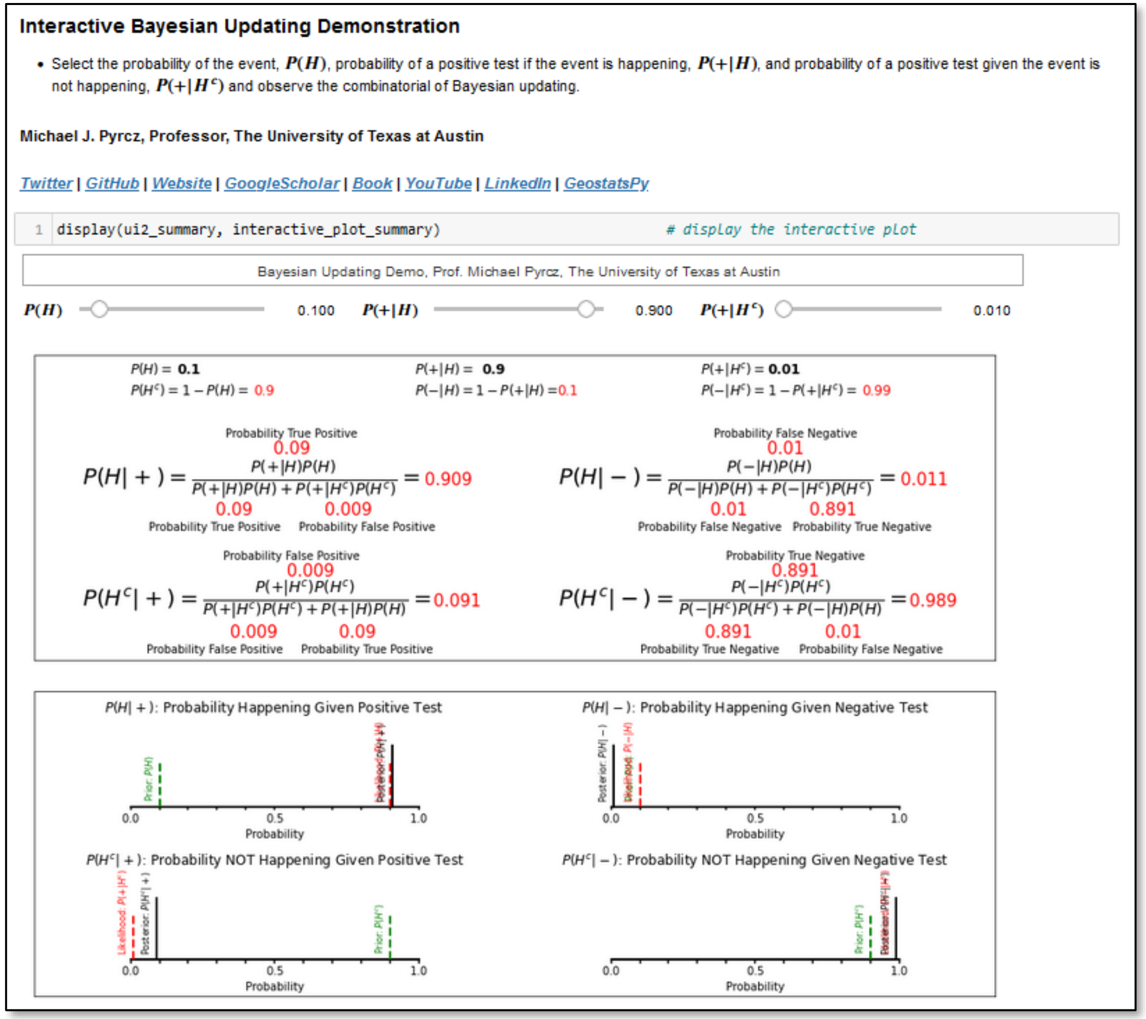
Bayesian Inversion, Value of Information:

Things to try:

1. False Positives:
Drop the false positive rate from 0.01% to 0.001%?

2. Rare Events:
What if probability of occurrence increased from 0.001% to 0.01%?

Observe the impact on the posterior, updated probability.



Interactive Bayesian updating demo (Interactive_Bayesian_Updating.ipynb).

