

Rules of Probability

Bayesian Modeling for Socio-Environmental Data

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

June 2019



Road map for today

- Rules of probability
- Factoring joint probabilities
- Directed acyclic graphs (a.k.a. Bayesian networks)

All of Bayesian inference extends from three rules of probability

- 1 Conditional probability (and independence)
- 2 The law of total probability
- 3 The chain rule of probability

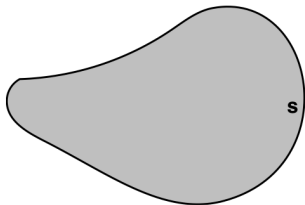
Random variables

The world can be divided into things that are observed and things that are unobserved.

- ① Bayesians treat all unobserved quantities as *random variables*.
- ② The values of random variables are governed by chance.
- ③ Probability distributions describe “governed by chance.”
- ④ A specific value of a random variable is called an event or an outcome.

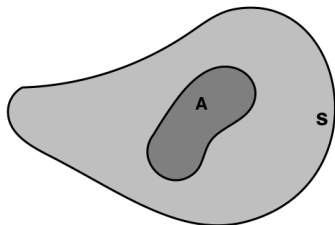
S =Sample Space

- The set of all possible values of a random variable.
- The sample space, S has a specific area.

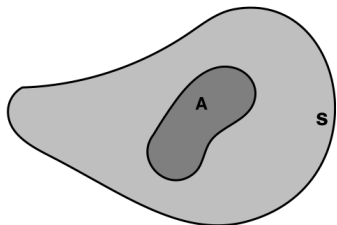


Events in S

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



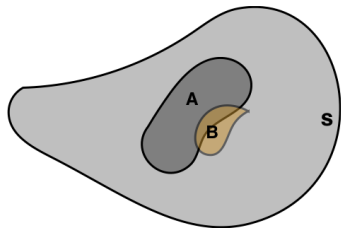
What is the probability of event A?



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

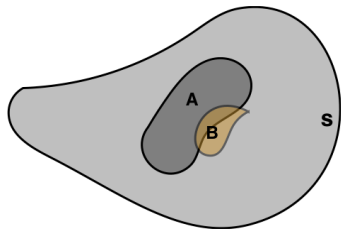
Conditional Probability

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



Conditional Probability

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

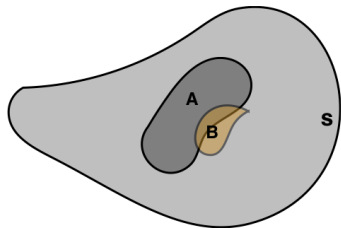


$\Pr(B|A)$ = probability of B conditional on knowing A has occurred

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

Conditional Probability

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



Independence

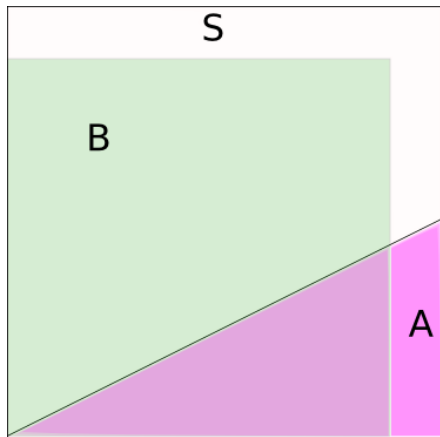
Event A and B are *independent* If the occurrence of event A does not tell us anything about event B .

Events are independent if and only if:

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

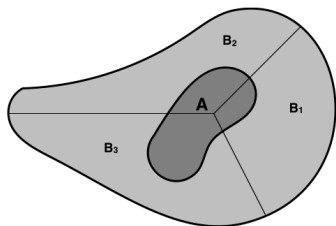
$$\Pr(B|A) = \Pr(B)$$

Independence



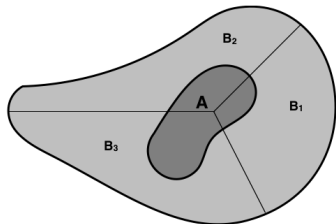
$$\Pr(A|B) = \frac{\text{area of A and B}}{\text{area of B}} = \frac{\text{area of A}}{\text{area of S}}$$

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?



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

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

Chain rule of probability

Board work

The Chain Rule of Probability

The chain rule of probability allows us to write joint distributions as a product of conditional distributions.

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

Notice the pattern here.

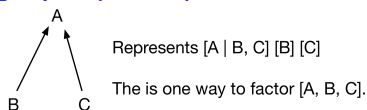
- z 's can be scalars or vectors.
- Sequence of conditioning does not matter.
- When we build models, we choose a sequence that makes sense.

Factoring joint probabilities

Why is factoring useful?

- Factoring joint distributions is how we build Bayesian models.
- The rules of probability allow us to simplify complicated joint distributions, breaking them down into chunks.
- Chunks can be analyzed one at a time.

Consider a factored joint distribution represented by a directed acyclic graph (DAG)



- Directed acyclic graphs (aka 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.
- Nodes at heads of arrows are called “children”; at tails, “parents.”

Factoring joint probabilities

Illustrate with simple regression model on board.

Work on lab

Complete Probability Lab #1