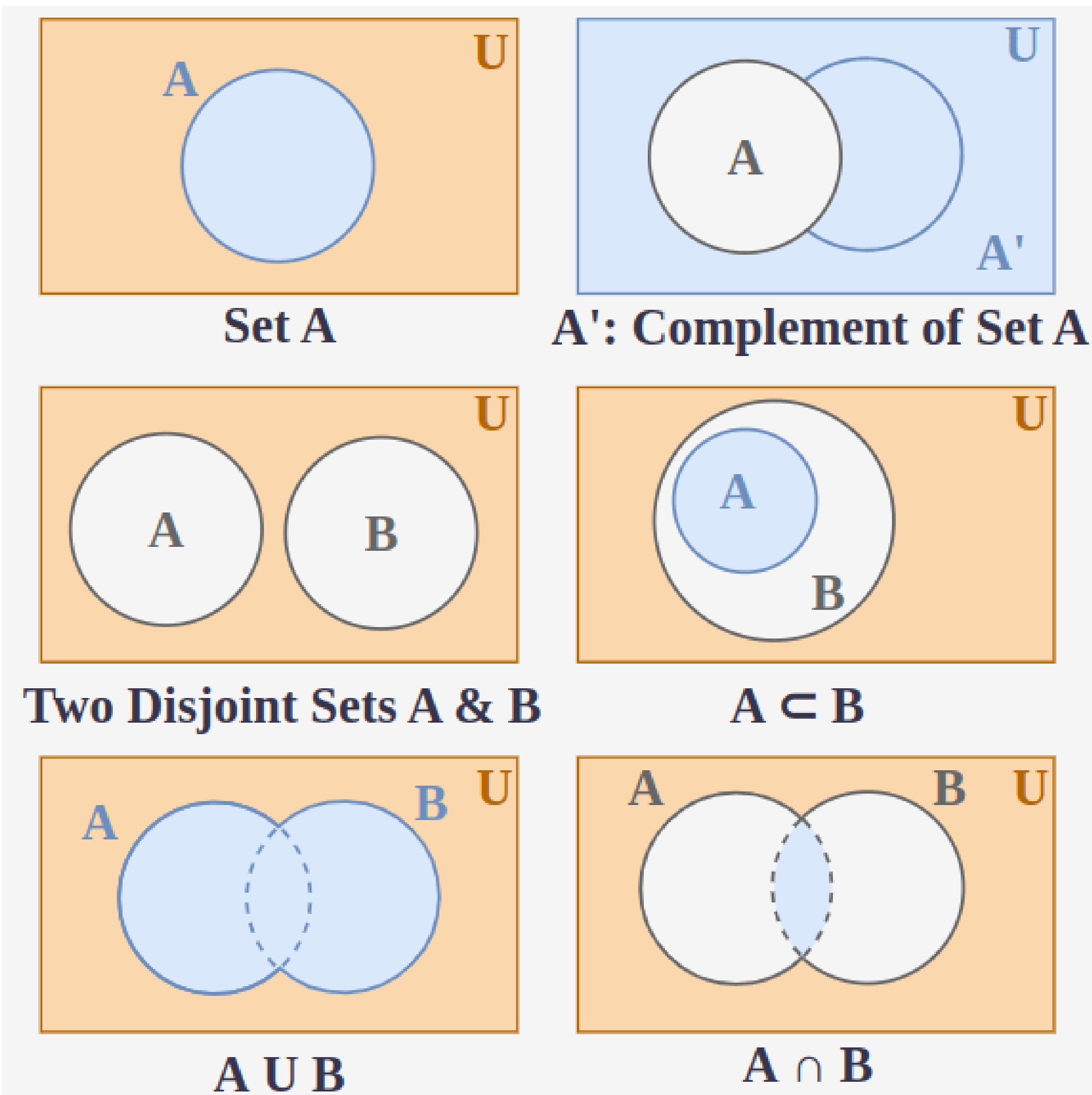
Methods 3, Week 3:

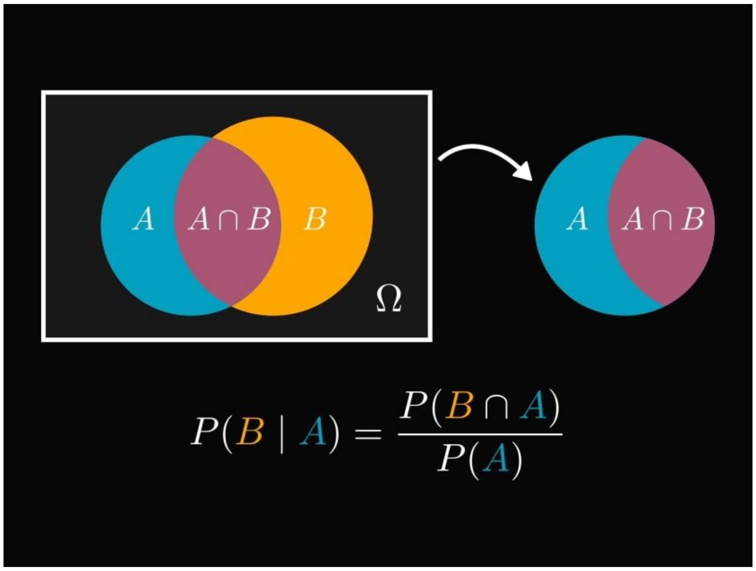
Bayesian Thinking: Experimental and Implementational Aspects

Maxime Sainte-Marie

A quick look back at set theory



A quick look back at conditional probability



A quick look back at Bayes' Theorem

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

- $Pr(B|A)$ = Conditional probability that needs to be reversed
- $Pr(A)$ = Categorical (marginal) probability of unknown event
- $Pr(B)$ = Categorical (marginal) probability of known event

$Pr(A)$ = Marginal probability of unknown event (Prior)

- Most analysts assume uninformative priors
  - no strong assumptions about the parameter estimates other than the shape of their distribution
  - the results are strongly determined by the current experiment's data
  - no strong rationale for bayesian
- Informative priors
  - Bigger effect on the posterior estimation
  - Effect larger if the data is limited or highly variable

$Pr(B)$ = Marginal probability of known event

- Probability of known event, regardless of whether the unknown event occurs
  - Sum of all possible joint probabilities of known event B with all possible unknown events A_i in sample space S_A

$Pr(B) = \sum_{A_i \in S_A} Pr(A_i \cap B) = \sum_{A_i \in S_A} Pr(B|A_i)Pr(A_i)$

Bayes' theorem becomes:

$Pr(A|B) = Pr(A)\frac{Pr(B|A)}{\sum_{A_i \in S_A} Pr(A_i \cap B)} = Pr(A)\frac{Pr(B|A)}{\sum_{A_i \in S_A} Pr(B|A_i)Pr(A_i)}$

Example: Prostate cancer test

- A 30 year-old man has a positive blood test for a prostate cancer marker (PSA). What is the probability he has prostate cancer (can) given one positive test (pos)?

- The test gives positive results to positive cases 90% of the time (TPR = .9)
- The test gives gives positive results to negative cases 50% of the time (FPR = .5).
- Prostate cancer incidence rate for 30 year-old men is 0.001%
- Possible unknown events in the sample space:
  - Cancer (can)
  - Not cancer ($\sim can$)

Estimating prostate cancer probability

Given

$$Pr(\text{pos}|\text{can}) = .9 \tag{1}$$
$$Pr(\text{pos}|\sim\text{can}) = .5 \tag{2}$$
$$Pr(can) = .00001 \tag{3}$$

Then

$$Pr(can|pos) = Pr(can)\frac{Pr(pos|can)}{(Pr(pos|can)Pr(can)) + (Pr(pos|\sim can)Pr(\sim can))} \tag{4}$$
$$Pr(can|pos) = .00001 \times \frac{.9}{(.9 \times .00001) + (.5 \times (1 - .00001))} \tag{5}$$
$$Pr(can|pos) = .000018 \tag{6}$$
$$\tag{7}$$

Bayesian Updating of Prostate Cancer Probability

- Second positive test?
  - New Prior = posterior probability for first test ($Pr(A) = .000018$)

Then

$$Pr(can|pos) = .000018 \times \frac{.9}{(.9 \times .000018) + (.5 \times .999982)} \tag{8}$$
$$Pr(can|pos) = .000032 \tag{9}$$

More positive tests?