Bayes' Theorem Hands-on
Prob(Event | Indicator of the Event)

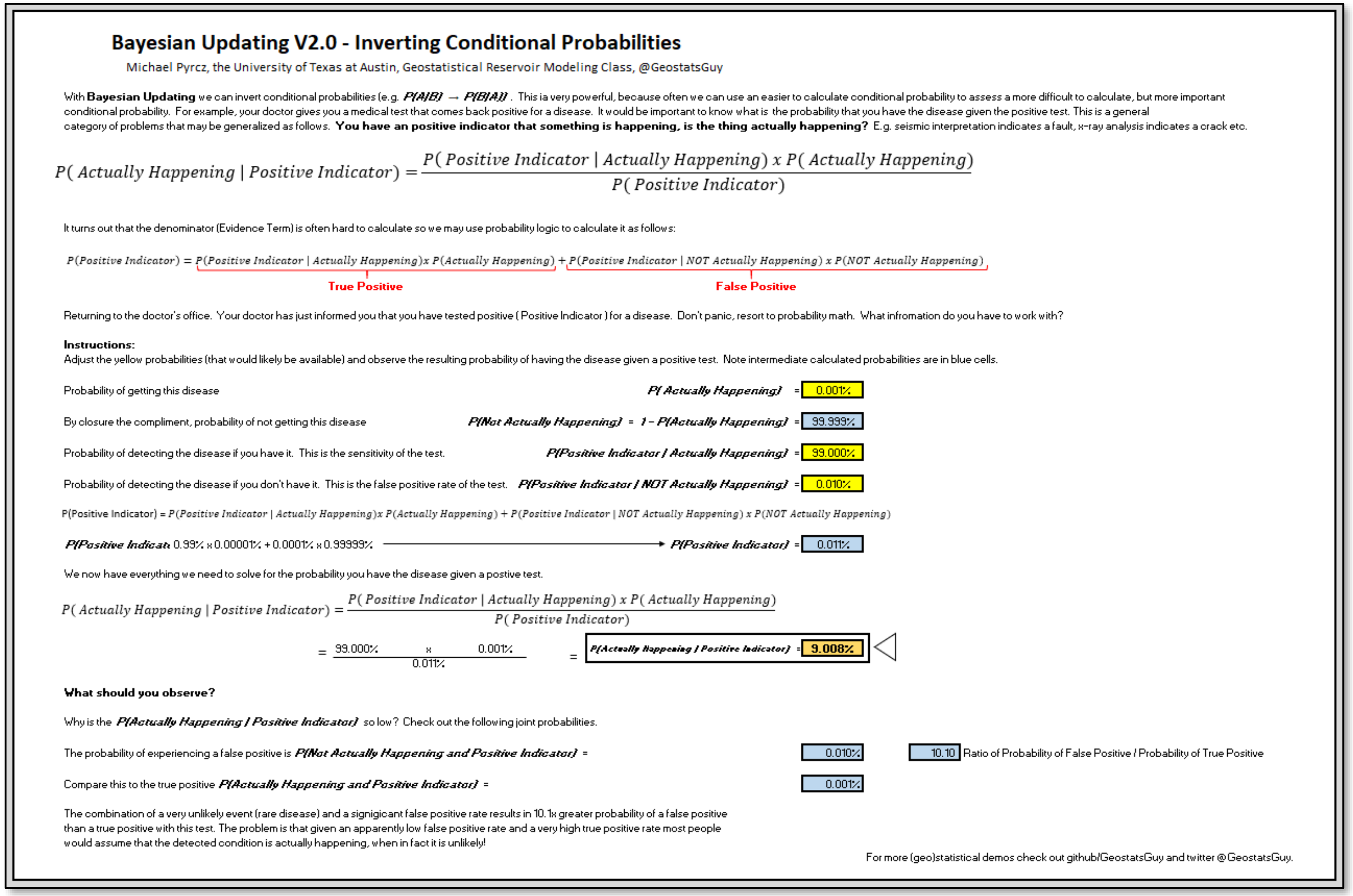
Bayesian Inversion, Value of Information:

Things to try:

1. False Positives:
Drop the false positive rate from 0.01% to 0.001%?

2. Rare Events:
What if probability of occurrence increased from 0.001% to 0.01%?

Observe the impact on the posterior, updated probability.



Bayesian Updating V2.0 - Inverting Conditional Probabilities

Michael Pyrcz, the University of Texas at Austin, Geostatistical Reservoir Modeling Class, @GeostatsGuy

With **Bayesian Updating** we can invert conditional probabilities (e.g. ***P(A|B) → P(B|A)***. This ia very powerful, because often we can use an easier to calculate conditional probability to assess a more difficult to calculate, but more important conditional probability. For example, your doctor gives you a medical test that comes back positive for a disease. It would be important to know what is the probability that you have the disease given the positive test. This is a general category of problems that may be generalized as follows. **You have an positive indicator that something is happening, is the thing actually happening?** E.g. seismic interpretation indicates a fault, x-ray analysis indicates a crack etc.

