Rules of Probability

Bayesian Models for Ecologists

Tom Hobbs

June 03, 2024

Middlebury Funga The University of Texas at Austin Colorado State University NSF DEB 2042028

Credit to:

Mary Collins
Stony Brook University

Chris Che-Castaldo
USGS and University of Wisconsin

SESYNC
DBI-1052875,1639145
DEB-1145200

Road map for today

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

Random variables

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

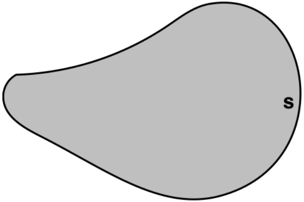
1. Bayesians treat all unobserved quantities as *random variables*.
2. The values of random variables are governed by chance.
3. Probability distributions quantify “governed by chance.”

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

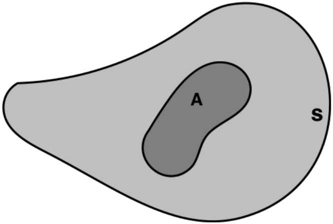
S=Sample space, a.k.a *support*

- The set of all possible values of a random variable.
- The sample space, S has a specific area.
- A specific value of a random variable within S is called an event or an outcome.

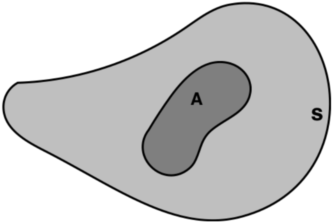


Events in S

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



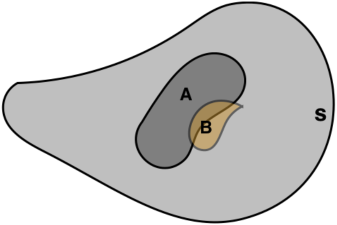
What is the probability of event A?



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

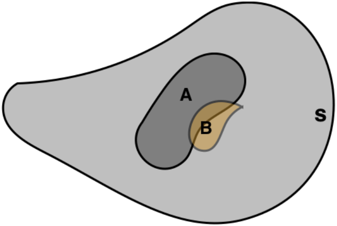
Conditional Probability

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?

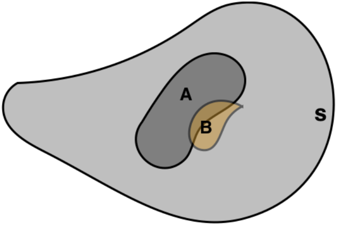


$\Pr(B|A)$ = probability of B conditional on knowing A has occurred

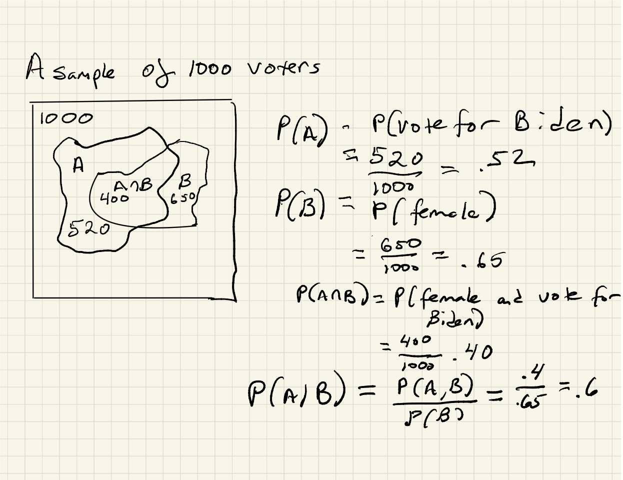
$Pr(B|A) = \frac{\text{Joint Probability}}{\text{Probability of A}} = \frac{\Pr(A,B)}{\Pr(A)}$

Conditional Probability

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



Example



The knowledge that a voter was female changed the probability of a vote for Biden from .52 to .6

Independence

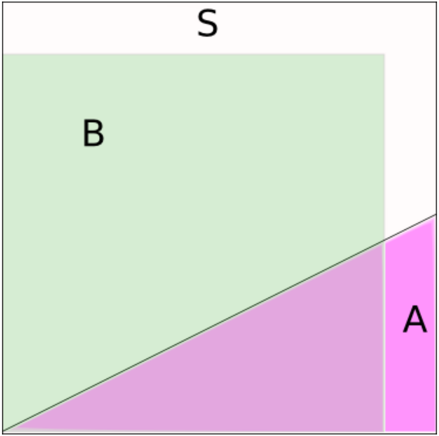
Event A and B are *independent* if the occurrence of event A does not tell us anything about event B.

Events are independent if and only if:

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

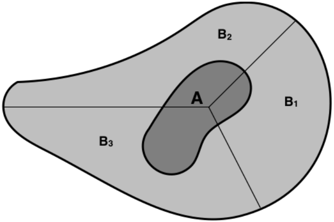
$\Pr(B|A) = \Pr(B)$

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



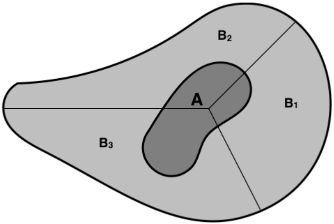
Pr(A) is unknown, but can be calculated using the known probabilities of several events.

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

Rearraging the expression for conditional probability

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

What is the probability of event A?



$\Pr(A) = \sum_n \Pr(A|B_n)\Pr(B_n) = \sum_n \Pr(A, B_n)$ (discrete case)

$\Pr(A) = \int \Pr(A|B)\Pr(B)dB = \int \Pr(A, B)dB$ (continuous case)

Intution about the chain rule of probability

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

We have *factored* $\Pr(A,B)$ into $\Pr(A|B)\Pr(B)$. How else could it be factored?

The Chain Rule of Probability

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

$$\Pr(z_1, z_2, ..., z_n) =$$
$$\Pr(z_n | z_{n-1}, ..., z_1) \Pr(z_{n-1} | z_{n-2}, ..., z_1), ..., \Pr(z_3 | z_2, z_1) \Pr(z_2 | z_1) \Pr(z_1)$$

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.

Example

$$\Pr(z_1, z_2, z_3, z_4, z_5) =$$
$$\Pr(z_5|z_4, z_3, z_2, z_1)\Pr(z_4|z_3, z_2, z_1)\Pr(z_3|z_2, z_1)\Pr(z_2|z_1)\Pr(z_1)$$

How else could this be written?

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 in MCMC.

Problem

Factor $[\sigma, \theta, \alpha, \gamma]$

Decisions on independce allow us to simplify factored expressions

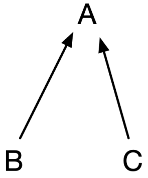
Factoring the joint distribution of A, B, and C:

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

Assuming B and C are independent we can simplify:

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

Consider the directed acyclic graph (DAG)



Represents [A | B, C] [B] [C]

This is *one* way to factor [A, B, C].

- 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 conditioning.
- 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 using DAGs

Some board work.

Work on lab

Complete Probability Lab #1