$3^{rd} = .000058, \;\; 4^{th} = .0001, \;\; 5^{th} = .00018, \;\; 6^{th} = .00032 \;\; ...$

Incremental Science

- We continue to gather data to evaluate a particular scientific hypothesis
  - We do not begin anew each time we attempt to answer a hypothesis, because prior research provides us with a priori information concerning the merit of the hypothesis
  - This approach moves psychology from a bunch of individual studies that are occasionally combined using meta-analysis to one that approaches the best estimate for parameters by having earlier studies inform the current one

From probability points to probability distributions

- Exact probabilities are rare -> multiple factors at play
  - Prostate cancer: Family history, genetics, obesity, height, pesticides, ethnicity, hormones…
  - Representativity of samples, relative frequencies
  - Population quantities are not known or fixed
- Distributions allow to represent uncertainty regarding the true value of specific parameters
  - Represent prior uncertainty about model parameters with a probability distribution, and update prior uncertainty with evidence to get less uncertain posterior probability distributions for parameters

Distributional Bayes

$f(\theta|data) = f(\theta)\frac{f(data|\theta)}{f(data)}$

$$
\begin{aligned}
f(\theta|data) &= \text{posterior distribution of the parameter } \theta && (10)\\
f(data|\theta) &= \text{sampling density for the data} && (11)\\
&\quad \text{(not a proper density function,} && (12)\\
&\quad \text{but proportional to likelihood function)} && (13)\\
f(\theta) &= \text{Prior distribution for the parameter} && (14)\\
f(data) &= \text{Marginal probability of the data} = \int f(data|\theta)f(\theta)d\theta && (15)\\
&\quad \text{(aka marginal likelihood, acts as normalizing constant} && (16)\\
&\quad \text{to make posterior density proper)} && (17)
\end{aligned}
$$

Marginal probability/likelihood

$f(data) = \int f(data|\theta)f(\theta)d\theta$

- Normalizing constant making posterior density proper
  - Simply scales the posterior density, so often left out

$Posterior \propto Prior \times Likelihood$

- Proportionality
  - If a is proportional to b, then a and b differ only by a multiplicative constant
    - Not a problem for Bayesian analysis
      - Not taking into account the marginal likelihood does not alter the relative frequencies of different values of a random variable

Implementation of Bayesian inference

- Obtaining the analytical form of posterior distributions is generally intractable
  - No closed form expression for the product of the likelihood function by the prior distribution
    - A closed-form expression is a mathematical expression that can be evaluated in a finite number of expressions
      - Involves some kind of limit/infinity
    - Data distribution cannot be written down on paper
- Posterior distributions most often approximated
  - Exception: Conjugate priors and prior mixtures

MCMC Posterior Distribution Approximation

- Computer–driven sampling method allowing to characterize a distribution without knowing all of the distribution's mathematical properties by randomly sampling values out of the distribution.
- Monte-Carlo
  - Practice of estimating the properties of a distribution by examining random samples from the distribution
  - As the number of samples increases, it converges to an accurate representation of the posterior
- Markov Chain = Special sequential process
  - Chain: Each random sample is used as a stepping stone to generate the next random sample (hence the chain)
    - Markovian: each new sample depends only on the one before it

BRMS (Bayesian Regression Models using Stan)

- Attempt at bridging frequentist and Bayesian practices by taking up and extending lme4's syntax
- Gateway to STAN (more flexible, covered in Advanced Cognitive Modeling)

Bayes? Why?... BA*WHY*ES?

```
## PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed, please
```

Advantages

- Incorporation of prior knowledge
  - Useful when historical data or expert knowledge can provide insights
- Flexibility
  - Can handle many statistical models: complex hierarchical models, non-linear models, models with many parameters
- Uncertainty quantification
  - Measure, shrink, propagate uncertainty in hypotheses and parameters
- Small sample sizes
  - Incorporation of prior information stabilizes estimates when samples are small

More advantages

- Sequential analysis
  - Bayesian updating allows for incremental inferences based on gradual accumulation of data
- Posterior distribution
  - Provide more information than point estimates
- Handling missing data
  - Treat missing values as parameters to estimate
- Bayesian Machine Learning
  - Model generalization, uncertainty in predictions, model selection and hyperparameter tuning

Challenges

- Complexity
  - less ritualistic
    - justification-based
    - polymorphic
  - longer
  - fragile
- conducive to awareness and thoroughness

– Background knowledge, assumptions, implications