$$P(\text{Actually Happening} \mid \text{Positive Indicator}) = \frac{P(\text{Positive Indicator} \mid \text{Actually Happening}) \text{ x } P(\text{Actually Happening})}{P(\text{Positive Indicator})}$$

It turns out that the denominator (Evidence Term) is often hard to calculate so we may use probability logic to calculate it as follows:

$$P(\text{Positive Indicator}) = \underbrace{P(\text{Positive Indicator} \mid \text{Actually Happening}) \text{x } P(\text{Actually Happening})}_{\text{True Positive}} + \underbrace{P(\text{Positive Indicator} \mid \text{NOT Actually Happening}) \text{ x } P(\text{NOT Actually Happening})}_{\text{False Positive}}$$

Returning to the doctor's office. Your doctor has just informed you that you have tested positive (Positive Indicator) for a disease. Don't panic, resort to probability math. What infromation do you have to work with?

Instructions:
Adjust the yellow probabilities (that would likely be available) and observe the resulting probability of having the disease given a positive test. Note intermediate calculated probabilities are in blue cells.

Probability of getting this disease	***P(Actually Happening)*** =	0.001%
By closure the compliment, probability of not getting this disease	***P(Not Actually Happening) = 1 - P(Actually Happening)*** =	99.999%
Probability of detecting the disease if you have it. This is the sensitivity of the test.	***P(Positive Indicator / Actually Happening)*** =	99.000%
Probability of detecting the disease if you don't have it. This is the false positive rate of the test.	***P(Positive Indicator / NOT Actually Happening)*** =	0.010%

P(Positive Indicator) = $P(\text{Positive Indicator} \mid \text{Actually Happening}) \text{x } P(\text{Actually Happening}) + P(\text{Positive Indicator} \mid \text{NOT Actually Happening}) \text{ x } P(\text{NOT Actually Happening})$

P(Positive Indicat 0.99% x 0.00001% + 0.0001% x 0.99999% ⟶ ***P(Positive Indicator)*** = 0.011%

We now have everything we need to solve for the probability you have the disease given a postive test.

$$P(\text{Actually Happening} \mid \text{Positive Indicator}) = \frac{P(\text{Positive Indicator} \mid \text{Actually Happening}) \text{ x } P(\text{Actually Happening})}{P(\text{Positive Indicator})}$$

$$= \frac{99.000\% \quad \text{x} \quad 0.001\%}{0.011\%} = \boxed{P(\text{Actually Happening} / \text{Positive Indicator}) = 9.008\%}$$

What should you observe?

Why is the ***P(Actually Happening / Positive Indicator)*** so low? Check out the following joint probabilities.

The probability of experiencing a false positive is ***P(Not Actually Happening and Positive Indicator)*** = 0.010% | 10.10 Ratio of Probability of False Positive / Probability of True Positive

Compare this to the true positive ***P(Actually Happening and Positive Indicator)*** = 0.001%

The combination of a very unlikely event (rare disease) and a signigicant false positive rate results in 10.1x greater probability of a false positive than a true positive with this test. The problem is that given an apparently low false positive rate and a very high true positive rate most people would assume that the detected condition is actually happening, when in fact it is unlikely!

For more (geo)statistical demos check out github/GeostatsGuy and twitter @GeostatsGuy.

Bayesian updating demo (BayesianUpdatingInversion_Demo.xlsx).

Probability Definitions
Bayesian Statistics Examples

What did we learn?

- we can solve many general, important problems if we define the terms and use them consistently in Bayes' theorem
- use marginalization to solve for the evidence term
- combination of rare events and high false positive rates can make the conditional probability of an event given an indication of the event low!
- we can calculate the posterior and compare to the prior and use this to assess the value of information of a test!

$$P\left(\begin{matrix}\text{Something is}\\\text{Happening}\end{matrix} \middle| \begin{matrix}\text{Looks like}\\\text{it's happening}\end{matrix}\right) = \frac{P\left(\begin{matrix}\text{Looks like}\\\text{it's happening}\end{matrix} \middle| \begin{matrix}\text{Something is}\\\text{Happening}\end{matrix}\right) P\left(\begin{matrix}\text{Something is}\\\text{Happening}\end{matrix}\right)}{P\left(\begin{matrix}\text{Looks like}\\\text{it's happening}\end{matrix}\right)}$$

Bayesian Example #3

You have 3 machines making the same product (product lines). They have different volumes and errors.

