

Rules of Probability

Bayesian Modeling for Socio-Environmental Data

Chris Che-Castaldo, Mary B. Collins, N. Thompson Hobbs

Summer 2018



Road map

- Rules of probability
 - ▶ Conditional probability
 - ▶ Chain rule
 - ▶ Independence
 - ▶ The law of total probability
- Factoring joint probabilities

Why do we need to know this stuff?

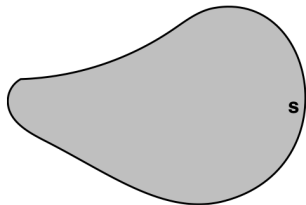
- ① **Conditional probability** foundational for all the inferences that we make.
- ② **The law of total probability** is the denominator of Bayes' Theorem.
- ③ **Factoring** joint distributions is how we deal with complexity, reducing high dimensional problems (the chain rule allows us to do this).
- ④ **Independence** allows us to simplify fully factored joint distributions.

Random variables

- are quantities governed by chance.
- have a specific value called an *events* or *outcomes*.
- are summarized by probability distributions.
- *Bayesians treat every unobserved quantity as a random variable.*

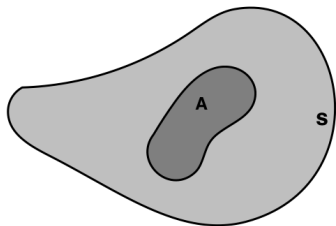
S =Sample Space

- The set of all possible events or outcomes of an experiment or survey.
- The sample space, S has a specific area.

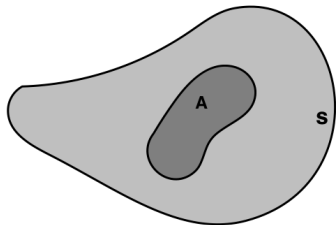


Events in S

- Can define an event, A .
- The area of event A is less than S .



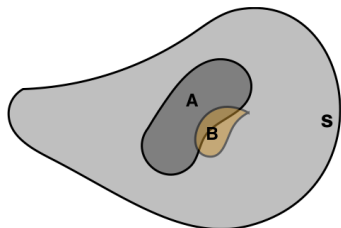
What is the probability of event A?



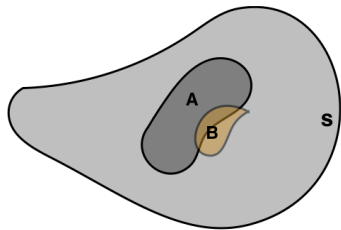
$$\Pr(A) = \frac{\text{Area of } A}{\text{Area of } S}$$

Conditional Probability

Conditional probability: the probability of an event given that *we know* another event has occurred.

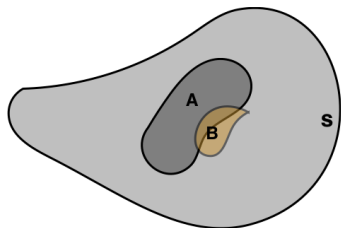


What is the probability of event B , given that event A has occurred?



$\Pr(B|A)$ = probability of B conditional on knowing A has occurred

What is the probability of event B , given that event A has occurred?



$$Pr(B|A) = \frac{\text{Joint Probability}}{\text{Probability of A}} = \frac{Pr(A,B)}{Pr(A)}$$

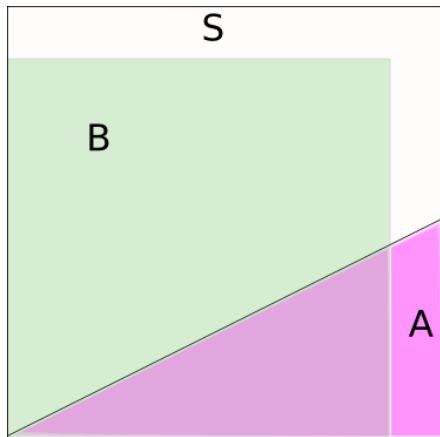
If the occurrence of event A does not tell us anything about event B ?

*In this case, events A and B are said to be **independent***

Events are independent if and only if. . .

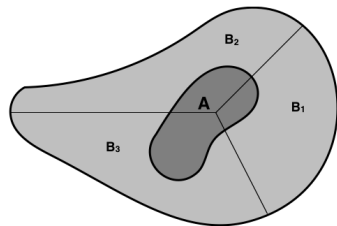
$$\Pr(A|B) = \Pr(A)$$

Assuming independence, the joint probability of event A and event B



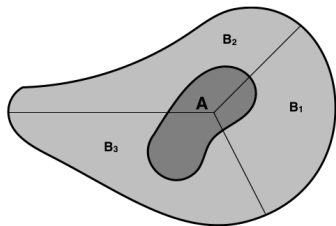
$$\Pr(A, B) = \Pr(A)\Pr(B)$$

The Law of Total Probability



We can define a set of events $\{B_n : n = 1, 2, 3, \dots\}$, which taken together define the entire sample space, $\sum_n B_n = S$.

What is the probability of event A?



$$\Pr(A) = \sum_n \Pr(A|B_n) \Pr(B_n) \text{ (discrete case)}$$

$$\Pr(A) = \int \Pr(A|B) \Pr(B) dB \text{ (continuous case)}$$

The Chain Rule of Probability

The chain rule of probability allows us to calculate any number of joint distributions using only conditional probabilities.

$$\Pr(z_1, z_2, \dots, z_n) = \Pr(z_n | z_{n-1}, \dots, z_1) \dots \Pr(z_3 | z_2, z_1) \Pr(z_2 | z_1) \Pr(z_1)$$

Notice the pattern here.

- z 's can be scalars or vectors.
- Sequence of conditioning doesn't matter.
- When we build models, we choose a sequence that makes sense.

Chain rule of probability board work and independence

Factoring joint probabilities

Why is factoring useful?

- The rules of probability allow us to simplify complicated joint distributions, breaking them down into chunks.
- Chunks can be analyzed one at a time.
- Provide a usable graphical and mathematical foundation, *critical* for the model specification step.

Consider a Bayesian Network (represented by a directed acyclic graph or DAG)



- Bayesian networks specify how joint distributions are factored into conditional distributions using nodes to represent RV's and arrows to represent dependencies.
- Nodes at the heads of arrows *must* be on the left hand side of the conditioning symbols;
- Nodes at the tails of arrows are on the right hand side of the conditioning symbols.
- Any node at the tail of an arrow without an arrow leading into it must be expressed unconditionally.

Factoring with DAGs at the board

Work on lab

Complete parts I-VI