Machine 1	Machine 2	Machine 3
$P(X_1)$, **20% Production**	$P(X_2)$, **30% Production**	$P(X_3)$, **50% Production**
$P(Y\|X_1)$, **5% Error Rate**	$P(Y\|X_2)$, **3% Error Rate**	$P(Y\|X_3)$, **1% Error Rate**

Events - Y: Error, X_1: Machine 1, X_2: Machine 2, X_3: Machine 3 source machines, respectively.

Example: Probability of an error in the product, $P(Y)$?

Hint: Calculate Marginal $P(Y) = \sum_{i=1}^{n} P(Y, X_i) = \sum_{i=1}^{n} P(Y|X_i)P(X_i)$ since exhaustive and mutually exclusive events.

Example from https://en.wikipedia.org/wiki/Bayes%27_theorem

Bayesian Example #3

You have 3 machines making the same product (product lines). They have different volumes and errors.

Machine 1	Machine 2	Machine 3
$P(X_1)$, **20% Production**	$P(X_2)$, **30% Production**	$P(X_3)$, **50% Production**
$P(Y\|X_1)$, **5% Error Rate**	$P(Y\|X_2)$, **3% Error Rate**	$P(Y\|X_3)$, **1% Error Rate**

Events - Y**: Error,** X_1**: Machine 1,** X_2**: Machine 2,** X_3**: Machine 3 source machines, respectively.**

Example: Probability of an error in the product, $P(Y)$?

$$P(Y) = \sum_{i=1}^{n} P(Y, X_i) = \sum_{i=1}^{n} P(Y|X_i)P(X_i)$$

$$P(Y) = P(Y|X_1)P(X_1) + P(Y|X_2)P(X_2) + P(Y|X_3)P(X_3)$$

$$P(Y) = (0.20)(0.05) + (0.30)(0.03) + (0.50)(0.01)$$

$$P(Y) = 0.024 = 2.4\%$$

Example from https://en.wikipedia.org/wiki/Bayes%27_theorem

Bayesian Example #4

You have 3 machines making the same product (product lines). They have different volumes and errors.

Machine 1	Machine 2	Machine 3
$P(X_1)$, **20% Production**	$P(X_2)$, **30% Production**	$P(X_3)$, **50% Production**
$P(Y\|X_1)$, **5% Error Rate**	$P(Y\|X_2)$, **3% Error Rate**	$P(Y\|X_3)$, **1% Error Rate**

Events - Y: Error, X_1: Machine 1, X_2: Machine 2, X_3: Machine 3 source machines, respectively.

Example: Probability product came each machine given an error is observed, $P(X_i|Y)$?

Note: From the previous slide: $P(Y) = 0.024 = 2.4\%$

Hint: calculate the conditional: $P(X_i|Y) = \frac{P(Y|X_i)P(X_i)}{P(Y)}$

Example from https://en.wikipedia.org/wiki/Bayes%27_theorem

Bayesian Example #4

You have 3 machines making the same product (product lines). They have different volumes and errors.

Machine 1	Machine 2	Machine 3
$P(X_1)$, **20% Production**	$P(X_2)$, **30% Production**	$P(X_3)$, **50% Production**
$P(Y\|X_1)$, **5% Error Rate**	$P(Y\|X_2)$, **3% Error Rate**	$P(Y\|X_3)$, **1% Error Rate**

Events - Y: Error, X_1: Machine 1, X_2: Machine 2, X_3: Machine 3 source machines, respectively.

Example: Probability product came each machine given an error is observed, $P(X_i|Y)$?

$$P(X_i|Y) = \frac{P(Y|X_i)P(X_i)}{P(Y)}$$

$$P(X_1|Y) = \frac{(0.05)(0.2)}{(0.024)} = 0.41$$

$$P(X_2|Y) = \frac{(0.03)(0.3)}{(0.024)} = 0.38$$

$$P(X_3|Y) = \frac{(0.01)(0.5)}{(0.024)} = 0.21$$

Note we can check closure:

$$P(X_1|Y) + P(X_2|Y) + P(X_3|Y) = 0.41 + 0.38 + 0.21 = 1.0$$

Example from https://en.wikipedia.org/wiki/Bayes%27_theorem

Bayesian Hands-on
Updating Exploration Success

Exploration Bayesian Updating for Exploration Drilling:

Things to try:

1. Change the Recent Drilling Success from 33% to 10% or 80%.

2. Change the Prior from mode of 45% to 10% or 50%.

Observe the impact on the posterior, updated probability.

Bayesian Updating, Bayes' Theorem for Updating Exploration Success Rate with New Exploration Drilling Results

Michael Pyrcz, Associate Professor, the University of Texas at Austin

Problem: update the assumed exploration success rate with new exploration drilling results. Update the prior exploration probability of exploration success with n_s drilling successes out of n new exploration wells.

$$Prob\{\text{Model}|Result\} = \frac{Prob\{\text{Result}|\text{Model}\} \cdot Prob\{Model\}}{Prob\{\text{Result}\}}$$

we can use Bayes' Theorem go from Prob{ Result | Model } (probability of exploration drilling outcome given exploration model) that is easy to calculate to the Prob { Model | Outcome } (probabilty of the exploration model success rate given drilling outcomes) that is not available.

$$Posterior = \frac{Likelihood \cdot \text{Prior}}{Evidence}$$

the prior is our belief of the probability of each possible exploration success rate (an uniform probability distribution is a naïve prior - we don't know) before drilling the new exploration wells.
Likelihood comes from the binomial distribution. Evidence is the normalization constant such that the resulting posteriori PDF sums to 1.0.

$$Prob\{\text{Result}|\text{Model}\} = \binom{n_s}{n} P(n_s)^{n_s} \cdot (1 - P(n_s))^{n-n_s}, n_s$$

where n_s is the number successes, n is the total number of wells and $P(n_s)$ is the probability of exploration success.

$$Prob\{Result\} = \text{k}$$

the evidence term is a constant to ensure closure (all posterior probabilities sum to 1.0)

1. Data results - Exploration Outcome

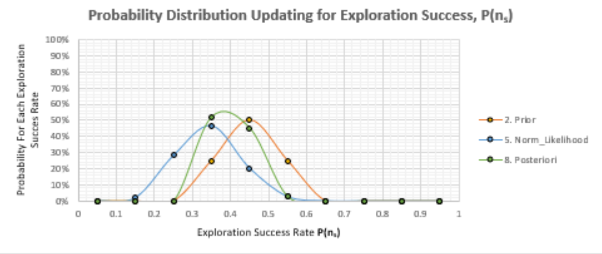
Success, n_s	10
Failures, n_f	20
Total Wells, n	30

Experimental Exploration Success Rate: 33.3%

Rate from these recent wells.

Prior, Likelihood and Posterior probabilities Binned by Probability of Exploration Drilling Success

Exploration Success Rate, $P(n_s)$	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95	Sum
2. Prior	0.00000	0.00000	0.00000	0.10000	0.20000	0.10000	0.00000	0.00000	0.00000	0.00000	0.4000
3. Norm_Prior	0.00000	0.00000	0.00000	0.25000	0.50000	0.25000	0.00000	0.00000	0.00000	0.00000	
4. Likelihood	0.00000	0.00672	0.09087	0.15022	0.06564	0.00882	0.00031	0.00000	0.00000	0.00000	0.3226
5. Norm_Likelihood	0.00000	0.02082	0.28169	0.46569	0.20348	0.02735	0.00095	0.00000	0.00000	0.00000	
6. Prior x Likelihood	0.00000	0.00000	0.00000	0.01502	0.01313	0.00088	0.00000	0.00000	0.00000	0.00000	0.0290
7. Evidence	0.02903	0.02903	0.02903	0.02903	0.02903	0.02903	0.02903	0.02903	0.02903	0.02903	
8. Posteriori	0.00000	0.00000	0.00000	0.51743	0.45218	0.03039	0.00000	0.00000	0.00000	0.00000	1.0000



Instructions for Bayes' Theorem Excel Demo

1. Set any data outcome, Data. Where Heads, n_s, is the number of exploration sucesses and $n - n_s$ is the number of epxloration failures and n is the total number of exploration wells.
2. Set the prior to any set of relative probabilities to reflect prior belief concerning the exploration drilling success rate prior to drilling the new exploration wells. Constant is a naïve prior (no idea) or higher for 0.4 reflects a prior / belief in a 40% exploration success rate.
3. The prior probabilities for each exploration success rate bin are standardized to sum to 1.0 as expected for a PDF.
4. The likelihood calculated from the binomial distribution based on the exploration drilling outcome.
5. The likelihood normalized sum to 1.0 as expected for a PDF (for plotting).
6. The product of the prior and the likelihood.
7. The evidence term as the sum of the product of prior and likelihood to ensure the posteriori sums to 1.0 over the exploration success rate bins as expected for a PDF.
8. The posterior as the product of prior and likelihood standardized by evidence for each exploration success rate bin.

What did we learn?

1. Bayes' Theorem may be applied to calculated conditional probabilities that otherwise would be difficult to assess.
2. The prior model has a significant impact on the posterior and must be selected carefully.
3. For a naïve prior the posterior is equal to the likelihood.

Based on Sivia, D.S., 1996, Data Analysis, A Bayesian Tutorial, Oxford Science Publications, 189 p.

Bayeisan updating with Gaussian distribution demo (Bayesian_Exploration_Demo.xlsx).

Bayesian Updating with Gaussian Distributions

There is an analytical solution for working with Gaussian parametric distributions for Bayesian updating (Sivia, 1996).

- Calculate the mean of the posterior from the prior and likelihood mean and variance.

$$\bar{x}_{posterior} = \frac{\bar{x}_{likelihood} \cdot \sigma^2_{prior} + \bar{x}_{prior} \cdot \sigma^2_{likelihood}}{\left[1 - \sigma^2_{likelihood}\right]\left[\sigma^2_{prior} - 1\right] + 1}$$

- Calculate the variance of the posterior form the prior and likelihood variances (no means, homoscedastic!).

$$\sigma^2_{posterior} = \frac{\sigma^2_{prior} \cdot \sigma^2_{likelihood}}{\left[1 - \sigma^2_{likelihood}\right]\left[\sigma^2_{prior} - 1\right] + 1}$$

We will formalize mean (arithmetic average) and variance next lecture and the Gaussian parametric distribution later.

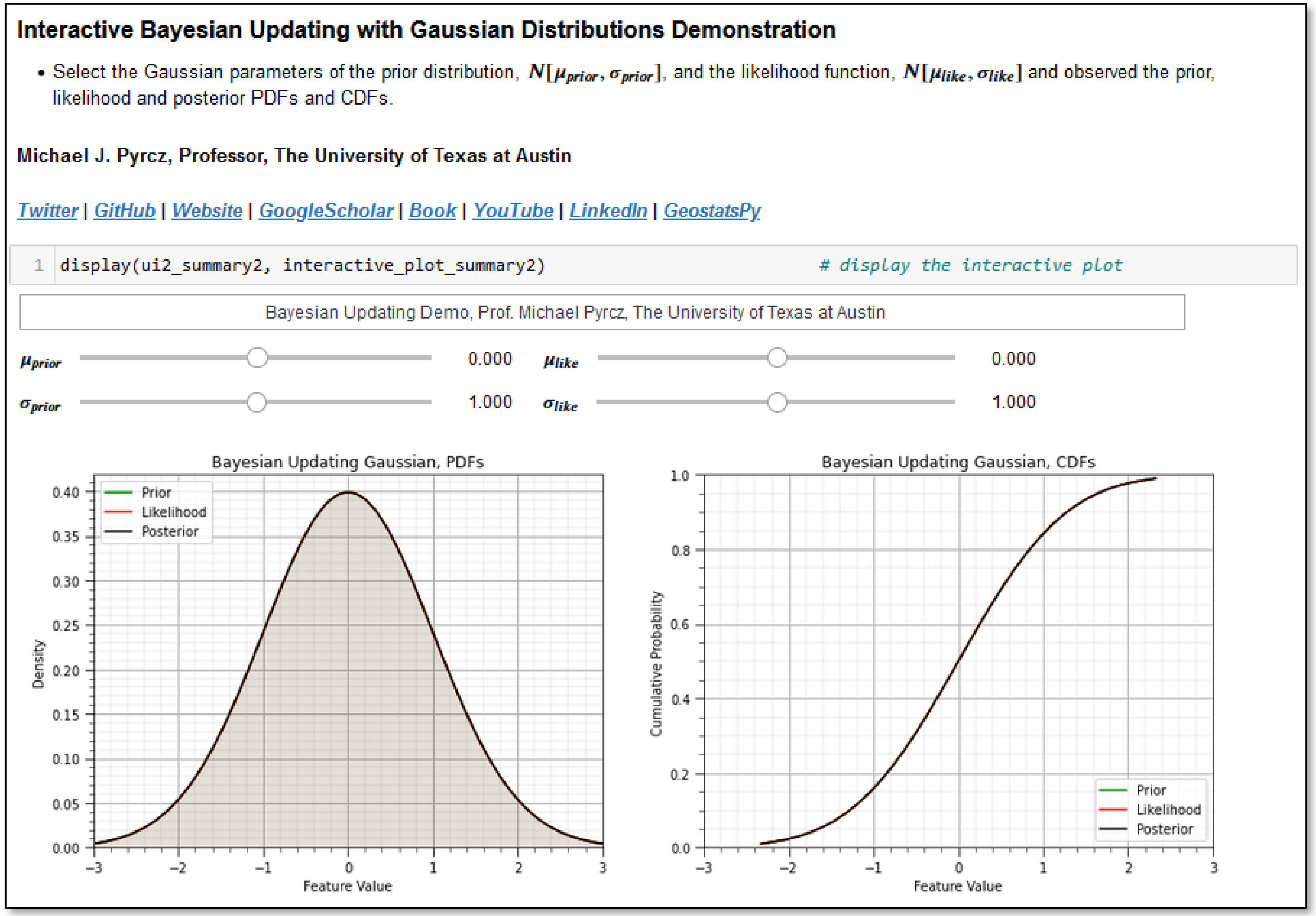
Bayesian Hands-on
Updating with Gaussian Distributions

Bayesian Updating with Gaussian Distributions

Jupyter Notebook Python Demonstration

Things to try:

1. Try a naïve prior.
2. Try of very specific prior.
3. Try a naïve and specific likelihood function.
4. Include contradiction between prior and likelihood.



Bayeisan updating with Gaussian distributions (Interactive_Bayesian_Updating.ipynb).

Bayesian Hands-on
Updating with Gaussian Distributions

Bayesian Updating with Gaussian Parametric Distribution for Exploration Drilling:

Things to try:

1. Use a large variance for the prior.

2. Make the prior and likelihood distributions the same.

3. Make the prior low, and the likelihood even lower!

Observe the impact on the posterior, updated probability.

Bayesian Gaussian Analytical Example Demo

Michael Pyrcz, the University of Texas at Austin, Geostatistical Reservoir Modeling Class

Formulation from Sivia, 1996.

$$\bar{x}_{\text{updated}} = \frac{\bar{x}_{\text{likelihood}}(\mathbf{u}) \cdot \sigma^2_{\text{prior}}(\mathbf{u}) + \bar{x}_{\text{prior}}(\mathbf{u}) \cdot \sigma^2_{\text{likelihood}}(\mathbf{u})}{[1 - \sigma^2_{\text{likelihood}}(\mathbf{u})][\sigma^2_{\text{prior}}(\mathbf{u}) - 1] + 1}$$

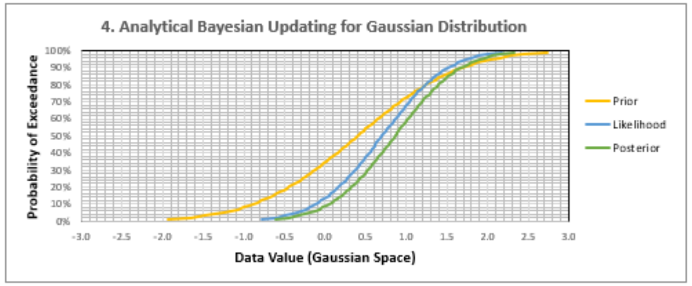
$$\sigma^2_{\text{updated}}(\mathbf{u}) = \frac{\sigma^2_{\text{prior}}(\mathbf{u})\, \sigma^2_{\text{likelihood}}(\mathbf{u})}{[1 - \sigma^2_{\text{likelihood}}(\mathbf{u})][\sigma^2_{\text{prior}}(\mathbf{u}) - 1] + 1}$$

1. Prior Distribution	
average	0.40
variance	1.00

2. Likelihood Distribution	
average	0.70
variance	0.40

3. Posterior Distribution	
average	0.86
variance	0.40

Percentile	Prior	Likelihood	Posterior
0.01	-1.926	-0.771	-0.611
0.02	-1.654	-0.599	-0.439
0.03	-1.481	-0.490	-0.330
0.04	-1.351	-0.407	-0.247
0.05	-1.245	-0.340	-0.180
0.06	-1.155	-0.283	-0.123
0.07	-1.076	-0.233	-0.073
0.08	-1.005	-0.189	-0.029
0.09	-0.941	-0.148	0.012
0.1	-0.882	-0.111	0.049
0.11	-0.827	-0.076	0.084
0.12	-0.775	-0.043	0.117
0.13	-0.726	-0.012	0.148
0.14	-0.680	0.017	0.177
0.15	-0.636	0.045	0.205
0.16	-0.594	0.071	0.231
0.17	-0.554	0.097	0.257
0.18	-0.515	0.121	0.281
0.19	-0.478	0.145	0.305
0.2	-0.442	0.168	0.328
0.21	-0.406	0.190	0.350
0.22	-0.372	0.212	0.372
0.23	-0.339	0.233	0.393
0.24	-0.306	0.253	0.413
0.25	-0.274	0.273	0.433
0.26	-0.243	0.293	0.453
0.27	-0.213	0.312	0.472
0.28	-0.183	0.331	0.491
0.29	-0.153	0.350	0.510
0.3	-0.124	0.368	0.528
0.31	-0.096	0.386	0.546
0.32	-0.068	0.404	0.564
0.33	-0.040	0.422	0.582
0.34	-0.012	0.439	0.599
0.35	0.015	0.456	0.616
0.36	0.042	0.473	0.633



Instructions for Analytical Bayesian Updating for Gaussian Distributions

1. Set the average and the variance of the prior distribution (Gaussian parametric distribution).
2. Set the average and the variance of the likelihood distribution (Gaussian parametric distribution).
3. Observed the updated average and variance of the posterior distribution (Gaussian parametric distribution).
4. Observed the prior, likelihoof and posterior cumulative distribution functions (CDFs).

What did we learn?

1. The posterior variance is only a function of the prior and likelihood variances. The prior and likelihood means have no influence.
2. In general updating results in a reduction variance. Posterior variance is equal to or less than the greater of the prior and the likelihood variance.
3. High certainty in either prior or likelihood distribution (very low variance) causes either term to dominate the updated posterior.

Sivia, D.S., 1996, Data Analysis, A Bayesian Tutorial, Oxford Science Publications, 189 p.

Bayeisan updating with Gaussian distributions (Bayesian_Gaussian_Demo.xlsx).

Probability Definitions

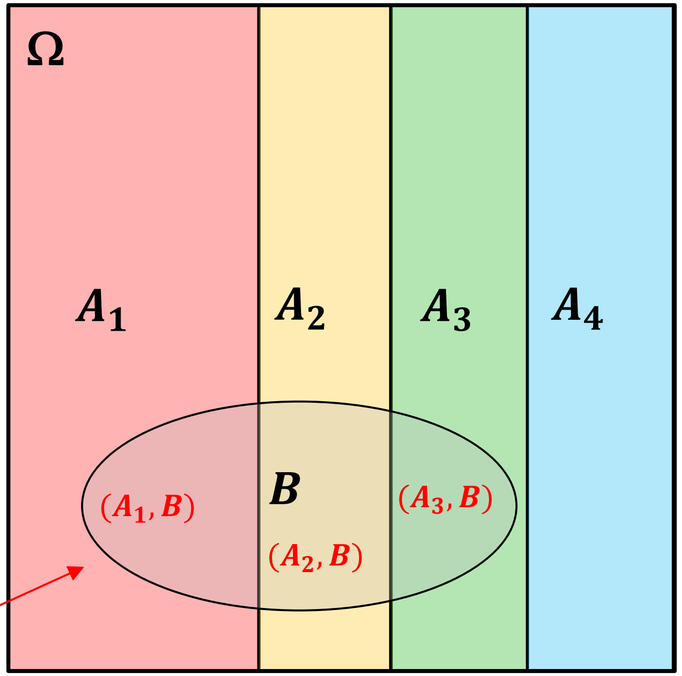
Bayesian Theorem General Form

Bayesian probability, expanding beyond 2 mutually exclusive, exhaustive events.

General Form:

$$P(A_k \mid B) = \frac{P(B|A_k)\, P(A_k)}{P(B)}$$

if non-overlapping $A_i \bigcap A_j = \emptyset, \forall i, \forall j, i \neq j$ and exhaustive $\bigcup_{k=1}^{K} A_k = \Omega$



Venn Diagram – illustrating exhaustive, mutually exclusive series.

then: $$P(B) = \sum_{k=1}^{K} P(B|A_k)\, P(A_k) = \sum_{k=1}^{K} P(B, A_k)$$

we substitute: $$P(A_k \mid B) = \frac{P(B|A_k)\, P(A_k)}{\sum_{k=1}^{K} P(B, A_k)}$$

Careful, can't do this if not mutually exclusive and exhaustive events.

- More complicated to calculate evidence, $P(B)$

Probability New Tools

Topic	Application to Subsurface Modeling
Frequentist Concepts	When sufficient observations are available use (long-run) counting to access the required probabilities. *Predict reservoir average porosity by pooling analogous fields.*
Bayesian Concepts Inversion of Conditionals	Calculate a difficult to access conditional probability from an accessible one. Probability of event given indicator from probability indicator given event. *Calculate probability of sealing fault given indicator of sealing fault.*
Bayesian Concepts Bayesian Updating	Update prior belief with new information. *Calculate probability of exploration success rate given prior model and outcomes from exploration drilling program.*

PGE 383 Subsurface Machine Learning

Lecture 2: Probability

Lecture outline:

- **Probability Definitions**
- **Venn Diagrams, Probability Operations, Frequentist Concepts**
- **Bayesian Concepts**